Kooperatives Experimentieren im Chemieunterricht

Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik

DISSERTATION
zur Erlangung des Doktorgrades
der Naturwissenschaften
- Dr. rer. nat. -

Vorgelegt dem Fachbereich Chemie
der Universität Duisburg – Essen

von
Stefan Rumann
aus Essen

Essen 2004

1. Gutachterin: **Prof. Dr. Elke Sumfleth**

2. Gutachterin: **Prof. Dr. Gisela Lück**

3. Gutachterin: **Prof. Dr. Karin Stachelscheid**

Tag der Disputation: 23.12.2004

Die vorliegende Arbeit wurde in der Zeit von August 2000 bis September 2004 am Institut für Didaktik der Chemie im Fachbereich Chemie der Universität Duisburg-Essen durchgeführt.

Frau Professor Dr. Elke Sumfleth danke ich für die interessante Themenstellung, die umfangreiche und intensive Unterstützung meiner Arbeit sowie die Bereitstellung exzellenter Arbeitsbedingungen.

Frau Professor Dr. Karin Stachelscheid und Frau Professor Dr. Gisela Lück danke ich für die Übernahme der Korreferate sowie ihre Diskussions- und Hilfsbereitschaft.

Herrn Professor Dr. Eckart Hasselbrink danke ich für die Übernahme des Prüfungsvorsitzes.

Ferner danke ich Frau Professor Dr. Elke Wild sowie ihrem Mitarbeiter Josef Exeler von der Arbeitseinheit Pädagogische Psychologie der Universität Bielefeld für unsere gemeinsamen Kooperationsjahre im Schwerpunktprogramm „BIQUA“ sowie den kostenlosen Nachhilfeunterricht in empirisch-statistischer Methodenlehre.

Mein Dank gilt weiterhin den mutigen und neugierigen Lehrern, Schülern und Schulleitern der teilnehmenden Schulen, ohne die diese Arbeit nicht möglich gewesen wäre.

Den Mitarbeitern und Mitarbeiterinnen des Instituts für Didaktik der Chemie danke ich für das freundschaftliche Arbeitsklima und die interessanten Diskussionen.

Herrn Robert Hüllen danke ich für die überaus hilfreiche Unterstützung auf der Zielgeraden.

Abschließend danke ich von ganzem Herzen meinem alten Lehrer, Herrn Prof. Altfrid Gramm, der sich vor vielen Jahren dazu entschloss mir didaktisches Denken beizubringen.

Für Katja und Paul

„Lediglich deskriptiv vorzugehen, Befunde zu notieren und sie in der Hoffnung an die Schulen zurückzuspiegeln, diese könnten und würden dann das Notwendige veranlassen, reicht nicht. Sie haben Anspruch auf Übersetzungshilfe, Entwicklungsarbeit und Unterstützung durch die Wissenschaft. Insofern ist die Phase der Vergleichsuntersuchungen auch die Stunde der Fachdidaktiken, die sich dieser Aufgabe stellen müssen, darin aber auch eine großartige Chance ihrer weiteren Entwicklung bekommen."

Hermann Lange, 2001

Inhaltsverzeichnis

Seite

1. Einleitung 7
2. Unterrichtsqualität als Konstrukt der Unterrichtsforschung 9
   - 2.1 Unterrichtsqualität – was ist guter Unterricht? 10
   - 2.2 Ansätze der Lehr-Lernforschung zur Unterrichtsqualität 11
   - 2.3 Paradigmen und Determinanten zur Erfassung der Unterrichtsqualität 13
3. Kooperatives Arbeiten im Unterricht 21
   - 3.1 Anspruch und Wirklichkeit kooperativer Lernformen 23
     - 3.1.1 Theoretische Annahmen – Warum ist kooperatives Arbeiten erfolgreich? 24
       - 3.1.1.1 Motivationale Perspektiven 24
       - 3.1.1.2 Kognitive Perspektiven 25
     - 3.1.2 Effekte kooperativen Arbeitens 27
       - 3.1.2.1 Empirische Befunde 29
       - 3.1.2.2 Schwierigkeiten in kooperativen Lernarrangements 31
   - 3.2 Formen kooperativen Arbeitens 32
     - 3.2.1 Learning together 33
     - 3.2.2 Schülerteams und Leistungsdifferenzierung 34
     - 3.2.3 Jigsaw I und Group Investigation 35
     - 3.2.4 Jigsaw II 37
4. Zielsetzung und Design der Interventionsstudie 39
   - 4.1 Standards der Untersuchungsplanung 40
   - 4.2 Forschungsfragen und Variablen der Untersuchung 44
     - 4.2.1 Forschungsfragen 44
     - 4.2.2 Untersuchungsvariablen 45
     - 4.2.3 Personengebundene und untersuchungsbedingte Störvariablen 46
   - 4.3 Untersuchungsdesign, Testinstrumentarium und Arbeitsplan 48
     - 4.3.1 Stichprobe und Untersuchungsdesign 48
     - 4.3.2 Erhebungsinstrumentarium 50
     - 4.3.3 Arbeitsplan 55
5. Darstellung der Unterrichtsreihe 57
   - 5.1 Schülerexperimente mit Interaktionsboxen 57
   - 5.2 Die Unterrichtsreihe „saure und basische Lösungen“ 59
     - 5.2.1 Die Vorstunde 63
     - 5.2.2 Die erste Unterrichtsstunde 65
     - 5.2.3 Die zweite Unterrichtsstunde 67
     - 5.2.4 Die dritte Unterrichtsstunde 70

Seite

5.2.5 Die vierte Unterrichtsstunde 72
5.2.6 Die fünfte Unterrichtsstunde 75
5.2.7 Die sechste Unterrichtsstunde 77

6. Auswertung der Schulleistungstests 81

6.1 Eingangsbedingungen und Treatment-Check 81
6.2 Inferenzstatistische Auswertung der offenen Aufgaben 83
6.2.1 Offene Fragen - Haupteffekt >Zeit< 83
6.2.2 Interaktionseffekt 84
6.2.2.1 Interaktionsdiagramme 87
6.2.2.2 Post-hoc-Analysen 88
6.2.3 Analyse ausgewählter offener Transferaufgaben 90
6.3 Inferenzstatistische Auswertung des Multiple-Choice-Tests 93
6.3.1 Multiple-Choice-Test – Haupteffekt >Zeit< 94
6.3.2 Interaktionseffekt 95
6.4 Gruppenvergleiche 96
6.4.1 Kooperative Kleingruppenarbeit vs. lehrerzentrierte Gesprächsführung 98
6.4.2 Treatmenteffekte in Abhängigkeit von der Fachleistungsperformanz 100
6.4.3 Einfluss von Kontrollvariablen 103
6.5 Zusammenfassung der Ergebnisse der Schulleistungstests 107

7. Auswertung der Kleingruppenvideos 111

7.1 Inhaltliche Progressionsanalyse (IPA) der Gruppenarbeitsphase 112
7.1.1 Bezugsrahmen 116
7.1.2 Theoriegehalt 117
7.1.3 Problemelemente 118
7.2 Ergebnisse der inhaltlichen Progressionsanalyse 119
7.2.1 Ergebnisse der Facette >Fokus< 121
7.2.2 Ergebnisse der Facette >Bezugsrahmen< 123
7.2.3 Ergebnisse der Facette >Theoriegehalt< 125
7.2.4 Ergebnisse der Facette >Problemelemente< 126
7.3 Prozessgrafiken zur qualitativen Analyse der Gruppenarbeitsphase 129

8. Zusammenfassung und Ausblick 133

9. Literaturverzeichnis 137

10. Abbildungsverzeichnis 149

11. Tabellenverzeichnis 151

12. Anhang 153

1. Einleitung

„Fast alle Innovationen der letzten dreißig Jahre sind ohne vorherige Erforschung von Lernprozessen eingeführt worden. […] Fast alle haben sich als voreilige Generalisierungen erwiesen, die ein Detail zum Allheilmittel hochstilisierten.“ Derart ernüchternd bilanziert Melenk (2000) den Ertrag unterrichtlicher Reformansätze in den verschiedensten fachdidaktischen Disziplinen der Vergangenheit. Aus diesem Defizit leitet Melenk die Forderung an die fachdidaktische Forschung ab, curriculare Lehr- und Lernmittel künftig stärker als dies bislang der Fall war empirisch zu überprüfen und ihre Bewährung im schulpraktischen Feld zu testen.

Die implizite Forderung nach einer Wirkungskontrolle fachdidaktischer Arbeitsergebnisse zeigt hierbei eine große Nähe zu der in Folge der TIMSS und PISA-Ergebnisse geführten Debatte um unterrichtliche Qualitätssicherung sowie der daraus resultierenden Formulierung von Bildungsstandards und Kerncurricula. Wie, so ließe sich fragen, ist eine Qualitätssicherung fachdidaktischer Forschung zu gewährleisten, so sie selbst den Anspruch erhebt, mit ihren Ergebnissen einen Beitrag zur Verbesserung der Qualität des (naturwissenschaftlichen) Unterrichts zu erzielen? Vereinzelt wird inzwischen versucht, Antworten auf diese Frage zu geben (Ralle & Eilks 2002), dennoch gilt sie zum gegenwärtigen Zeitpunkt aus chemiedidaktischer Sicht als offen.

Das Bemühen indes, auf der Ebene des Unterrichts eine Verbesserung der Situation zu erzielen ist groß. So wird von der Deutschen Forschungsgemeinschaft (DFG) als Reaktion auf den TIMSS-Schock seit dem Jahr 2000 das Schwerpunktprogramm „Bildungsqualität von Schule“ (BIQUA) gefördert, dessen Ziel es ist, die Bedingungen für eine Förderung fachlicher und überfachlicher Kompetenzen im schulischen und außerschulischen Kontext zu erforschen (Doll & Prenzel 2001). Im Rahmen dieses Schwerpunktprogramms wurde ein chemiedidaktisch-psychologisches Kooperationsprojekt bewilligt, welches den Titel trägt: „Schulische und familiale Bedingungen des Lernens und der Lernmotivation im Fach Chemie: Evaluation eines integrierten Interventionskonzeptes zur Säure-Base-Thematik“ (Sumfleth & Wild 2000). Das Projekt zielt auf eine verschränkte Intervention im schulischen und häuslichen Bereich, um so die Verbindungen zwischen dem Schule-Familie-Mesosystem zu verstärken und zu verbessern. Die vorliegende Studie wurde im Rahmen dieses Forschungsvorhabens durchgeführt und fokussiert auf die schulische Intervention sowie die damit verbundene Lernleistung der Schüler[1].

Gegenstand dieser Arbeit ist somit die Entwicklung, Durchführung und Evaluation einer Interventionsstudie zum Themenbereich Säure-Base im Gymnasium der Jahrgangsstufe sieben. Die Zielsetzung der Intervention liegt hierbei nicht auf der Entwicklung neuer Experimente für diesen Themenbereich sondern auf der Erprobung einer spezifischen, kooperativen Arbeitsform des Schülerexperiments. Mit der Evaluation einer Unterrichts*methode* widmet sich die Arbeit einem Forschungsfeld, wel-

[1] Wenn im Folgenden für Personenbezeichnungen nur die männliche Substantivform gewählt wird, so geschieht dies ohne Bezug zum Geschlecht.

ches in den letzten Jahren zunehmend in den Fokus der Chemiedidaktik rückt, wie sowohl die Arbeiten zum offenen Chemieunterricht von Woest (1995) als auch die umfangreichen Publikationen des Ansatzes „Chemie im Kontext“ (s. bspw. Parchmann, Ralle, Demuth 2000) im Allgemeinen belegen.

Die speziell dieser Arbeit zugrunde liegende Methodik des kooperativen Arbeitens im naturwissenschaftlichen Unterricht wurde von Lazarowitz (1998) in die Diskussion gebracht und für den deutschsprachigen Raum erst jüngst von Eilks (2003) aufgearbeitet.
Dabei liegen bislang erst wenige gesicherte Ergebnisse bezüglich der Determinanten vor, die die Unterrichtsqualität im kooperativen Unterricht bestimmen, wie Einsiedler (1997) darlegt: „Forschung ist ganz stark auf die Qualität des lehrergesteuerten Klassenunterrichts zentriert. Wir wissen wenig über die Qualität von Einzelarbeit und Gruppenarbeit, obwohl hier erhebliche Unterschiede hinsichtlich kognitiver Niveaus und aktiver Beteiligung bestehen dürften.“

Theoretisch fußt die Arbeit demnach auf den pädagogisch-psychologischen Forschungsergebnissen zur Unterrichtsqualität im Allgemeinen und zum kooperativen Arbeiten im Besonderen (s. Kp. 2 und 3). Es folgt eine Darstellung des Untersuchungsdesigns der Intervention sowie der intervenierten Unterrichtsreihe selbst (Kp. 4 und 5). Eine Diskussion der Interventionsergebnisse sowohl an Hand quantitativer Leistungsdaten wie auch qualitativer Videoprozessdaten schließt sich in den Kapiteln 6 und 7 an.

Die leitende Forschungsfrage dieser Arbeit sucht zu ergründen, inwieweit sich die Leistungsergebnisse von Schülern, die im Rahmen einer gegebenen Problemstellung einen experimentellen Untersuchungsplan selbstständig erstellen, durchführen und bewerten sollen, von den Leistungsergebnissen der Schüler unterscheiden, die einer Kontrollgruppe in einem lehrerzentrierten Unterrichtsarrangement zugeordnet sind. Gemäß der oben von Melenk angemahnten Defizite seitens der fachdidaktischen Evaluationsforschung, versucht diese Arbeit dabei den Kriterien, die an eine standardisierte Interventionsstudie zur Überprüfung einer Unterschiedshypothese zu stellen sind, gerecht zu werden.

2. Unterrichtsqualität als Konstrukt der Unterrichtsforschung

Seit jeher sieht es die fachdidaktische Forschung als einen Teil ihres Aufgabengebiets an, die Praxis des naturwissenschaftlichen Unterrichts unter verschiedenen Perspektiven zu reflektieren. Vielfach wurden dabei Defizite in der Praxis des Chemieunterrichts konstatiert und ein Reformbedarf angemahnt (vgl. bspw. Gramm 1991, Langensiepen 1995, Schallies 1999). Diese Diskussion wurde in der Vergangenheit meist jedoch allein von Experten – Fachdidaktikern und Fachlehrern – geführt, während ein öffentliches bildungspolitisches Interesse an der Qualität des mathematisch-naturwissenschaftlichen Unterrichts ausblieb.

Dass es sich hierbei zumindest partiell um die Folge eines *nationalen* Phänomens handelt, belegt das Teilnahmeverhalten der Bundesrepublik Deutschland an internationalen Schulleistungsvergleichsstudien (Baumert 1998), wie sie von der „International Association for the Evaluation of Educational Achievement“ (IEA) seit nunmehr fast vierzig Jahren durchgeführt werden: Während die Kultusminister der Länder in den Jahren 1970/71 noch eine Teilnahme an der „Six-Subject-Study“ in zumindest drei Fächern (Naturwissenschaften, Englisch als Fremdsprache und Staatsbürgerliche Erziehung) beschlossen hatten (Postlethwaite, Weiler, Roeder, 1980), so stagnierte in den folgenden zwanzig Jahren die Schulleistungsforschung in Deutschland, wie Ingenkamp und Schreiber (1989) mit Verweis auf die Nichtteilnahme an den IEA-Studien der achtziger Jahre bemerken.

Die Ergebnisse der Vergleichsstudien TIMSS/II (Baumert, Lehmann, Lehrke 1997) und TIMSS/III (Baumert, Bos, Watermann 1998) sowie der ersten PISA-Studie (Baumert, Klieme, Neubrand et al. 2001, Kirsch et al. 2002) haben diesen Sachverhalt kontinuierlich und grundlegend verändert und eine in einem bislang nicht bekanntem Ausmaß öffentliche Diskussion über die Bildungsqualität des mathematisch-naturwissenschaftlichen Unterrichts in Deutschland eröffnet. Themen wie „Bildung und Innovation“ bestimmen die politische Auseinandersetzung im deutschen Bundestag (Bulmahn, 2002). Die Ergebnisse der Ländervergleichsstudie (Baumert, Artelt, Klieme et al. 2002) werden umgehend von der Tagespresse rezipiert (vgl. bspw. FAZ, SZ, FR, Die Welt vom 24.06.2002 als Reaktion auf die Veröffentlichung von PISA-E vom Vortag) und häufig schon vor der offiziellen Pressekonferenz in Tageszeitungen publiziert.

Betrachtet man die Ergebnisse der Large-Scale-Assessments seit TIMSS/II, so fällt die Konsistenz der zentralen Aussagen auf, woraus resultiert, dass eine infolge der ersten TIMS-Studie noch lebhaft geäußerte Kritik (vgl. z. B. Hagemeister 1999) gegenwärtig kaum mehr zu verzeichnen ist, ferner ein Konsens bezüglich der Validität der Ergebnisse vorliegt und sich somit „jede Diskussion über die Richtigkeit und Relevanz der Ergebnisse oder über die Angemessenheit der Aufgaben erübrigt“ (Fischer & Sumfleth 2002).

Inhaltlich bescheinigen die Schulleistungsstudien der letzten Jahre in ihrer Quintessenz den deutschen Lernenden insbesondere mangelnde Kompetenzen beim Transfer sowie beim Problemlösen, also bei Aufgaben, die ein konzeptuelles Verständnis

und flexibel anwendbares Wissen erfordern. Damit wird letztlich die von den Fachdidaktikern schon lange postulierte „Krise“ des mathematisch-naturwissenschaftlichen Unterrichts an Hand einer breiten Datenbasis abgesichert, empirisch belegt und spezifiziert (vgl. Bühler & Graf 2003).

Als unmittelbare Reaktion auf das erwartungswidrig schlechte Abschneiden der deutschen Lernenden ist gegenwärtig in der Unterrichtsforschung ein Paradigmenwechsel zu verzeichnen, der sich methodisch durch eine stärkere Ausrichtung der Forschungsprogramme auf an messbaren Schulleistungen orientierte Evaluationsstudien bei einer inhaltlichen Fokussierung auf die Unterrichtsqualität auszeichnet. Damit, so Lange (1999), gilt es die „empirische Wende“, wie sie bereits von Heinrich Roth vor etwa dreißig Jahren gefordert wurde, in Bildungspolitik und Schulalltag zu vollziehen. Weiter ordnet Lange (2001) demzufolge das „Entwickeln“ dem „Messen“ unter und erklärt: „Pädagogisches Handeln verfolgt Ziele und will Wirkungen erzeugen. Die Ziele müssen formuliert und damit überprüfbar, die Wirkungen geklärt und damit auch notfalls kritisierbar und veränderbar gemacht werden.“

Für die Chemiedidaktik stellt die Umsetzung dieser Forderung einen programmatischen Kurswechsel dar. Dominierte doch in der Vergangenheit die Vorstellung, dass eine Verbesserung des Unterrichts vor allem durch eine Veränderung inhaltlicher und curricularer Rahmenvorgaben zu erwirken sei, so gilt es nun zu formulieren, welchen Beitrag die Fachdidaktik zur normativen Definition und zur evaluativen Kontrolle (*monitoring*) von Unterrichtsqualität im Chemieunterricht zu leisten hat. Helmke (2002: 261) prognostiziert den Wechsel in der bildungspolitischen Diskussion „von einer input- und ressourcenorientierten Sichtweise […] hin zu einer „output“-orientierten Perspektive“, mit Verweis auf den durch die PISA-Ergebnisse ausgelösten Schock, gar als „nicht mehr rückgängig zu machen“.

Insofern verpflichtet sich chemiedidaktische Unterrichtsforschung nach PISA intensiver noch als bislang, an Hand von deskriptiven wie auch präskriptiven, wirkungskontrollierten Forschungsstudien ihren Beitrag zur Verbesserung der Unterrichtsqualität im Chemieunterricht zu leisten. Die Reforminitiativen auf allgemein-naturwissenschaftlicher Ebene wie auch speziell auf chemiedidaktischer Ebene, die im Zuge der PISA-Folgeforschung bislang ergriffen wurden, werden noch näher dargestellt. Zunächst soll jedoch das bereits erwähnte, zentrale theoretische Metakonstrukt der „Unterrichtsqualität“ näher thematisiert werden.

2.1 Unterrichtsqualität – was ist guter Unterricht?

Die Frage nach einer Definition von „Unterrichtsqualität“ ist genuin verbunden mit einer normativ-wertenden Vorstellung dessen, was guter Unterricht an sich sei (Ditton 2002, Weinert 1998). Wer über Unterrichtsqualität spricht, eine Verbesserung der Unterrichtsqualität einfordert, muss ferner verdeutlichen, ob er sich mit dieser Forderung an den Zielen oder an der Effektivität der zur Erreichung der Ziele eingesetzten Mittel orientiert. Das wiederum zieht die Frage nach sich, ob der „gute“ Unterricht an normativen Zielvorstellungen oder aber an Hand der gemessenen Wirkungen und

Effekte zu beurteilen ist. Aus dieser Zweiteilung kann gegebenenfalls ein „Wertigkeitsdilemma“ erwachsen (Ditton 2002), sofern einem intendierten, normativen Unterrichtsanspruch nur wenig Wirkung nachgewiesen werden kann, oder aber, im umgekehrten Falle, wenn sich ein „wenig wünschenswerter - z.B. rigide-direktiver - Unterricht als effektiver erweisen sollte, z.B. bezüglich der Leistungsentwicklung der Schüler“ (a.a.O.: 199).

Wenn im Folgenden normative Aspekte der Unterrichtsqualität zunächst weitestgehend ausgeblendet werden sollen, so gilt es dennoch zu berücksichtigen, dass bei prozessorientierten Zielsetzungen, Mittel und (normative) Ziele häufig kongruieren. So kann beispielsweise der Unterrichtsaspekt „Förderung des sozialen Lernens der Schüler“ sowohl Zieldimension als auch das im Unterricht zu erfassende Mittel selbst sein (vgl. Clausen 2002: 16).

2.2 Ansätze der Lehr-Lernforschung zur Unterrichtsqualität

Die Frage, wie sinnvolles, lernwirksames Vermitteln von Chemie zu gestalten sei, ist vergleichsweise alt. Nur wenige Jahrzehnte nachdem die Chemie eine eigenständige Wissenschaft wurde, gab es bereits die ersten Publikationen (Marcet 1839), die eine Auseinandersetzung mit diesem Thema dokumentieren. Man mag diesen Punkt als die Geburtsstunde der Chemiedidaktik betrachten, empirisch kontrollierte Evaluationsstudien bezüglich des Qualitätsaspektes beim Vermitteln von Chemie gingen damit noch nicht einher. Vielmehr ist bis in die jüngste Gegenwart der Topos „Unterrichtsqualität“ originär eine Domäne der empirisch-pädagogischen bzw. pädagogisch-psychologischen Forschung, die sich seit den frühen sechziger Jahren des letzten Jahrhunderts dieser Thematik widmet (vgl. hierzu bspw. Bloom et al. 1956, Flanders 1965, Carroll 1973).

Carroll (1973) legt in seinem „Modell schulischen Lernens“ zur näherungsweisen Quantifizierung des Lernerfolgs („*degree of learning*“) ein Gleichungsmodell vor, in der Unterrichtsqualität als eine von fünf Variablen eingeht. In Abgrenzung zu den weiteren vier von Caroll betrachteten Determinanten des Lernerfolgs, der zugestandenen *Lernzeit*, der schülerspezifischen *Ausdauer*, der Fähigkeit *Unterricht zu folgen* sowie der *Begabung* des Lernenden, bezeichnet Carroll die Unterrichtsqualität als die am schwierigsten zu messende Größe (Caroll 1973). Als Merkmale von Unterrichtsqualität nennt Carroll: instruktionale *Verständlichkeit* des Lehrenden, inhaltliche *Sequenzierung* der Aufgabenstellung und die voraussetzungsgebundene *Adaptivität* der Instruktion.

Das dem Modell des schulischen Lernens zugrundeliegende Verständnis der Erfassung von Lernergebnissen an Hand der Lernrate war zunächst noch sehr enggefasst und wurde in den siebziger Jahren von Bloom im Rahmen seines Konzepts des *zielerreichenden Lernens* (learning for mastery, 1976) maßgeblich erweitert. Bloom berücksichtigt in diesem Modell neben einer effizienzorientierten Lernrate auch quantitative und qualitative Aspekte (Leistungs*höhe* und Leistungs*art*) sowie affektive Lernergebnisse. Aus der so formulierten Zielperspektive resultiert, dass, neben dem

bereits bei Caroll postulierten Einfluss von Unterrichtsqualität und kognitiven Eingangsvoraussetzungen der Schüler, auch motivationale Schülermerkmale als Eingangsvoraussetzung berücksichtigt werden müssen (vgl. Abb. 2.1). Ähnlich wie Caroll definiert Bloom Unterrichtsqualität als Näherungswert manifester Variablen der Lehrer-Schüler-Interaktion. Die drei entscheidenden Aspekte der Interaktion sind hiernach die Qualität der Instruktion, die Quantität der aktiven Schülerbeteiligung sowie das Verwenden positiver und negativer Bestärkungen im Lernprozess (Bloom1971).

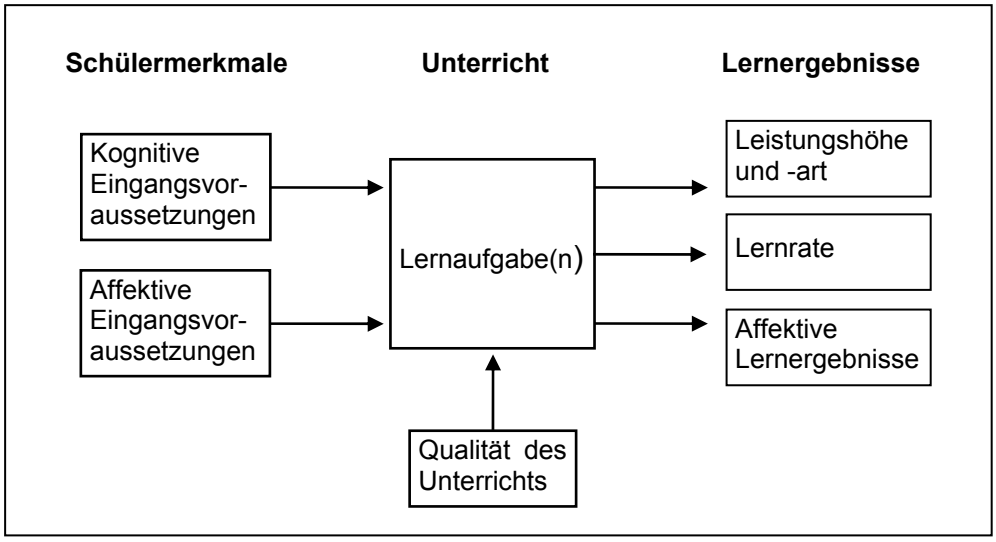


Abbildung 2.1: Modell des schulischen Lernens von Bloom (nach Helmke, Weinert 1997)

Mit der Umsetzung der dargestellten Wirkzusammenhänge im Konzept des mastery learnings verbindet Bloom den programmatischen Anspruch, nahezu jeder Schüler einer Klasse (explizit bis zu 95%) könne, bei entsprechender Gewährung und Nutzung der individuell benötigten Lernzeit, ein definiertes, anspruchvolles Leistungsniveau erreichen. Wenngleich diese Zielsetzung empirisch nicht bestätigt werden konnte, so verband sich implizit damit die prospektive Vision von einem Unterrichtsmodell mit dem Potential einer interpersonellen Divergenzminimierung. Damit wurde ein normativer Aspekt der Unterrichtsqualität antizipiert, der infolge der drastischen Leistungsvarianz deutscher Schüler bei den internationalen Vergleichsstudien erst in jüngster Zeit wieder in die Diskussion eingebracht wird (vgl. Clausen 2002).

Im weiteren Verlauf der siebziger und achtziger Jahre wurde die Forschung zum effektiven Lehrerhandeln zunehmend intensiviert, wie die zahlreichen Metaanalysen[2], ohne die ein repräsentativer Überblick über den jeweiligen Forschungsstand längst nicht mehr zu gewährleisten wäre, belegen (Walberg 1986, Brophy & Good 1986, Wang, Haertel, Walberg 1990, 1993, Shuell 1996). Dabei sind die Forschungsergebnisse ebenso vielschichtig wie in ihren zentralen Aussagen inkonsistent (vgl. Kp. 2.1.2). Einsiedler (1997) beklagt beispielsweise die endlose Zahl der untersuchten Variablen, die nach mehr oder weniger willkürlichen, meist jedoch theoriefernen Kri-

[2] Glass (1990) definiert eine Metaanalyse als Aggregierung der Befunde unabhängiger, inhaltlich homogener Primäruntersuchungen, mit dem ziel der integrativen, komprimierten Ergebnisdarstellung in einem Forschungsfeld (vgl. auch Bortz,Döring 1995).

terien gebündelt werden. Konsens bezüglich der Schulleistungsdeterminanten besteht lediglich in der überragenden Bedeutung der individuellen – kognitiven und motivationalen – Eingangsvoraussetzungen (zu diesen Konstrukten werden mittlere Korrelationen von 0.5 bis 0.6 angegeben) sowie der familiären Einflussfaktoren. Vor diesem Hintergrund nimmt sich der Einfluss der Unterrichtsqualität auf die Schulleistung nur ausgesprochen bescheiden aus, wenngleich der varianzanalytische Anteil der Unterrichtsqualität an der Schulleistung inzwischen als bedeutsamer erachtet wird als noch vor einigen Jahrzehnten. So kommt Jencks (1972, nach Einsiedler 1997) zu dem Ergebnis, dass unter Kontrolle des familialen Hintergrunds die Qualität des Schulunterrichts mit einer Varianzaufklärung von nur drei Prozent in die Leistungsentwicklung der Schüler eingeht. Bloom (1971) erklärt auf Grund untersuchter Interaktionsprozesse zwischen Lehrern und Schülern hingegen 25% der Schulleistungsvariation über die Unterrichtsqualität, welche allerdings nur unspezifisch zu diagnostizieren, nicht jedoch kausal über singuläre Lehrervariablen determiniert werden kann. Auch in Blooms Modell wird die große Bedeutung dessen deutlich, was der Schüler prädispositionell in den Unterricht einbringt, da Bloom durch die kombinierte Berücksichtigung kognitiver und affektiver Eingangsvoraussetzungen bis zu zwei Dritteln der Varianz der Leistungsentwicklung erklärt.

Als weitere Quelle zur Bedeutung von Unterrichtsqualität auf die Leistungsentwicklung sei auf Walbergs Modell der „Educational Productivity“ verwiesen (Walberg 1981, nach Clausen 2002), welches der Unterrichtsqualität neben der Unterrichtsquantität jeweils 15% Varianzaufklärung zuspricht.

Vor dem Hintergrund der geschilderten Datenlage mögen die Bemühungen um eine Optimierung schulischer Lehr-Lernprozesse in ihrer Auswirkung auf den individuellen Lernerfolg als Marginalie erscheinen, Unterrichtsqualität ist jedoch, im Gegensatz zu den dargestellten Eingangsvoraussetzungen sowie des sozialen Hintergrunds derjenige Prädiktor, der einer externen Gestaltung durch den einzelnen Lehrer am effektivsten zugänglich ist und somit im Rahmen der Unterrichtsforschung einen bedeutsamen Ansatzpunkt für eine Steigerung des Lernerfolgs darstellt (vgl. Klieme 1999).

2.3 Paradigmen und Determinanten zur Erfassung der Unterrichtsqualität

Versucht man die Determinanten der Unterrichtsqualität näher zu spezifizieren, so muss berücksichtigt werden, dass die Entwicklung der empirischen Unterrichtsforschung durch eine Abfolge von Forschungsparadigmen zum erfolgreichen Unterrichten gekennzeichnet ist, die ihrerseits durch unterschiedliche Grundannahmen verschiedene Kategorien von Determinanten bedingen. Nach Rheinberg et al. (2001) lauten die Forschungsparadigmen:

a) Das Persönlichkeitsparadigma

b) Das Prozess-Produkt-Paradigma

c) Das Experten-Paradigma

Das Persönlichkeitsparadigma war durch die Suche nach allgemeinen Persönlichkeitsmerkmalen des „guten Lehrers“ gekennzeichnet. Methodisch versuchte man die mit psychometrischen Tests erhobenen Lehrermerkmale (z.B. Intelligenz, Führungsstil, fachliche und didaktische Fertigkeiten) mit Leistungsdaten der Lernenden zu korrelieren. Die erhobenen statischen Zusammenhängen waren jedoch durchweg schwach und lieferten nur eine varianzanalytische Aufklärung von ca. 10% (a.a.O. 298), so dass dieser Forschungsansatz heute als gescheitert gilt.

Bedeutsamer erweisen sich die aus dem angloamerikanischen Raum stammenden, dem Prozess-Produkt-Paradigma verhafteten Forschungen, welches sich durch folgende Schritte auszeichnet (nach Helmke 2003):

1. Unter *Prozessen* werden bestimmte, zu erfassende Aspekte des Unterrichtsverhaltens, meist Verhaltensweisen der Lehrkraft, verstanden.
2. Die *Produkte* werden an Hand formulierter Zielkriterien erfasst, meist als Maß der Schulleistung in dem betreffenden Fach.
3. Schließlich wird zwischen dem *Prozess* der Unterrichtsmaßnahme und dem *Produkt*kriterium ein (vorzugsweise korrelativer) *Zusammenhang* berechnet.

Übereinstimmend definieren Anderson, Everston und Brophy (nach Helmke & Weinert 1997: 128) die zentrale Aufgabe der Prozess-Produkt-Forschung als Ermittlung von „relationships between what teachers do in the classroom (the process of teaching) and what happens in their students (the product of learning). One product that has received much attention is achievement in the basic skills.” Die explizite Betonung des Lehrerhandelns (“what teachers do”) als Prozessvariable macht ferner die starke Fokussierung auf einen lehrerzentrierten Klassenunterricht als dominierende Unterrichtsform der Prozess-Produkt-Forschung deutlich. Als prototypisches Beispiel für diese Sichtweise kann das Sammelreferat von Rosenshine (1979) zur so genannten direkten Instruktion (direct instruction) herangezogen werden. Das Konzept der direkten Instruktion beschreibt einen Unterrichtsstil, der als zentrales Kriterium für den Lehrerfolg den Lernzuwachs heranzieht und der in Bezug auf dieses Kriterium in seiner Effizienz anderen Unterrichtsformen deutlich überlegen ist. Kennzeichnend für das Konzept der direkten Instruktion ist – neben einer straffen Unterrichtsführung durch den Lehrer – eine Einschränkung der Handlungsfreiheit der Schüler, ein intensives Monitoring sowie eine stark strukturierte Lehrer-Schüler-Interaktion. Keineswegs ist dies jedoch gleichzusetzen mit einem dogmatisch-dozierenden Lehrervortrag, da der Lehrer in zweckmäßiger Weise auch unterschiedliche Sozialformen in das Planungsraster einbezieht, Übungsphasen anbietet und die individuellen Lernfortschritte kontrolliert. Begünstigend kommt durch die starke Lehrerzentrierung eine optimale Nutzung der Unterrichtszeit hinzu (Einsiedler 1997, Helmke & Weinert 1997).

Wenngleich die Reduzierung des Gütekriteriums auf den Lehrerfolg häufig ein Ansatzpunkt der Kritik ist, so liegt die besondere Bedeutung dieses Unterrichtskonzept in seiner Erklärungsmächtigkeit hinsichtlich der Determinanten einer optimalen Strukturierung des Unterrichts. Die Lehrwirksamkeit der direkten Instruktion gehört zu den

wissenschaftlich am besten gesicherten Erkenntnissen der Unterrichtsqualitätsforschung, weshalb Clausen (2002) auch auf die Referenzfunktion für andere Unterrichtsformen hinweist.

Zu der in der pädagogischen Psychologie am häufigsten rekurrierten Gesamtdarstellung zum Zusammenhang zwischen Lehrerhandeln und Schülerleistung gehört das bereits erwähnte Sammelreferat von Brophy & Good (1986). Aus einer chronologischen Darstellung der Prozess-Produkt-Forschungsprogramme der sechziger, siebziger und frühen achtziger Jahre leiten die Autoren eine systematische Integration stabiler Variablen der Unterrichtsqualität ab. Die Variablen sind ferner dadurch bestimmt, dass sie sich unabhängig vom jeweiligen Inhaltsbereich sowohl nieder- als auch hoch-inferent als Unterrichtsinteraktionen erfassen lassen und damit eine allgemeine Beschreibungsform von Unterrichtsqualität gestatten (vgl. Bromme 1995). In der folgenden Tabelle sind die zentralen leistungsförderlichen Strukturmerkmale von Unterricht nach Brophy und Good aufgeführt:

Allgemeiner Unterrichtsaspekt		**Unterrichtsdeterminanten**
Unterrichts-quantität	Pacing/ Lerntempo	▪ Lerngelegenheit/ Inhaltsbereich ▪ Rollenselbstverständnis/ Erwartungshaltung/ zeitliche Gestaltung ▪ Unterrichtsmanagement/ Organisationsvermögen ▪ konsistente Zielorientierung ▪ aktives Lehren
Unterrichtsqualität	Informationsdarbietung	▪ Strukturierung ▪ Redundanz/ Sequenzierung ▪ Klarheit (Prägnanz) ▪ Enthusiasmus ▪ Informationsrate/ -tempo
	Gestaltung/ Moderation der Lehrerfragen	▪ Schwierigkeitsgrad ▪ Kognitive Anforderung ▪ Klarheit ▪ Wartezeit bis zum Aufrufen ▪ Auswahl des antwortenden Schülers ▪ Wartezeit nach Aufruf eines Schülers
	Reaktion auf Schülerbeiträge	▪ Reaktion auf richtige Antworten ▪ Reaktion auf teilrichtige Antworten ▪ Reaktion auf falsche Antworten ▪ Reaktion auf ausbleibende Antworten ▪ Reaktion auf Schülerfragen, -kommentare
	Einzel-/ Hausarbeit	▪ Aufgabenerläuterung
	Klassen- vs. Gruppen- vs. individuelle Instruktion	▪ *keine empirisch gesicherten Determinanten vorhanden*

Tabelle 2.1: Schulleistungsrelevante Instruktionsdeterminanten (nach Brophy & Good 1986; vgl. Weinert, Schrader, Helmke, 1989, Einsiedler, 1997, Seidel 2002)

Obschon zentrale Aussagen dieses Referats durch neuere Studien bestätigt wurden (vgl. im deutschsprachigen Raum bspw. Helmke & Weinert 1997), wird in der gegenwärtigen Forschung die Bildung linearer, unidirektionaler Zusammenhänge als prinzipielle Methode zur Erkenntnisgewinnung über Unterrichtsprozesse kritisch diskutiert. Helmke und Schrader (1998) betonen die Notwendigkeit, Unterrichtseffekte nicht auf einzelne, isolierte Variablen zurückzuführen (variablenorientierter Ansatz), sondern vielmehr das interdependente Zusammenwirken mehrere Unterrichtsdeterminanten zu berücksichtigen. Ergänzend weist Shuell (1996: 739) auf die Gefahr der Ausblendung von, gerade aus einer modernen konstruktivistischen Sicht bedeutsamen, *Mediationsprozessen* hin: “one of the more serious limitations of process-product research is its failure to investigate the cognitive (and other psychological) processes of students that mediate the teacher’s behaviour and its effect on student outcomes“.

Als Beispiel für die komplexe Determination der Schulleistung sei auf ein Ergebnis der Hauptschulstudie „Unterrichtsqualität und Leistungszuwachs“ von Weinert, Schrader und Helmke (1989) verwiesen. Zwar konnte in Übereinstimmung mit den Annahmen der Theorie der „direct instruction“ gezeigt werden, dass eine effiziente Klassenführung, verbunden mit einer gezielten Kontrolle des Lernverhaltens der Schüler, sich als sehr leistungsförderlich erweist. Überraschend ist jedoch der Einfluss der didaktischen Strukturierungshilfen des Lehrers auf den Lernerfolg, da diese mit der diagnostischen Kompetenz des Lehrers konfundiert, wie folgende Abbildung zeigt (aus: Max-Planck-Gesellschaft 1986).

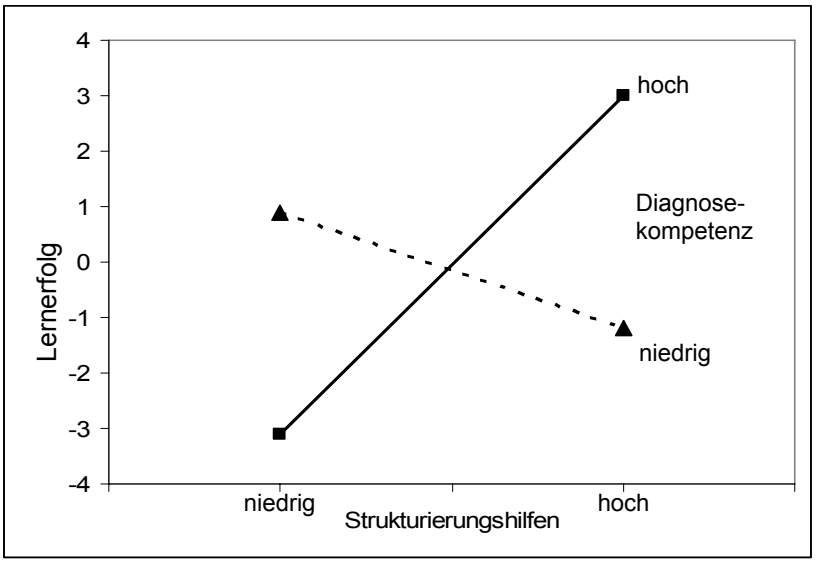


Abb.: 2.2 Lernerfolg in Abhängigkeit von der Diagnosefähigkeit des Lehrers und dem Strukturierungsgrad des Unterrichts.

Aus der Abbildung geht hervor, dass sich die Unterrichtsdeterminanten „Strukturierungshilfen“ und „diagnostische Kompetenz“ nur in Kombination als lernförderlich erweisen, wohingegen der intensive Einsatz von Strukturierungshilfen bei niedriger diagnostischer Kompetenz des Lehrenden sogar kontraproduktiv auf die Lernleistung wirkt. Die Autoren interpretieren dies, indem sie auf die Notwendigkeit der „Passung“ von angebotener Strukturierungshilfe in Abhängigkeit von der Lernausgangslage der einzelnen Schüler hinweisen: „the positive effects of structuring cues depends on

wether the teacher adapts them to students needs“ (Weinert, Schrader, Helmke 1989: 904).

Eine isolierte Betrachtung der einzelnen Unterrichtsvariablen führt darüber hinaus in unabhängigen Studien häufig zu einer völlig unterschiedlichen Gruppierung gleicher Variablen, weshalb Einsiedler (1997) sowohl das Fehlen einer systematischen Kategorisierung von Unterrichtsmerkmalen als auch die weitestgehend atheoretische Herangehensweise in den Forschungsprogrammen bemängelt. Diese Kritik betrifft auch die fortschreitende Identifizierung und Differenzierung lernprozessrelevanter Variablen, ohne diese jedoch in ein umfassendes, konsistentes Konzept einbinden zu können. So tragen beispielsweise Wang, Haertel und Walberg (1990) auf der Basis einer Metaanalyse nicht weniger als 228 (!) den Lernprozess beeinflussende Variablen zusammen, die sie auf dreißig Skalen in sechs Kategorien verteilen. Vergleichbare Befunde finden sich auch in anderen Übersichtsartikeln. Helmke und Weinert (1997: 125) pointieren diesen Sachverhalt mit dem Hinweis, man könne zu dem Eindruck gelangen „im Unterricht wäre alles und jedes irgendwie wichtig und zugleich auch wieder unwichtig“.

Schließlich sei auch noch der geringe Aufklärungsgrad des variablenorientierten Ansatzes für die Schulleistung erwähnt, da der ermittelte korrelative Zusammenhang im Allgemeinen sehr gering ist: So weisen Fraser et al. (1987) in einer Synthese mehrerer Metaanalysen (eine so genannte „Megaanalyse“) keiner untersuchten Unterrichtsvariablen einen höheren korrelativen Zusammenhang mit der Schulleistung nach als 0.29. Zusammenfassend bilanzieren Helmke und Weinert (1997: 130) vor dem Hintergrund der Vielzahl der unabhängig beschriebenen Einflussfaktoren lakonisch die Gesamtsituation zur Unterrichtsqualitätsforschung mit den zwei folgenden Postulaten:

„Postulat 1: Lehrer können sowohl guten als auch schlechten Unterricht auf eine sehr verschiedene Weise halten.

Postulat 2: Gleich erscheinende Verhaltensweisen des Lehrers können unter verschiedenen Bedingungskonstellationen […] völlig unterschiedliche Wirkungen haben.“

Aus den genannten Gründen ist es daher wenig verwunderlich, dass die alleinig dem Prozess-Produkt-Paradigma verhafteten Forschungsansätze sich zunehmend der Kritik ausgesetzt sehen. Neben den dargestellten Schwächen bei der eindeutigen Identifizierung von Wirkzusammenhängen gründet dies auch in der Ausblendung so genannter moderierender Prozesse, die zwischen den Instruktionsvariablen und dem Produkt in Form der Lernleistung geschaltet sind und die zunehmend als lernbedeutsam erachtet werden. Ditton (2002) spricht daher auch von einem *black-box* Charakter des Modells.

Die dargestellte und sicherlich auch berechtigte Kritik an der Prozess-Produkt-Forschung darf jedoch nicht darüber hinwegtäuschen, dass die inzwischen postulierte Komplexität des Systems „Unterricht“ sowie die Interdependenz der zugrundeliegenden Variablen überhaupt erst durch die intensive Prozess-Produkt-Forschung

ermöglicht wurde. Damit wird auch deutlich, dass ein Zurückgreifen auf dieses Paradigma bei einer – bedingt durch die Forschungsfrage – Beschränkung auf einzeln isoliert zu analysierenden Unterrichtsphasen (z.B. Gruppenarbeit) aufgrund der damit verbundenen Dimensionsreduzierung durchaus eine Berechtigung haben kann. Darüber hinaus gewährleistet die Prozess-Produkt-Forschung eo ipso eine hohe ökologische Validität und erscheint daher in unterrichtsnahen Feldstudien als geeigneter Ansatz, wie Seidel (2002) darlegt.

Die ungelösten Probleme des Prozess-Produkt-Paradigmas – wie auch die durch diese Forschungsposition überhaupt erst aufgeworfenen Fragen – führten zu Beginn der neunziger Jahre zur Entwicklung des Expertenparadigmas, das den zunächst letzten abgeschlossenen Stand im Theoriewandel der Lehrerbildung markiert (Rheinberg et al. 2001). Kennzeichnend für das Expertenparadigma sind die Abkehr von einer behavioristisch-technizistisch überhöhten Erwartungshaltung an die Lernwirksamkeit kontextunabhängiger Lehr*fertigkeiten* des Lehrers und eine erneute Hinwendung zu seiner *Person*. Im Gegensatz zum Persönlichkeitsparadigma der fünfziger und sechziger Jahre fokussiert das Expertenparadigma jedoch nicht auf die Identifikation (vermeintlich) unterrichtsrelevanter Charaktereigenschaften des Lehrers, sondern betrachtet den erfolgreichen Lehrer als kompetenten *Experten* für die Gestaltung von Lern*gelegenheiten*.

Ausgangspunkt dieses Paradigmenwechsels ist die (schmerzliche) Erkenntnis, wonach die im Prozess-Produkt-Paradigma untersuchten Handlungsfertigkeiten des Lehrers nur auf einer sichtbaren, niedrig-inferenten Beobachtungsebene die Verhaltensweisen der Schüler beeinflussen, ohne jedoch eine direkte Aussage über deren Lernleistung treffen zu können. Demzufolge legt Slavin (1989) ein Unterrichtsmodell vor, in dem die Funktion des Lehrers nicht länger allein in der Tätigkeit des Lehrens, denn viel mehr im angemessenen Bereitstellen von Lerngelegenheiten gesehen wird.

Gemeinsam ist Prozess-Produkt- und Expertenparadigma die Suche nach dem erfolgreichen Lehrer. Zerfiel jedoch im Prozess-Produkt-Paradigma „die Person des Lehrers in ein Bündel von Teilfertigkeiten, von denen man gar nicht erwarten konnte, dass sie alle in einer Person realisiert sein konnten, [so rückt durch] die erhebliche Kontextabhängigkeit erfolgreichen Lehrerhandels […] wiederum die Person des Lehrers in den Blickpunkt der Analyse“ (Bromme 1997). ’Welches Wissen, welches Können zeichnet den erfolgreichen Lehrer aus?’ lautet demnach die leitmotivische Fragestellung des Expertenparadigmas. Rheinberg et al. (2001) sehen hierin mit Hinblick auf Fragestellung, Forschungsmethoden und Sichtweisen eine programmatische Nähe zur kognitionspsychologischen Expertiseforschung.

In der jüngsten Entwicklung der Unterrichtsforschung scheint sich eine Abkehr von paradigmatischen Grundpositionen hin zu pragmatischen Modellbildungen anzudeuten. So legt Helmke (2002) ein Angebots-Nutzungs-Modell des Unterrichts vor, das gleichermaßen die Wirkungen des Unterrichts als auch Aspekte der Lehrerpersönlichkeit in Betracht zieht.

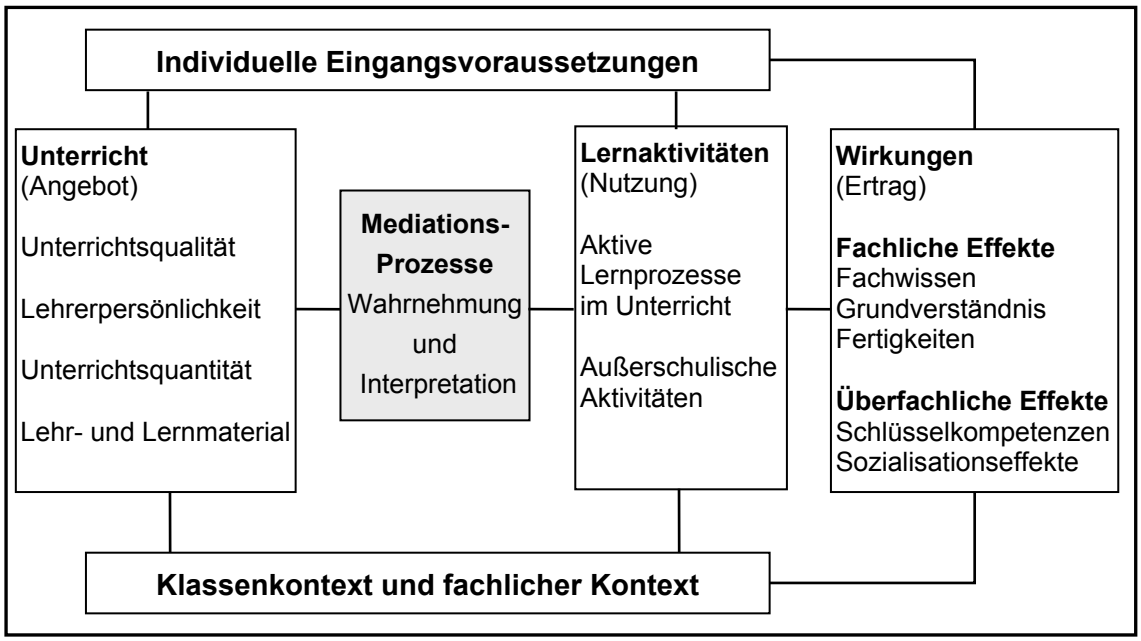


Abb. 2.3: Angebots-Nutzungs-Modell des Unterrichts nach Helmke 2002

Helmke betont, dass die intendierte Wirkung des Unterrichts nicht notwendigerweise allein durch einen konkreten Unterricht erzielt werden kann, sondern dass Unterricht in diesem Modell als *Angebot* zu verstehen ist, dessen Wirkung von einer Vielzahl an Faktoren abhängig ist. Besondere Bedeutung kommt hierbei den Mediationsprozessen zu, welche über individuelle Verarbeitungsprozesse die Wirkung von Unterricht auf Schülerebene vermitteln. Helmke (2003) unterscheidet hierbei bei den Vermittlungsprozessen auf Seiten der Schüler zwischen solchen, die auf einer kognitiven Wahrnehmung beruhen (werden die unterrichtlichen Maßnahmen überhaupt vom Schüler wahrgenommen) und solchen Vermittlungsprozessen, die zu motivationalen, emotionalen und volitionalen Reaktionen der Schüler führen. Ob und welche Lernaktivitäten auf Schülerseite als Folge des Unterrichts resultieren, so Helmke, hängt in entscheidendem Maße von dem Einfluss dieser beiden Formen der Mediationsprozesse ab.

3. Kooperatives Arbeiten im Unterricht

Wie wohl kaum ein anderer Begriff im unterrichtsmethodischen Sprachgebrauch wird der Terminus der *Kooperation* in Lehr-Lern-Arrangements gleichsam häufig wie auch unscharf verwendet. Entsprechend vielschichtig fällt bei Sichtung der Literaturlage die Bandbreite der Definitionen aus. Häcker und Stapf (1998) definieren eine Kooperation aus einer ergebnisorientierten Perspektive, indem sie ausführen, „das Erreichen eines individuellen Ziels erhöht die Wahrscheinlichkeit, dass andere ebenfalls ein entsprechendes Ziel erreichen“. Slavins Verständnis von Kooperation ist hingegen stärker auf den schulischen Kontext ausgerichtet und charakterisiert kooperatives Lernen durch das Zusammenarbeiten von Schülern in kleinen Gruppen, um sich beim Lernen des Stoffes gegenseitig zu helfen (Slavin 1989). Diese Definition von kooperativem Lernen macht deutlich, dass neben der fachlichen Leistungsentwicklung an diese unterrichtliche Organisationsform eine Erwartungshaltung hinsichtlich des Erwerbs prosozialer Kompetenzen geknüpft wird. Ähnlich äußern sich auch Johnson und Johnson (1994a), die *cooperation* bzw. *cooperative learning* wie folgt definieren: „*Cooperation* means working together to accomplish shared goals. Within cooperative activities individuals seek outcomes that are beneficial to themselves *and* beneficial to all other group members. Cooperative learning is the instructional use of small groups that allows student to work together to maximize their own and each other’s learning.” Allgemein lassen sich an Hand dieser Definitionen folgende Kennzeichen kooperativen Arbeitens identifizieren:

- Zusammenarbeit in kleinen Gruppen (klein genug, dass jedes Mitglied an der Kommunikation teilhaben kann, wodurch eine maximale Obergrenze von fünf bis sechs Lernenden gegeben ist) sowie eine
- gemeinsam zu lösende Aufgabe bzw. zu erreichende Zielsetzung.

Darüber hinaus weist Cohen (1993) ergänzend auf die Notwendigkeit einer selbstständigen Aufgabenbearbeitung hin, wodurch als dritter Aspekt

- eine direkte und unmittelbare Supervision durch den Lehrer ausgeschlossen werden sollte.

Spezifiziert man die an kooperatives Lernen gestellten Erwartungshaltungen, so ist festzustellen, dass hinsichtlich der Entwicklung kognitiver Kompetenzen weniger der Erwerb von Faktenwissen denn mehr das *Verstehen von Zusammenhängen* sowie die Fähigkeit, *komplexe Probleme zu lösen* begünstigt werden sollen. Neben dem Aspekt der Leistungsentwicklung wird mit Blick auf die Entwicklung prosozialer Lernziele insbesondere ein Zuwachs an *kommunikativen und kooperativen Kompetenzen* als Begründung für kooperatives Arbeiten angeführt (Gräsel & Gruber 2000), wonach sich verdichtet die zentrale Zielsetzung kooperativen Arbeitens als kommunikativ-kooperatives Lösen komplexer Probleme zusammenfassen lässt.

Neben der erwünschten Begriffsklärung verdeutlichen die oben angeführten Definitionen kooperativen Lernens von Slavin bzw. Johnson und Johnson mit dem eher beiläufigen Verweis auf das Lernen in „kleinen Gruppen“ (small groups) als Ort koopera-

tiven Arbeitens aber auch eine Problematik, die auf die unzulässige Gleichsetzung von kooperativem Arbeiten und „Gruppenarbeit“ gründet: Gruppenarbeit stellt neben Partner-, Einzelarbeit und Klassen- bzw. Frontalunterricht eine *Sozial*form des Unterrichts dar und regelt als solche die Beziehungsstruktur des Unterrichts (Meyer 1987). Kooperatives Arbeiten ist hingegen eine Unterrichts*methode*, die, in Abhängigkeit von der jeweiligen Form des kooperativen Arbeitens, eine spezifische Unterrichtsstruktur bedingt. Somit sind die meisten Formen kooperativen Arbeitens an die Sozialform der Gruppenarbeit gebunden (allerdings nicht zwingend, da auch Partnerarbeit Möglichkeiten des kooperativen Arbeitens bietet), während die Ausweisung einer Gruppenarbeitsphase im Planungsraster einer Unterrichtsstunde noch keine Auskunft über die gewählte Form des kooperativen Arbeitens gibt. So beinhalten auch lehrerzentrierte, direktive Unterrichtsformen häufig Gruppenarbeitssequenzen, wie Cohen (1994a) darlegt: “(…) cooperative learning should not be confused with small groups that teachers often compose for the purpose of intense, direct instruction - for example, reading groups.” Derartige Arbeitsphasen erheben nur geringe kooperativ-kommunikative Anforderungen an die Lernenden und sind aus methodischer Sicht als die kaschierte Form eines ergebnisorientierten Unterrichtsgesprächs zu bewerten. Erschwert beispielsweise die unterschiedliche Lesekompetenz der Lernenden die Bearbeitung eines Textes im Klassenplenum, so kann der Arbeitsauftrag, im Sinne einer inneren Differenzierung, in leistungshomogenen Kleingruppen mit unterschiedlichem Anforderungsbereich formuliert werden.

Auf der anderen Seite terminiert eine konkrete kooperative Unterrichtsmethode maßgeblich die Artikulation und damit auch die Strukturierung einer Stunde. Methodische Überlegungen zeichnen sich daher im Vergleich zu Entscheidungen zur Sozialform durch eine größere Komplexität aus und stehen in einer unmittelbaren Interdependenz zu den Zielen, Inhalten, Aktionsformen und Medien des Unterrichts (Gudjons 1993). Abb. 3.1 veranschaulicht die zahlreichen Aspekte der Unterrichtsmethodik:

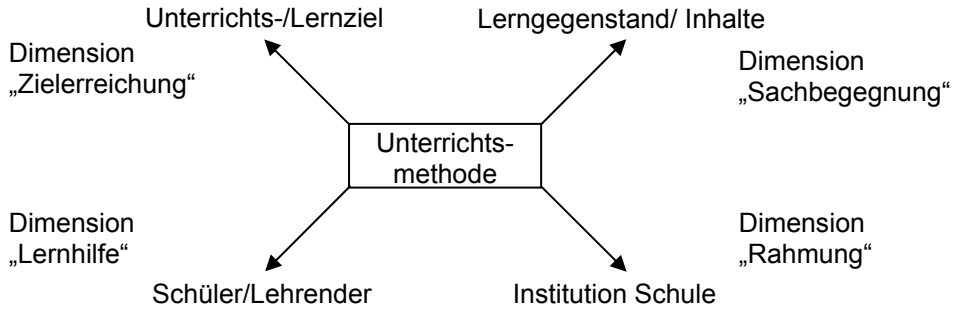


Abb. 3.1: Dimensionen der Unterrichtsmethode (Terhart, 2000)

Terhart ordnet der Lehr-Lernmethode hiernach vier zentrale Dimensionen zu, die eine unterschiedliche Definition von Methode bedingen:

- *Zielerreichung*: Methode wird definiert als ein Mittel zur Erreichung von Unterrichtszielen. Die Reduktion des Methodenbegriffs auf eine technische Instrumentalisie-

rung bedingt eine plausible Begründung und Legitimation der übergeordneten Lernziele.

- *Sachbegegnung*: Methode wird gleichsam definiert als Mediator, zwischen lernendem Subjekt und dem zu erlernenden Objekt, d.h. dem Lerngegenstand.
- *Lernhilfe*: Diese Dimension fokussiert auf das lernpsychologisch nutzbare Potenzial der Unterrichtsmethode und bezeichnet den gezielten Einsatz von Methode zur Schaffung von lernförderlichen Umgebungen.
- *Rahmung*: Durch die institutionellen Rahmenbedingungen unterrichtlichen Lernens ist ein Handlungsspielraum des Lehrenden in Hinblick auf seine Methodenentscheidungen zwar vorhanden, wohl aber begrenzt.

Die ausgeführten Definitionen verdeutlichen die Komplexität methodischer Entscheidungen und unterstreichen die Unzulässigkeit der Gleichsetzung von Sozialform und Methode. Wenn im Folgenden aus Gründen der Lesbarkeit ohne eine weitere Spezifizierung pauschal von „Gruppenunterricht“ gesprochen wird, so bezieht sich dies methodisch immer auf ein kooperatives Lehr-Lernarrangement, so wie sich auch Ausführungen über kooperatives Arbeiten immer auf die Sozialform der Gruppenarbeit erstrecken sollen, sofern es nicht explizit anders ausgewiesen wird.

3.1 Anspruch und Wirklichkeit kooperativer Lernformen

Neben sprachlich-definitorischen Unklarheiten ist ein weiteres, dem Gruppenunterricht eigenes Phänomen augenfällig: Wie keine andere Sozialform wird Gruppenunterricht von allen Vertretern des Bildungsbetriebs in wissenschaftlichen Publikationen wertgeschätzt und empfohlen. So schätzt Meyer (1987) die Zahl der Monographien zum Thema Gruppenunterricht auf über einhundert, wohingegen die entsprechende Zahl zum Frontalunterricht unverhältnismäßig gering ausfällt. Demgegenüber steht der verschwindend geringe Anteil an effektiver Unterrichtszeit, in der Gruppenarbeit praktiziert wird. Hage et al. (1985) weisen mit einer Studie für den deutschsprachigen Raum nach, dass je nach Schulstufe lediglich zwischen zwei und elf Prozent der Unterrichtszeit auf Gruppenarbeit entfällt, wohingegen der so genannte „Klassenunterricht“ (im Sinne eines lehrerzentrierten Frontalunterrichts) mit über siebenundsiebzig Prozent den Unterricht dominiert. Zu vergleichbaren Zahlen kommen Rotering-Steinberg und von Kügelen (1986), die auf der Grundlage einer Lehrerbefragung feststellen, dass im Mittel nur sieben Prozent der Lehrkräfte Gruppenunterricht regelmäßig durchführen. Wie ist diese offenkundige Diskrepanz zwischen pädagogischer Wertschätzung und realisierter Umsetzung in der Alltagspraxis zu erklären? Um dieser Frage nachzugehen sollen im Folgenden theoretische Annahmen und empirische Befunde zum kooperativen Lernen sowie mit kooperativen Lernformen verbundene Schwierigkeiten betrachtet werden.

3.1.1 Theoretische Annahmen – Warum ist kooperatives Arbeiten erfolgreich?

Die theoriegeleitete Entwicklung kooperativer Lernformen wird maßgeblich geprägt durch die Arbeit von Deutsch (1949), der als zentrales Merkmal einer Gruppe die *Interdependenz* der Gruppenmitglieder postuliert, welche selbst wiederum erst durch eine *gemeinsame Zielsetzung* hergestellt wird. Hinsichtlich der Interdependenz unterscheidet Deutsch positive und negative Interdependenz, wodurch unterschiedliche Qualitäten in der Zielstruktur festgelegt werden: Positive Interdependenz liegt vor, wenn der individuelle Erfolg dazu beiträgt, dass auch andere Gruppenmitglieder das Aufgabenziel erreichen. In diesem Fall spricht man von einer *kooperativen Zielstruktur.* Korrelieren hingegen im Falle einer negativen Interdependenz die individuellen Wahrscheinlichkeiten ein Ziel zu erreichen negativ miteinander, so konkurrieren die Gruppenmitglieder untereinander, es liegt eine *kompetitive Zielstruktur* vor. Ist schließlich der Erfolg der individuellen Zielerreichung ohne Konsequenzen für die Zielerreichung der übrigen Personen, so ist keine Interdependenz vorhanden und man spricht von einer *individualistischen* Zielstruktur.

Fußend auf den dargestellten Grundannahmen zur kooperativen Zielstruktur beschreibt Slavin (1995, 1997) unter verschiedenen Perspektiven positive Auswirkungen kooperativer Lernmethoden. Er unterscheidet hierbei motivationale und kognitive Perspektiven.

3.1.1.1 Motivationale Perspektiven

Motivationale Perspektive im engeren Sinne

Die motivationale Perspektive setzt eine externe Anreizstruktur in Form einer interpersonellen Belohnung (Lob, Noten, Würdigung etc.) auf der Basis der gesamten Gruppenleistung voraus. Ist dieser Anreiz gegeben, wächst sowohl die Bereitschaft der Gruppenmitglieder einander zu helfen, als auch die Erwartungshaltung an jedes Gruppenmitglied, sich maximal anzustrengen. Dieser motivationstheoretische Ansatz resultiert als Umkehrschluss aus der Erfahrung, wonach unter kompetitiven Bedingungen individuelle Leistungsbereitschaft und Erfolg als sozial unerwünscht angesehen wird, da hierdurch die Norm für alle Schüler erhöht wird. (Der Leistungserfolg des Einzelnen („Strebers“) bedingt somit den Misserfolg des anderen Schülers.) Arbeiten Schüler hingegen kooperativ auf ein Ziel hin, so motiviert es sie, ihre Anstrengungsbereitschaft zu verstärken und leistungsförderliche Normen zu generieren.

Perspektive der sozialen Kohäsion

Die Perspektive der sozialen Kohäsion stellt eine Variante der motivationalen Perspektive dar und rekurriert ebenfalls auf motivationstheoretischen Annahmen, beinhaltet aber eine stärker sozialpsychologisch geprägte Komponente. Damit tritt die im Rahmen der motivationalen Perspektive geforderte Schaffung von Gruppenanreizen (Noten) und individueller Verantwortlichkeit in den Hintergrund, während als entscheidende Triebkraft für die konstruktive Kooperationsbereitschaft der Gruppenmitglieder die persönliche Wertschätzung der Gruppe an sich, respektive ihrer Funkti-

onsfähigkeit gesehen wird (Cohen 1994b, Sharan & Sharan 1992). Vorausgesetzt die Aufgabenstellung ist hinreichend klar und interessant formuliert, so Cohen (1994b), und die Gruppenmitglieder sind ausreichend auf ihre Aufgabe vorbereitet, „students will experience the process of groupwork itself as highly rewarding“. Diese Position kongruiert mit der Selbstbestimmungstheorie der Motivation (Deci & Ryan 1993), wonach das Bedürfnis nach sozialer Einbindung (neben Autonomie- und Kompetenzerleben) eine motivationsfördernde Variable darstellt.

Die Bandbreite aller Theorien zum kooperativen Lernen verdeutlicht jedoch, dass sich motivationale und sozial-kohäsive Perspektiven keineswegs disjunkt verhalten. Während einige Vertreter eher eine puristische Position einnehmen (z.B. Cohen, Sharan & Sharan zugunsten der sozial-kohäsiven Theorie, Slavin zugunsten der motivationalen Perspektive), repräsentiert die Lernform des Learning-Together (s. Kp. 3.2) von Johnson & Johnson (1989) einen Ansatz, der beide Positionen vereint.

3.1.1.2 Kognitive Perspektiven

Als Begründung für die Steigerung der Schulleistung durch kooperative Lernformen werden neben motivationstheoretischen Argumenten auch kognitive Argumente angeführt. Letztere unterscheiden sich von den bislang dargestellten Perspektiven dadurch, dass sie als zentralen Wirkmechanismus auf die mentale Verarbeitung von Informationen fokussieren. Zu den bedeutendsten kognitiven Theorien zählt Slavin (1993, 1996) die Entwicklungsperspektive und die Perspektive der kognitiven Elaboration.

Entwicklungsperspektive

Unter dem Begriff „Entwicklungsperspektive“ lässt sich eine Gruppe kognitiver Theorien subsumieren, die sich auf den theoretischen Annahmen des sozio-genetischen Konstruktivismus von Piaget (1926) sowie, als alternative Erklärung zum Piagetschen Ansatz, auf die Arbeit von Wygotski (1978) zurückführen lassen.

Piaget gründet seine Theorie auf der zentralen Annahme, wonach sozial-arbiträres Wissen in Form von Sprache, Werten, Symbolsystemen nur in der Interaktion mit anderen gelernt werden kann (Slavin 1993). Die für Piaget essentiellen *kognitiven Konflikte* gewinnen ihre Authentizität und damit ihr kognitives Entwicklungspotential nicht aus der Diskussion mit Erwachsenen (Eltern, Lehrer), sondern mit gleichaltrigen Interaktionspartnern, da Heranwachsende vor allem mit egalitären Diskussionspartnern sozial induzierte kognitive Konflikte bearbeiten können: „Piaget […] regarded social exchanges between children and adults as unlikely to lead to the kinds of cognitive development that exchanges with child peers promote. Because children and adults do not cooperate as equals and do not exercise mutual control over the interaction, the child cannot really share the adult’s point of view.” (Webb & Palinscar 1996: 844). Kritisiert wird am kognitiv-entwicklungspsychologischen Ansatz allerdings die ungeklärte Frage, wie Lernende unterstützt werden können, unter unterrichtlichen Bedingungen kognitive Konflikte zu generieren und konstruktiv zu lösen. Allzu oft, so O’Donnell & O’Kelly (1994), wirkt ein ungleicher sozialer Status sowie eine unausgewogene Beteiligung der Gruppenmitglieder an der Diskussion dem entgegen, wes-

halb externe, strukturierende Hilfen eingefordert werden. Huber (1999) verweist in diesem Zusammenhang auf die Lernmethode der „strukturierten akademischen Kontroverse“ von Johnson und Johnson (1994b) die dieses Defizit durch einen strukturierten und phasierten Argumentationsverlauf zu beheben suchen.

Wie Piaget, so betont auch Wygotski (1978) die Bedeutung der sozialen Interaktion für die kognitive Entwicklung, legt aber darüber hinaus eine differenzierte Anforderung an den Interaktionskontext vor. Interaktionen, so Wygotski, sollten sich hinsichtlich des kognitiven Anforderungsniveaus in der Zone der *proximalen Entwicklung* ereignen. Nach Slavin (1993) versteht man hierunter die *Differenz zwischen* dem *aktuellen Entwicklungsniveau*, definiert als Fähigkeit, selbstständig Probleme zu lösen, und dem *potentiellen Entwicklungsniveau*, welches durch Hilfe eines *kompetenteren* Kooperationspartners erreicht werden kann. Wygotski betont, dass diese Hilfe sowohl durch Erwachsene als auch durch Gleichaltrige („In cooperation with more capable peers.“) erbracht werden kann. Auf der Basis dieser Theorie wurden in den achtziger und neunziger Jahren verschiedene methodische Techniken entwickelt, die helfen sollen, die Kooperation in der notwendigen Art zu organisieren (z. B. „reciprocal teaching“ (Palinscar & Brown 1984), „scaffolding“ (O’Donnell & O’Kelly 1994)).

Kognitiv-elaborative Perspektive

Die Perspektive der kognitiven Elaboration betrachtet den Lernprozess als Aufbau einer vernetzten Wissensstruktur; neue Informationen müssen durch Umstrukturierung oder Elaboration kognitiver Schemata mit bereits vorhandenen Informationen verknüpft werden (Wittrock 1986, Renkl 1997). Eine besonders effektive Form der Elaboration besteht nach Slavin (1997) darin, den Lerninhalt einer anderen Person zu erklären. Untersuchungen so genannter tutorieller Lehr-Lernarrangements belegen gleichermaßen den Nutzen für den Tutor wie auch für den Schüler. Neben der bereits erwähnten Methode des „reciprocal teaching“ (Palinscar & Brown 1984) fußt die Methode der „kooperativen Lernskripte“ (Dansereau 1988) auf der Perspektive der kognitiven Elaboration. Huber (1999) merkt kritisch an, dass die kognitiv-elaborative Perspektive die spezifischen Bedingungen der Kleingruppenarbeit nicht reflektiert und stattdessen, losgelöst von der Sozialform, im Sinne einer allgemeinen Lerntheorie die Bedeutung der aktiven Auseinandersetzung des Einzelnen mit dem Lernstoff betont.

Ansatz der situierten Kognition – eine neue Perspektive

Schulisch vermitteltes Wissen, so Gruber, Mandl und Renkl (2000), stellt all zu oft „träges Wissen“ dar, das zwar einer Überprüfung in schulischen Testsituationen standhält, auf das aber in komplexen, außerschulischen Anwendungssituationen nicht zurückgegriffen wird. Renkl (1996, 1998) identifiziert drei Erklärungsansätze für die Entstehung trägen Wissens: Neben *metakognitiven Defiziten* seitens der Lernenden sowie *strukturellen Defiziten,* die im Wissen selbst angesiedelt sind und eine transferresistente Wissenskompartimentalisierung bedingen, führt Renkl *Situiertheitserklärungen* als dritte Kategorie an. Dieser Erklärungsansatz betrachtet eine fehlende Transferfähigkeit nicht als Defizit einer mangelnden propositionalen Struktur,

sondern als „Normalfall“, da Wissen prinzipiell an einen situativen Kontext gebunden ist (Renkl 1998). Dieser Aspekt wird von Vertretern der situierten Kognition auf die Methodendominanz des Frontalunterrichts zurückgeführt, da ein derartiges Unterrichtsarrangement durch eine partikuläre Thematisierung inhaltlicher Aspekte sowie einer isolierten Erarbeitung durch die Schüler gekennzeichnet ist und damit eo ipso einen späteren Anwendungstransfer des Gelernten erschwert. Folgt man stattdessen der Annahme, dass Lernsituationen späteren Anwendungssituationen möglichst ähnlich sein sollen, so resultiert hieraus eine Begünstigung kooperativer Lernformen, da diese sowohl mit Hinblick auf die Möglichkeit der Aufgabengestaltung als auch bedingt durch das Lernen in der sozialen Gemeinschaft eine hohe Alltagsauthentizität besitzen (Gräsel, Gruber 2000).

Kritisch anzumerken ist allerdings, dass der Ansatz des situierten Lernens nicht unumstritten ist, da ihm, wie häufig moniert, der Innovationscharakter fehle. Weinert (1998) betrachtet situiertes Lernen als Teilmenge aller bekannten kognitiven Lernprozesse und damit als einen „handfesten Eklektizismus“. Ähnlich äußert sich Klauer (1999), der die Frage aufwirft, was von der Theorie des situierten Lernens bleibe, „wenn man die Anregungen von Dewey, Piaget und Wygotski herausinterpretiere“.

3.1.2 Effekte kooperativen Arbeitens

Sucht man nach empirischen Ergebnissen, die Auskunft über die Effektivität kooperativen Arbeitens liefern, so ergibt sich ein uneinheitliches Bild. Zwar belegen die Arbeiten zur Effektivität von Gruppenarbeit (Slavin 1997, Renkl 1995, Johnson, Maruyama et al. 1981) in ihrer Gesamtheit die Vorteile kooperativer Lehr-Lern-Arrangements gegenüber individualisierten oder kompetitiven Lernsituationen – auch in Hinblick auf die Entwicklung von Problemlösefähigkeiten und den Erwerb naturwissenschaftlicher Kenntnisse. Allerdings gibt es zwischen den Studien auch substantielle Unterschiede (s. Kp. 3.1.2.1), deren Ursachen nicht auf einfache Prozess-Produkt-Kausalitäten zurückzuführen sind. Eine verhältnismäßig hohe Konsistenz in den Aussagen weisen die – überwiegend – angloamerikanischen Metastudien der achtziger und neunziger Jahre auf (Johnson, Maruyama et al. 1981, Qin, Johnson & Johnson 1995, Lazarowitz & Hertz-Lazarowitz 1998, Springer, Stanne & Donovan 1999), die sich jedoch vornehmlich auf Schulleistungsdaten erstrecken.

Verglichen mit den empirischen Ergebnissen zur Unterrichtsqualität bei direkter Instruktion ist die Datenlage zum kooperativen Arbeiten nur als unbefriedigend zu bezeichnen. Aussagen, die inhaltliche *und* motivationale Aspekte im Sinne einer multikriterialen Zielerreichung inte-grieren, wie sie für einen qualitativ hochwertigen, direktiven Unterricht bereits dokumentiert sind (Gruehn 1995, Baumert & Köller 2000), liegen für den kooperativen Unterricht bislang nicht vor. Diesem Defizit liegt eine genereller Mangel an systematischer Erkenntnis zur Erstellung von Kriterien erfolgreichen kooperativen Arbeitens zu Grunde, da sich die Unterrichtsqualitätsforschung bislang ausschließlich auf direktive Unterrichtsformen beschränkt, wie Einsiedler (1997) ausführt (vgl. auch Einsiedler 2000 sowie Pauli, Reusser et al. 2003): „Forschung ist ganz stark auf die Qualität des lehrergesteuerten Klassenunterrichts zent-

riert. Wir wissen wenig über die Qualität von Einzelarbeit und Gruppenarbeit, obwohl hier erhebliche Unterschiede hinsichtlich kognitiver Niveaus und aktiver Beteiligung bestehen dürften."

Eine umfassende Beurteilung des Erfolgs kooperativer Lernarrangements wird überdies durch die Bandbreite der Erwartungshaltung, die an diese Lernformen geknüpft wird, erschwert. Neben den aus dem direktiven Unterricht bekannten Variablen der Schulleistung und der Motivation sind beim kooperativen Lernen insbesondere die Vermittlung von Schlüsselqualifikationen bzw. die Erreichung sozialer Lernziele von Interesse, da diese Lernform hierfür aus theoretischer Sicht besonders geeignet erscheint (Huber 1991). Wenngleich Didi et al. (1993) an Hand ihrer Studie den abstrakten Begriff der „Schlüsselqualifikationen" dahingehend konkretisieren können, dass hierunter primär „Kooperationsfähigkeit" und „Kommunikationsfähigkeit" verstanden wird, so weisen sie gleichzeitig auch nach, dass diese beiden Konstrukte aus psychologischer Sicht nur äußerst schwer zu operationalisieren sind. Somit stehen kooperative Lernformen bislang partiell immer auch vor dem Dilemma, in Ermangelung geeigneter Testinstrumente die selbst gestellten Ansprüche nicht belegen zu können.

Hinweise auf die Problematik der Operationalisierung von Schlüsselqualifikationen geben auch die Ergebnisse der ersten PISA-Studie (Baumert et al. 2001), in deren Rahmen zur Ermittlung so genannter „cross curricularer Kompetenzen" auch die Kooperations- und Kommunikationsfähigkeit deutscher Schüler erhoben wurde. Die Befunde deuten darauf hin, dass Kooperations- und Kommunikationsfähigkeit überwiegend durch familiäre Einflussfaktoren terminiert werden bzw. auf Geschlechtsunterschiede zurückzuführen sind. Im Gegensatz zu den erhobenen Leistungsdaten lassen sich nur in einzelnen kooperations- und kommunikationsrelevanten Aspekten schulformabhängige Differenzen nachweisen. So konnte einerseits gezeigt werden, dass an Gymnasien eine im Vergleich zu allen andern Schulformen geringere Verantwortungsabwehr und eine signifikant höhere Normeneinhaltung unter Gleichaltrigen zu verzeichnen ist. Andererseits zeigten die für die Ermittlung der Kooperations- und Kommunikationsfähigkeit bedeutsamen Aspekte „Perspektivenübernahme", „soziale Selbstwirksamkeitsüberzeugungen" und „Emphatie" nur für den letztgenannten Aspekt einen signifikanten Unterschied zugunsten Gymnasialschüler gegenüber Schülern aller andern Schulformen. Der Anteil der auf die Schulform zurückzuführenden Varianz dieser drei Aspekte liegt bei nur 0,9% (Empathie), 0,2% (Soz. Selbstwirksamkeitsüberzeugung) und für die Fähigkeit der Perspektivenübernahmen sogar bei 0,0%! (Stanat & Kunter 2001). Im Vergleich dazu beträgt der Varianzanteil der Schulform an der Lesekompetenz 46,0%.

Wenn im Folgenden an Hand einiger ausgewählter Studien einseitig auf den Einfluss kooperativer Lernformen auf die Schulleistung fokussiert wird, so ist dies repräsentativ für die Datenlage der empirischen Forschung und unterstreicht die dargestellten Fehlstellen in der Forschung zum kooperativen Arbeiten.

3.1.2.1 Empirische Befunde

Die Aussagekraft quantitativer Studien zum Einfluss kooperativer Lernformen auf die Schulleistung muss sich immer auch an der Qualität der zu Grunde liegenden Untersuchungsmethode und des Untersuchungsdesigns messen lassen (Vgl. Kp. 4). Diesbezüglich weisen fachdidaktische Studien häufig noch einen suboptimalen Standard auf. Wenn beispielsweise Sisovic und Bojovic (2000) einen signifikanten Vorteil kooperativer Arbeitsformen für die Entwicklung des konzeptuellen Wissens zum Themenbereich „Säure-Base“ von Neuntklässern konstatieren, so wird dieses Ergebnis durch das Untersuchungsdesign relativiert, welches in der Kontrollgruppe und in der Interventionsgruppe keine Konstanthaltung der Lehrperson, sondern den Einsatz von zwei verschiedenen Lehrern in beiden Gruppen vorsieht. Überdies wurde in dieser Studie die intervenierte Gruppe von einem der beiden Wissenschaftler selbst unterrichtet, so dass neben interpersonellen Effekten zwischen beiden Lehrern auch Versuchsleitereffekte nicht auszuschließen sind.

Es überrascht daher nicht, dass Metaanalysen zum kooperativen Arbeiten häufig ein eher uneinheitliches Bild zeichnen. In einer der wenigen Metaanalysen zum Einfluss kooperativer Arbeitsformen auf die Lernleistung speziell im Chemieunterricht kommt Bowen (2000) zu folgendem Ergebnis (Abb. 3.2): Wenngleich aus der Analyse hervorgeht, dass 85% der untersuchten Studien einen positiven Effekt auf die Lernleistung im Chemieunterricht haben, so zeigt die Graphik der Verteilungshäufigkeit eine deutliche Abweichung von der Normalverteilung, was eine Inkohärenz der in die Untersuchung aufgenommenen Studien nahe legt.

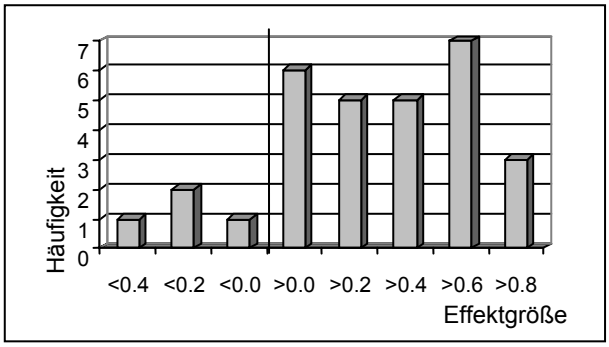


Abb. 3.2.: Häufigkeitsverteilung von 30 Effektgrößen zum Einfluss kooperativer Lernformen auf die Leistung (nach Bowen 2000)

Die Grafik verdeutlicht ferner, dass vereinzelt auch Ergebnisse vorliegen, die zu einer negativen Effektgröße führen und diese Interventionen damit schlechtere Leistungsergebnisse als die Kontrollgruppe aufweisen.

Der methodische Aufwand, der zur Erstellung von Metaanalysen betrieben werden muss, ist im Allgemeinen ernorm: Johnson und Johnson (1989) legen eine Metaanalyse zur Effektivität kooperativen Arbeitens vor, in die über 500 Studien eingehen. Für jede einzelne Studie wird dabei ein Index (5-16) berechnet, der unter Berücksichtigung von Bedingungskontrolle, Randomisierung der Probanden, Treatmentcheck, Konstanthaltung oder Rotation der Lehrenden und Konstanthaltung des Curriculums die methodologische Qualität der Studie angibt. Dadurch lässt sich ein relatives Maß für die Durchführungsobjektivität und damit für die Gesamtgüte der einzelnen Studie berechnen.

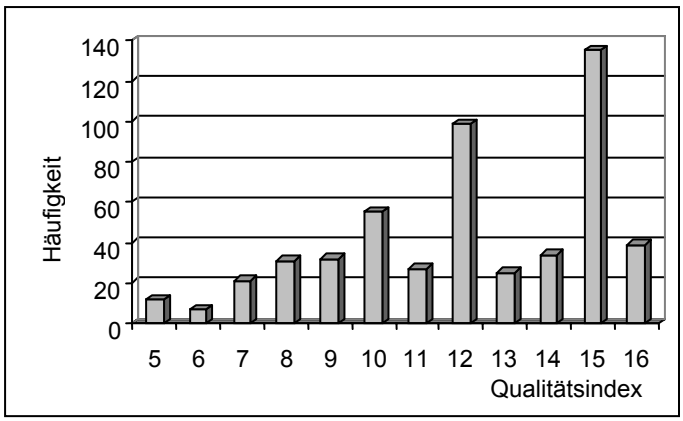


Abb. 3.3: Häufigkeitsverteilung der methodischen Qualität von Studien zum kooperativen Arbeiten

Abbildung 3.3 ist zu entnehmen, dass es eine Vielzahl an Studien gibt, die eine gute oder sehr gute methodische Struktur besitzen. Andererseits weisen etwa ein Drittel einen Qualitätsindex kleiner 10 auf, was methodisch auf nicht-standardisierte Untersuchungsbedingungen schließen lässt. Die Berücksichtigung dieser Studien oder deren Selektion hat damit einen großen Einfluss auf die Gesamtgüte einer Metaanalyse. Ferner weist Renkl (1997) nach, dass auch Analysen, die die Effektivität konkreter kooperativer Lernmethoden erheben, untereinander stark variieren. So wird beispielsweise sowohl für die Methode des Gruppenpuzzles als auch für Studien zur Effektivität der Gruppenrecherche gleichermaßen von ausbleibenden, stark positiven als auch substantiell negativen Effektgrößen berichtet. Selbst die vermeintlich hohe Reliabilität von Metaanalysen ist in Frage zustellen, so man beispielsweise die berechnete mittlere Effektgröße zum Gruppenpuzzle von Johnson und Johnson (1989), die mit einem Wert von 0,46 einen bedeutsamen Effekt nachweisen, mit dem aus der Metaanalyse von Slavin (1989) resultierenden Wert, der für die gleiche Methode nur eine moderate Effektgröße von 0,04 angibt, miteinander vergleicht.

Renkl (1997) führt dies auf die unterschiedliche Auswahl der berücksichtigten Studien (s.o.) zurück, weshalb eine Berechnung der „mittleren“ Effektivität ohnehin nur eine eingeschränkte Aussagekraft hat. Vielmehr, so Renkl, belegen die divergierenden Befunde die Existenz von Bedingungen, die die Effektivität der kooperativen Methoden moderieren. Aufgrund der bereits dargelegten lückenhaften Kenntnis über die qualitätsbestimmenden Prädiktoren von Gruppenarbeit liegen hierzu jedoch bislang keine gesicherten Erkenntnisse vor.

Es erscheint daher aus vielerlei Gründen angebracht, den Erfolg von kooperativen Lernformen ordinalskaliert anzugeben, so wie es in der Analyse von Johnson, Muryama et al. (1981) nach der „Auszählmethode“ erfolgt.

	Methode					
	Auszählen			Effektgröße		
Bedingung	N	KD	P	*M*	*SD*	*N*
kooperativ vs. gruppenkompetitiv	3	6	4	.00	.63	9
kooperativ vs. kompetitiv	**8**	**36**	**65**	**.78**	**.99**	**70**
gruppenkompetitiv vs. kompetitiv	3	22	19	.37	.78	16
kooperativ vs. individualisiert	6	42	108	.78	.91	104
gruppenkompetitiv vs. individualisiert	1	10	20	.50	.37	20
kompetitiv vs. individualisiert	12	38	9	.03	1.02	48

N = negativ, KD = keine Differenz, P = positiv

Tab. 3.1: Effekte kooperativer, kompetitiver und individualisierter Unterrichtsbedingungen auf die Lernleistung (nach Johnson, Maruyama et al. 1981)

Auf Grund der einzelnen Analysen wird lediglich erfasst, in wie weit in den kontrastierten Bedingungen ein negativer (N), positiver (P) oder kein differentieller Effekt (KD) zu verzeichnen ist.

Betrachtet man die Ergebnisse, so lässt sich zusammenfassend sagen, dass es - auch für den in dieser Studie durchgeführten Vergleich von kooperativen und kompetitiven Lernbedingungen - allgemein einen positiven Einfluss von kooperativen Lernformen auf die Lernleistung gibt, unabhängig von dessen Ausmaß.

3.1.2.2 Schwierigkeiten in kooperativen Lernarrangements

Jenseits der durch quantitative Leistungsdaten dokumentierten Erkenntnisse über den Erfolg kooperativer Lernarrangements liegen auf Basis qualitativer Einzelfallstudien Erfahrungsberichte vor, die dokumentieren, aus welchen Gründen Lernende kooperativem Arbeiten gegenüber häufig skeptisch eingestellt sind. Die im Folgenden dargestellten Phänomene veranschaulichen pointiert die Facetten potenzieller Schwierigkeiten, wie sie bei kooperativer Kleingruppenarbeit zu beobachten sind. Sofern nicht anders angegeben, folgen die Beispiele den Ausführungen von Renkl, Gruber, Mandl (1996):

- Das < Der-Hans-der-machts-dann-eh' >-Phänomen

Dieses Phänomen wird in der Literatur auch als „free rider effect“ (Slavin 1995) beschrieben und beschreibt eine Konfliktsituation während des kooperativen Arbeitens, in der einige Gruppenmitglieder sich zu Lasten anderer aus der Gruppenarbeit zurückziehen. Slavin (1983) führt dieses Phänomen auf eine „Diffusion der Verantwortlichkeit“ zurück, die durch eine Aufgabenspezialisierung der Gruppenmitglieder reduziert werden kann.

- Das < Ja-bin-ich-denn-der-Depp >-Phänomen

Als unmittelbare Folge aus dem „free-rider effect“ kann es zu Verärgerung, Motivationsverlust und damit auch zu einer Verringerung der Anstrengungsbereitschaft der Personen kommen, die maßgeblich den Gruppenarbeitsprozess gestalten. Dies Phänomen wird auch als „sucker-effect“ (Kerr 1983) bezeichnet.

- Das < Da-mach-ich's-doch-gleich-lieber-selbst >-Phänomen

Schüler mit höheren motivationalen und kognitiven Eingangsvoraussetzungen ziehen häufig zentrale Aufgaben an sich, da andere Gruppenmitglieder nicht in der Lage scheinen, gemäß ihren Ansprüchen zu arbeiten. Dieses Phänomen führt häufig zum unerwünschten „Matthäus-Effekt“ („the rich-get-richer effect“, Johnson, Johnson, Smith 1995), woraus eine Zunahme der Leistungsvarianz resultieren kann; gute Schüler werden immer besser, schlechte Schüler werden immer schlechter (daher auch als „Schereneffekt“ bezeichnet).

- Das < Kann-und-mag-ich-nicht-mach-du >-Phänomen

Das Phänomen stellt eine Variante des zuvor genannten Phänomens dar und wird auch als intrapersonaler Matthäus-Effekt bezeichnet: Die Aufgaben in einer Gruppenarbeit werden von den Mitgliedern so verteilt, dass jeder die Anforderung übernimmt, die er am besten zu erfüllen vermag. Dieses scheinbar vernünftige Verhalten führt allerdings bei dauerhafter Reproduktion zu einer immer weiteren Verbesserung individueller Stärken, während individuelle Defizite keine weitere Entwicklung erfahren.

- Das < Ich-hab-meinen-Teil-erledigt >-Phänomen

Dieses Phänomen tritt häufig in Folge des zuvor genannten Phänomens auf: Im Rahmen einer Kleingruppenarbeit mit Aufgabenteilung wird kein Interesse an der Qualität der Arbeitsergebnisse der anderen Gruppenmitglieder gezeigt und die Aufmerksamkeit allein auf den eigenen Beitrag gerichtet. Potenzielle Fehler anderer Gruppenmitglieder werden somit häufig wider besseren Wissens ignoriert. Dieses auch als *„ganging up effect“* bezeichnete Phänomen (Neber 1998) kann soweit führen, dass die gesamte Gruppe die Lösungswege präferiert, die mit der geringsten Anstrengung verbunden sind.

- Das < Gruppenarbeit-nein-danke >-Phänomen

Mit diesem Verhalten wird ein Phänomen charakterisiert, das auf einer Ansammlung verschiedener, auf die dargestellten Phänomene zurückzuführender Erfahrungen aus vergangen Gruppenarbeitsprozessen zurückzuführen ist. Aus den oben angeführten Gründen sind die Erfahrungen mit Gruppenarbeit häufig so negativ besetzt, dass es zu dieser Einstellung kommt.

3.2 Formen kooperativen Arbeitens

Das seit dem Beginn der siebziger Jahre andauernde Interesse gegenüber kooperativen Lernformen führt bis in die Gegenwart zu einer fortschreitenden Methodendis-

kussion, verbunden mit einer umfangreichen Variation vorhandener sowie der Entwicklung neuer Unterrichtsformen (bspw. Wiechmann 2000, Körner & Breuer 2004), so dass ein vollständiger Überblick über die Gesamtheit aller kooperativen Lernformen im Rahmen dieser Arbeit nicht gegeben werden kann. Es erfolgt daher eine Schwerpunktsetzung zugunsten der gängigsten kooperativen Lernformen.

Um die Vielzahl der kooperativen Lernformen systematisch darzustellen, wird häufig ein Kategorisierungsschema vorgeschlagen, welches zwischen den Dimensionen der Belohnungsstruktur und der Aufgabenstruktur differenziert (Gage, Berliner 1996, Bossert 1988).

Aufgabenstruktur	Belohnungsstruktur: (i) Keine Gruppenbelohnung bzw. Belohnung nicht aufgrund individueller Leistung	Belohnungsstruktur: (ii) Gruppenbelohnung aufgrund individueller Leistung
(iv) Keine Aufgabenspezifizierung	1. Learning-Together	2. Schülerteams und Leistungsdifferenzierung; Team-Spiele-Turniere
(iii) Aufgabenspezifizierung	3. Jigsaw I; *Group-Investigation*	4. Jigsaw II

Tab. 3.2: Typologie kooperativer Lernschemata (nach Bossert, 1988)

Die Differenzierung hinsichtlich der Aufgaben- bzw. Belohnungsstruktur stellt lediglich ein pragmatisches, typologisierendes Oberflächenmerkmal der verschiedenen Methoden dar, denen darüber hinaus unterschiedliche lern- und motivationstheoretische Annahmen zu Grunde liegen (s. Kp. 3.1.1). Die oben aufgeführten Methoden werden im Folgenden näher dargestellt.

3.2.1 Learning Together

Kooperatives Lernen nach der Methode des „Learning-Together" wurde in den siebziger Jahren als einer der ersten neo-reformpädagogischen Ansätze von Johnson und Johnson (1975) entwickelt. Der Learning-Together-Ansatz, in dem leistungsheterogene Gruppen von vier bis fünf Schülern eine vom Lehrer gestellte Aufgabe zur gemeinsamen Bearbeitung erhalten, stellt eine vergleichsweise gering vorstrukturierte Form kooperativen Arbeitens dar, die keine spezifische Phasierung vorsieht. Stattdessen formulieren Johnson und Johnson vier theoretisch motivierte, übergeordnete Kernmerkmale des Learning-Together (Johnson & Johnson 1994a & 1994c):

- Positive Interdependenz: Gemeinsames Erreichen des Aufgabenziels.
- „Face-to-face"-Interaktion: Beschränkung der Gruppengröße auf vier bis fünf Schüler.

- Individuelle Verantwortlichkeit: Jeder Schüler besitzt entsprechend seiner Fähigkeiten eine persönliche Verantwortlichkeit für das Erreichen des Gruppenziels. Diese Verantwortlichkeit kann auch in einem individuellen Leistungstest überprüft werden.
- Soziale Kompetenzen: Bedeutsame Kompetenzen für die Kleingruppenarbeit, wie Kommunikation, Konfliktlösefähigkeit, Methoden der Entscheidungsfällung müssen von Lehrer zunächst im Klassenverband vermittelt werden.

Eine Weiterentwicklung dieses Ansatzes besteht in der ebenfalls von Johnson und Johnson vorgelegten „akademischen Kontroverse“ (Johnson & Johnson 1994b). Hierzu wird die Kleingruppe in zwei Fraktionen geteilt, die dann nach einem vorstrukturierten Argumentationsverlauf ein gegebenes Thema diskursiv erörtern.

3.2.2 Schülerteams und Leistungsdifferenzierung

„Schülerteams und Leistungsdifferenzierung“ (engl. *Student Teams and Achievement Devisions*, *STAD,* Slavin 1978, 1995) ist Name und Programm einer weiteren kooperativen Lernmethode, die ebenso wie die Methode des Learning-Together auf eine Aufgabendifferenzierung der Gruppenmitglieder verzichtet, in der jedoch nach einem teambezogenen Lernabschnitt eine individuell differenzierte Leistungskontrolle stattfindet, deren Ergebnisse in ein Gesamtgruppenresultat eingehen. Die Organisation und Durchführung von STAD vollzieht sich in fünf konsekutiven Phasen:

1. *Präsentation des Materials*: Vor Beginn der Teamarbeit wird das Lernmaterial vom Lehrer im Klassenverband vorgestellt. Diese Phase folgt weitestgehend der direkten Instruktion.
2. *Teamarbeit*: Leistungsheterogene Kleingruppen bearbeiten das Lernmaterial. Hierbei sollen die Schüler – zuweilen auch durch peer-tutoring Gespräche – sicherstellen, dass alle Teammitglieder die Aufgaben konzeptionell verstanden haben und adäquat beantworten können.
3. *Leistungsquiz*: Nach der Teamarbeit wird jeder Schüler einzeln in einem Quiz über das erworbene Wissen getestet.
4. *Berechnung des individuellen Lernerfolgs*: Ausgehend von einem individuellen, leistungsabhängigen Basiswert wird ein Punkteschlüssel erstellt, der dem einzelnen Schüler sowohl bei einem perfekten Testergebnis als auch bei einem individuell nachgewiesenen substantiellen Lernzuwachs eine maximale Punktzahl zukommen lässt. Dadurch werden alle Gruppen unabhängig von der Leistungsfähigkeit ihrer Teilnehmer in die gleiche Ausgangslage versetzt und die soziale Akzeptanz leistungsschwächerer Schüler erhöht.
5. *Teamwürdigung*: Durch Aufsummierung der Individualscores wird ein Gruppenscore ermittelt und die Gruppenleistung gewürdigt (Slavin 1978).

Im Gegensatz zur STAD-Methode beruhen die ebenfalls von Slavin (1983) entwickelten Teams-Spiele-Tourniere (engl. *Teams-Games-Tournaments*, *TGT*) auf einer Kombination von leistungsheterogenen Basisgruppen und leistungshomogenen

Wettkampfgruppen. Die Phasen der Materialpräsentation, der Kleingruppenkonstitution sowie die letzte Phase der „team recognition“ entsprechen vollständig der Vorgehensweise der STAD-Methode.

Die Teilnehmer innerhalb einer Kleingruppe werden gemäß ihrem aktuellen Leistungsstand eingestuft und treten dann gegen leistungsgleiche Schüler anderer Kleingruppen in einem Wissensspiel gegeneinander an.

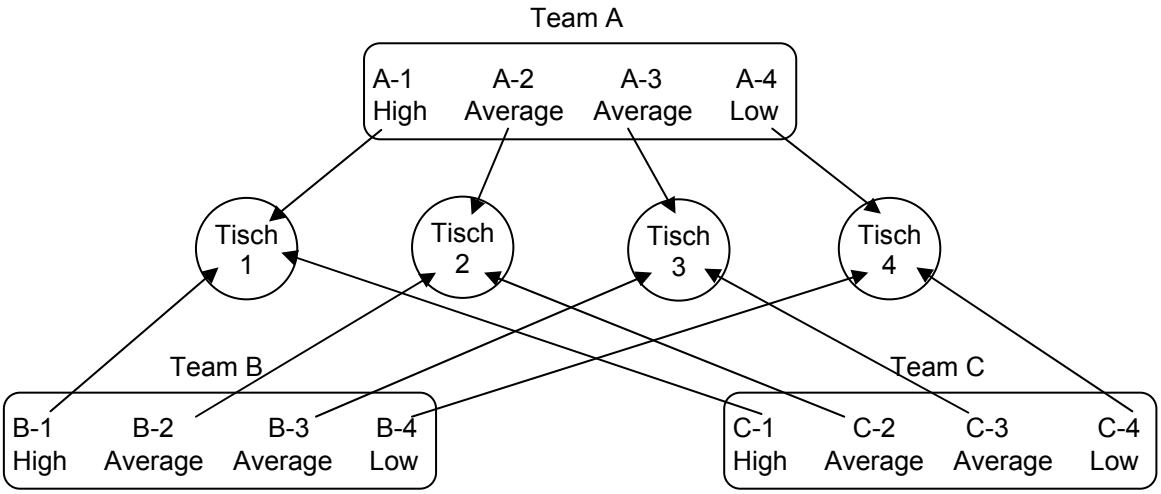


Abb. 3.4: Aufteilung der TGT-Turniertische (nach Slavin 1995)

Da die Beantwortung der Aufgaben primär deklaratives Wissen erfordert, eignet sich diese Methode nicht zu Erarbeitung neuer Inhaltsbereiche, wohl allerdings zur Lernzielkontrolle am Ende einer – beispielsweise nach der STAD-Methode konzeptionierten – Unterrichts-sequenz.

3.2.3 Jigsaw I und Group Investigation

Die Gemeinsamkeiten zwischen den kooperativen Arbeitsformen *Jigsaw I* und Group *Investigation* beschränken sich auf den Verzicht einer Gruppenbelohnung bei gleichzeitiger Aufgabendifferenzierung unter den Gruppenmitgliedern. Beide Methoden zielen jedoch auf unterschiedliche Facetten im schulischen Lernkontext, weshalb sie im Folgenden getrennt dargestellt werden.

Jigsaw I

Jigsaw I (Gruppenpuzzle), entwickelt von Aronson et al. (1978), ist eine kooperative Lernmethode, die sich besonders zur selbstständigen Aneignung von Wissen eignet. Die Methode vollzieht sich als Wechselspiel zwischen individueller Aufgabenspezifizierung in Basisgruppen, Erwerb von differenzierter Expertise in Expertengruppen und Rückkehr in die Basisgruppen zum wechselseitigen Austausch des erworbenen Fachwissens. Der hohe Grad des Expertenwissens des einzelnen Schülers soll in der Basisgruppe eine hohe positive Interdependenz bewirken.

Jigsaw I ist durch eine vierstufige Phasenabfolge gekennzeichnet (Clarke 1994):

1. *Einleitung*: Der Lehrer unterteilt das Thema zunächst in so viele sinnvolle Spezialthemen wie sich in jeder Kleingruppe Schüler befinden, teilt dann die

Klasse in leistungsheterogene Basisgruppen ein und verteilt die vorstrukturierten Arbeitsmaterialien zu den jeweiligen Spezialthemen.

2. *Fokussierte Exploration*: Die Schüler der Basisgruppen legen zu jedem Spezialthema einen verantwortlichen Experten fest. Alle Experten eines Spezialbereichs kommen zu einer Teamarbeit zusammen um sich die geforderte Expertise zu ihrem Thema anzueignen. Diese Phase des Jigsaw entspricht dem Learning-Together von Johnson und Johnson. Neben der fachlichen Komponente sollen in der Expertenrunde auch gemeinsame Strategien entwickelt werden, wie die jeweiligen Inhalte in den Basisgruppen zu vermitteln sind.
3. *Umorganisation und Berichterstattung*: Die Expertenrunden lösen sich auf, alle Schüler kehren zu ihren Basisgruppen zurück und berichten den anderen Gruppenmitgliedern aus den Expertenrunden. Dadurch wechseln alle Gruppenmitglieder zwischen der Rolle des Lernenden und des Lehrenden (peer-tutoring).
4. *Integration und Evaluation*: In den Kleingruppen oder auch im Plenumgespräch sollen an Hand der Bearbeitung einer komplexen Aufgabe alle Spezialgebiete integriert werden.

Group Investigation

Die Methode der „Group Investigation“, auch „Gruppenrecherche“ geht auf Sharan & Hertz-Lazarowitz (1980) zurück und fußt wie keine andere der hier dargestellten kooperativen Lernmethoden explizit auf dem reformpädagogischen Ansatz von Dewey und der von ihm begründeten Projektmethode. Im Gegensatz zum klassischen Projektunterricht zeichnet sich die Methode der Gruppenrecherche aber durch eine stärkere Vorstrukturierung durch den Lehrenden aus. Die Auswahl der Themen ist nicht von fächerübergreifenden Anforderungen bestimmt und erfolgt nicht durch die Schüler. Andererseits ist der Phasierung der Methode durch die Abfolge „Planung – Durchführung – Präsentation“ eine deutliche Nähe zur Handlungsorientierung zu entnehmen. Huber (1991) spricht daher auch von „Kleinprojekten“ und sieht die besondere Stärke dieser Methode darin, auf Grund des hohen Grads an Strukturierung auch komplexe, facettenreiche Aufgabenstellungen, die ein selbstständiges Problemlösen erfordern, in kooperativen Kleingruppen zu bearbeiten. Das problemorientierte Potenzial der Gruppenrecherche zeigt sich auch in den von Sharan und Sharan (1994) formulierten vier grundlegenden Charakteristika der Methode. Neben den auch aus anderen kooperativen Lernmethoden bekannten Prinzipien der *Interaktion* und der *intrinsischen Motivation* basiert diese Methode zusätzlich auf *Investigation* und *Interpretation*, deren besondere Bedeutung Sharan und Sharan (1994) mit Verweis auf Dewey wie folgt darstellen: „Meaningful learning proceeds through the steps of scientific inquiry, whereby students experience how knowledge is generated. In Dewey’s view, investigation of any subject can incorporate the essential features of the scientific method and thus can educate students in the spirit and method of scientific inquiry.”

Die formale Struktur der Gruppenrecherche umfasst die konsekutive Abfolge von sechs Phasen (Sharan & Sharan 1992 & 1994):

1. *Organisation der Gruppenarbeit*: Ähnlich der Vorgehensweise im Jigsaw werden die Schüler nach der Vorstellung des Themas in leistungsheterogene Kleingruppen eingeteilt. Sofern die Thematik eine arbeitsteilige Gruppenarbeit zulässt, können die Schüler gemäß ihrer Interessen eine Schwerpunktsetzung zur Bearbeitung des Themas vornehmen.
2. *Planung der Untersuchung/ Recherche*: Nachdem der Untersuchungsaspekt präzisiert und mit allen Gruppenteilnehmern vereinbart ist, entwickeln die Schüler einen Untersuchungsplan, der die weitere Vorgehensweise hinsichtlich der vorhandenen Ressourcen sowie der benötigten Materialien und Informationen strukturiert.
3. *Durchführung der Untersuchung/ Recherche*: Die Schüler führen die Untersuchung durch und analysieren ggf. Daten. Die Ergebnisse werden in der Kleingruppe diskutiert und mit dem Untersuchungsplan abgeglichen.
4. *Planung der Ergebnisdarstellung/Präsentation*: Die Gruppenmitglieder stellen die zentralen Ergebnisse ihrer Untersuchung fest und planen, was und in welcher Form sie dem Plenum berichtet werden.
5. *Ergebnisdarstellung der Kleingruppenrecherche*: Die Kleingruppen präsentieren ihre Ergebnisse der gesamten Klasse. Verständnisfragen seitens der zuhörenden Schüler werden in einer vom Lehrer geführten kurzen Diskussion geklärt; Schüler und Lehrer evaluieren die Qualität der Präsentation.
6. *Evaluation durch Lehrer und Schüler*: Für diese Phase ist keine feste Struktur vorgegeben. Denkbar ist die Beurteilung des Lernerfolgs durch Testleistungen, ähnlich der STAD-Methode. Der Lehrer kann aber auch schriftliche Auswertungen oder Berichte der Kleingruppenrecherche zur Bewertung heranziehen.

3.2.4 Jigsaw II

Jigsaw II (Slavin 1995) stellt eine Weiterentwicklung der Jigsaw I Methode dar, ergänzt um den Aspekt der individuellen Leistungskontrolle und der Gruppenbelohnung, so wie in STAD praktiziert. Von der Phasierung und vom Ablauf entspricht die Methode damit zunächst vollständig der Jigsaw I-Methode. Im Rahmen der Evaluationsphase (s.o.) müssen die Schüler einen Test absolvieren, der sich über alle Inhaltsbereiche erstreckt, um sicherzustellen, dass gleichermaßen die Lernziele ihres Spezialgebiets, also die von ihnen gelehrten Inhaltsbereiche, als auch die der gelernten Inhaltsbereiche erreicht wurden. Die Testauswertung erfolgt nach der aus der STAD-Methode bekannten kombinierten Messung von absoluter Testleistung und individuellem Lernzuwachs.

4. Zielsetzung und Design der Interventionsstudie

Gegenstand dieser Arbeit ist die Konzeption und Evaluation einer *Interventionsstudie* zum Themenbereich „Säure-Base“ für den Gymnasialbereich der Jahrgangsstufe sieben, NRW. Unter einer pädagogischen Interventionsstudie wird in diesem Kontext ein „theoriegeleiteter Eingriff in ein pädagogisches Feld zur Erreichung pädagogischer Ziele mit kontrollierter Wirkungsüberprüfung und Verallgemeinerungsabsicht“ verstanden (Doll et al. 2002). Folgt man dieser Definition, so sind für eine Interventionsstudie vier, die Versuchsplanung bestimmende, Elemente auszumachen:

1. Theoretische Fundierung (*theory-driven-approach*)
   Formulierung der zugrunde liegenden Theorie.
2. Zielformulierung
   Formulierung von Forschungsfragen bzw. Hypothesen; Konzeption eines auf die Forschungsfragen abgestimmten Designs zur Realisierung der unabhängigen Variablen.
3. Wirkungsforschung
   Erfassung der abhängigen Variablen und Berücksichtigung von Kovariablen; Berechnung von Signifikanzen und Effektgrößen.
4. Verallgemeinerungsabsicht
   Generalisierbarkeit der Ergebnisse, hohe externe Validität zur Übertragbarkeit der Ergebnisse auf andere Situationen, Personen oder Zeitpunkte.

Die vier Elemente lassen zum einen eine Orientierung an die klassische Phasierung der Grundlagenforschung - Konzeptualisierung, Implementation, Wirkungsforschung - erkennen, weisen aber mit Akzentuierung auf eine Verallgemeinerungsabsicht auch darüber hinaus. Eine Differenz zur Grundlagenforschung wird auch noch aus einer anderen Perspektive deutlich: Da eine Interventionsstudie nicht den kausalen Zusammenhang zweier theoretischer Konstrukte ermittelt, sondern ein technologisches *Produkt* (die Interventionsmaßnahme) mit einer theoretischen Variablen in Beziehung setzt, ist in Anlehnung an Rost (2000) statt von Grundlagenforschung hier treffender von Evaluationsforschung zu sprechen.

Neben der oben dargestellten Definition werden in der Literatur zum Teil deutlich komplexere Kennzeichenkataloge für Interventionsmaßnahmen formuliert (vgl. das zehnstufige Gütekriteriensystem von Hager und Hasselhorn 2000), die insbesondere den Aspekt der Wirkungsgeneralisierung stärker differenzieren.

Obschon große strukturelle Gemeinsamkeiten zwischen Grundlagen- und Evaluationsforschung sowie ein unscharfer Übergang zwischen beiden Designs zu konstatieren sind, zieht eine Verortung der hier dargestellten Interventionsstudie zur Evaluationsforschung Konsequenzen für die Qualität der Forschungshypothesen wie auch die Wahl des adäquaten Versuchsplans nach sich: Mit Hinblick auf die Forschungshypothese führt Rost (2000) an, dass eher von Forschungs*fragen* zu reden ist, die zwar wünschenswerter Weise einen theoretischen Hintergrund aufweisen, sich jedoch nicht unmittelbar, wie im Falle der Überprüfung zweier theoretische Konstrukte,

aus einer Theorie deduzieren lassen, sondern auf die technische Überprüfung eines Produktes fokussieren. Schließlich bedingt die Forderung nach Übertragbarkeit und Verallgemeinerbarkeit eine stärkere Gewichtung der externen Validität zu Lasten der internen Validität, weshalb sich das klassische Laborexperiment als gebotener Versuchsplan ausschließt (s. Kp. 5.2).

4.1 Standards der Untersuchungsplanung

Die empirische Unterrichtsforschung sucht nach Erkenntnissen über die Wirkmechanismen unterrichtlichen Handelns. Sie bedient sich hierzu der systematischen Auswertung von Untersuchungsdaten (genau genommen von ‚Erfahrungen', von griech: empeira = Erfahrung), die aus forschungsrelevanten *Merkmalen* der zugrundeliegenden Forschungsobjekte (Schüler, Lehrer, Klasse, Schule, Elternhaus) abgeleitet werden. Dabei gilt es zu berücksichtigen, dass die Untersuchungsdaten keineswegs, entgegen etwaiger Implikationen des wissenschaftlichen Sprachjargons, von den Forschungsobjekten selbst produziert werden. Vielmehr werden die Daten in einem wissenschaftlichen Transformationsprozess, beispielsweise der Diagnose des räumlichen Vorstellungsvermögens (Objekteigenschaft) an Hand einer bestimmten Testleistung (Objektmerkmal), generiert. Quantitative Daten liefern daher immer nur Informationen sekundärer Art, deren Güte maßgeblich davon abhängt, inwieweit das erfasste Merkmal die fokussierte Eigenschaft auch tatsächlich abbildet. Aus diesem Grunde wird in der Testtheorie die Validität eines Tests, seine Gültigkeit (genauer: „Ausmaß, in dem der Test das misst, was er messen soll" Rost 1996), als das entscheidende Kriterium angesehen, dem die weiteren Gütekriterien Reliabilität und Objektivität nachgeordnet sind.

Zur Differenzierung der zu quantifizierenden Forschungsmerkmale und zur näheren Beschreibung der Merkmalsunterschiede in einer Untersuchungsgruppe hat sich der Begriff der Untersuchungs*variable* eingebürgert. Unter einer Variablen versteht man somit ein Symbol für eine Menge von Merkmalsausprägungen (Bortz, Döring 1995). Im Folgenden sollen die in der vorliegenden Untersuchung relevanten Variablen an Hand der in der Literatur beschriebenen Typologisierung dargestellt werden. Nach Bortz und Döring (1995) lassen sich Variablen hinsichtlich ihrer Dimensionierung differenzieren in:

a) der Art der Merkmalsausprägung,

b) der empirischen Zugänglichkeit,

c) dem Stellenwert innerhalb der Untersuchung.

a) Art der Merkmalausprägung

Quantitative Variablen lassen sich hinsichtlich der *Ausprägungsform* des untersuchten Merkmals in *stetige* und *diskrete* Variabeln unterscheiden. Stetige Variablen zeichnen sich durch kontinuierliche und (theoretisch) unendlich viele Merkmalsausprägungen aus. So können beispielsweise die im Rahmen einer Videoanalyse vor-

genommenen Merkmalsauszählungen innerhalb eines Zeitintervalls jeden beliebigen ganzzahligen Wert annehmen. Dahingegen beschreiben diskrete Variablen eine diskontinuierliche, endliche Abstufung von Merkmalsausprägungen (z.B. Schulnoten). Diskrete Variablen werden weiterhin abhängig von der Anzahl ihrer Abstufungen in dichotome (zweifach gestufte) und polytome (mehrfach gestufte) Variablen unterschieden.

b) Empirische Zugänglichkeit

Sofern eine Merkmalsausprägung einer direkten empirischen Beobachtung zugänglich ist, spricht man von einer *manifesten* Variable. So lässt sich beispielsweise die Lernleistung eines Schülers direkt aus dem Ergebnis eines - validen - Tests ableiten. Anders verhalten sich Variablen, denen theoretische Konstrukte zugrunde liegen und die einer direkten Beobachtung nicht zugänglich sind. Beispielsweise ist das Konstrukt „Freizeitverhalten“ einer unmittelbaren Bewertung nicht zugänglich, sondern muss durch andere beobachtbare, manifeste Variablen indirekt ermittelt werden (sportliche Aktivität, kulturelle Interessen, zeitliche Verfügung etc.) Derartige Ausprägungen werden auf Grund ihrer nicht direkten Zugänglichkeit als *latente* Variablen bezeichnet. Krauth (1995) geht so weit, dass er latente Variablen auch als fiktive Variablen bezeichnet, da es keinen plausiblen Beweis für Ihre Existenz gibt.

c) Stellenwert innerhalb der Untersuchung

Art und empirische Zugänglichkeit der Merkmalsausprägungen sind gegenüber der wissenschaftlichen Fragestellung invariant, da es sich hierbei um ontologische *Eigenschaften* des Merkmals handelt. Bei einer gegebenen Fragestellung liegen stetige und diskrete sowie latente und manifeste Variablen daher gegeben vor. Diese Variablen werden in der vorgelegten Untersuchung aus diesem Grund hinsichtlich beider Aspekte an den entsprechen Stellen der Datenauswertung spezifiziert.

Dahingegen erfordert die dritte Dimensionierung von Untersuchungsvariablen, der Stellenwert der Variablen in der Untersuchung, ein hohes Maß an Vorüberlegungen in der Planungsphase, da es sich hierbei um zugeordnete *Funktionen* des Merkmals handelt. Ein möglichst hoher Grad an Vollständigkeit in der Erfassung, Zuordnung und Beschreibung aller relevanten Variablen ist daher ein wesentliches Gütekennzeichen einer empirischen Erhebung. Im Folgenden werden die Variablen der Untersuchung hinsichtlich ihres Stellenwertes ausführlich dargestellt. Einen Überblick über die Klassifizierung von Variablen in Abhängigkeit ihres Stellenwertes im Untersuchungsplan gibt Abbildung 4.1 (in Anlehnung an Bortz 1999)

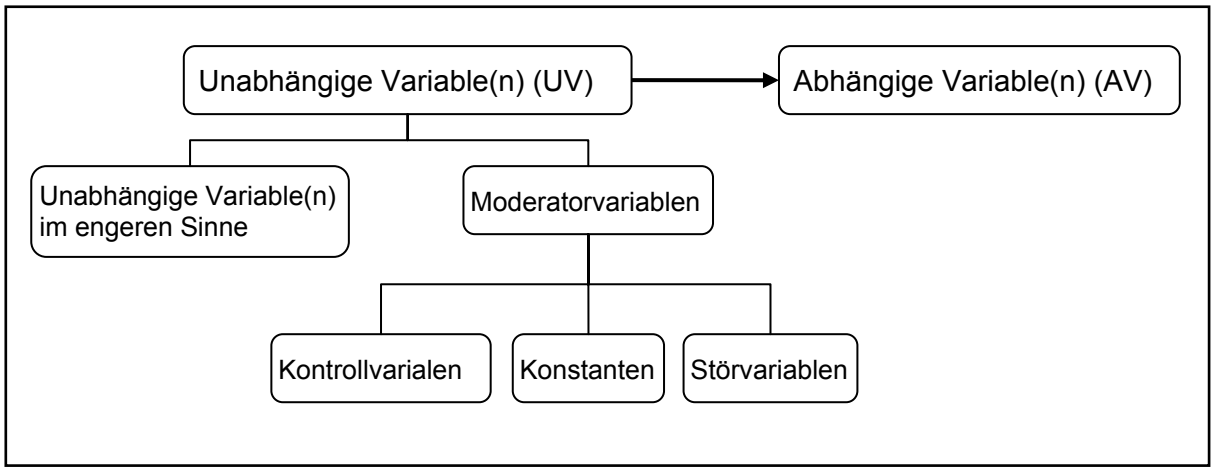


Abb. 4.1: Stellenwert von Variablen in einer Untersuchung

Die eindeutige Trennbarkeit der unabhängigen Variablen von den abhängigen Variablen sowie deren konsekutive Sequenzierung (UV → AV) ist eine zentrale Voraussetzung für ein *experimentelles Untersuchungsdesign*. Die besondere Stärke experimenteller Versuchsanordnungen liegt in der Möglichkeit begründet, einen vermuteten Zusammenhang zweier Ereignisse als kausale Folge von Ursache (der unabhängigen Variable) und Wirkung (auf die abhängige Variable) aufzuzeigen und inferenzstatistisch zu überprüfen. Kerlinger (1986) weist darauf hin, dass die Unterscheidung von unabhängigen und abhängigen Variablen damit die bedeutendste Variablengruppierung in den Sozialwissenschaften darstellt.

Ist die dargestellte Trennung und Sequenzierung von abhängigen und unabhängigen Variablen nicht gegeben, so liegt eine *Korrelationsstudie* vor; bei einer eher unsicheren Sequenzierung, wenn etwa die abhängigen Variablen gegeben sind und die zugrunde liegenden unabhängigen Variablen gesucht werden, spricht man von einer „*Ex-post-facto-Studie*“ (vgl. Musahl, Schwennen 2001).

Ein in der fachdidaktischen Unterrichtsforschung im Rahmen von unterrichtlichen Interventionsstudien häufig realisierter Versuchsplan zeichnet sich durch die alleinige Testung einer Versuchsgruppe ohne Paralleltestung einer Kontrollgruppe aus (Vorher-Nachher-Messung an einer einzigen Versuchsgruppe).

Versuchsgruppe	Vorher-Mesung	Treatment *X*	Nach-Messung
1	$\overline{Y}_{vor}$	X	$\overline{Y}_{nach}$

Abb. 4.2: One-group before-after design (nach Sarris 1992)

Wenngleich dieser Versuchsplan eine strikte Trennung von unabhängiger Variable (Treatment X) und abhängige Variable (Y) vorsieht und die mit diesem Design erzielten Zeiteffekte häufig hochsignifikant und hinsichtlich der erzielten Effektstärke beträchtlich sind, so kann aus verschiedenen Gründen dennoch nicht von einem experimentellen Design gesprochen werden: Aus der fehlenden Kontrollgruppe und dem unkontrollierten Einfluss der Randbedingungen (Störvariabeln) sowie der fehlenden

Randomisierung von Untersuchungsbedingungen und Probanden resultieren entscheidende Defizite, die die interne Validität dieser Studien massiv gefährden. Sarris (1992) weist darauf hin, dass ein derartiger Versuchsplan bei chemischen und physikalischen Laborexperimenten legitimiert werden kann, da in diesen Fällen von einer unverzerrten Datenerhebung ausgegangen werden kann und die numerische Differenz der abhängigen Variablen über das Treatment als valider Indikator für den experimentellen Effekt angesehen werden kann. Diese Vorrausetzung ist bei psychologisch-sozialwissenschaftlichen Untersuchungen aus den oben genannten Gründen allerdings nicht gegeben. Gleiches gilt auch für Versuchspläne, die gänzlich auf eine Vorher-Messung verzichten (one-shot case study bzw. Schrotschuss-Design) sowie für Versuchspläne, die einem statischen Gruppenvergleich (mehrer Versuchsgruppen zu einem Messzeitpunkt) folgen. Da auf diese Weise keine kausaltheoretisch begründeten Hypothesen vor Versuchsbeginn postuliert werden können und die gewonnenen Messwerte somit nicht schlüssig (konklusiv) zu interpretieren sind, bezeichnet man diese Untersuchungsdesigns kategorial als *vorexperimentelle* bzw. *ungültige Versuchspläne*. Vorexperimentellen Versuchsplänen scheint zur quantitativen Abschätzung des experimentellen Faktors im Rahmen von Pilotstudien eine gewisse Bedeutung zuzukommen, auf Grund des invaliden Designs und der nicht auszuschließenden Alternativerklärungen wird vor dem unkritischen Gebrauch jedoch gewarnt (Sarris 1992).

Wie oben erwähnt, stellt die Technik der Randomisierung ein weiteres Kriterium für einen experimentellen Versuchsplan dar. Unter Randomisierung versteht man die zufällige Aufteilung der Untersuchungsprobanden auf die verschiedenen Untersuchungsbedingungen mit dem Ziel der Einflussminimierung von personenbezogenen Störvariablen. Dadurch erklärt sich, dass mit zunehmender Randomisierung einer Studie die interne Validität steigt. Die folgende Abbildung gibt einen Überblick über die Typologisierung der gängigen experimentellen Versuchspläne in Abhängigkeit von Randomisierung und Kontrolle weiterer Randbedingungen:

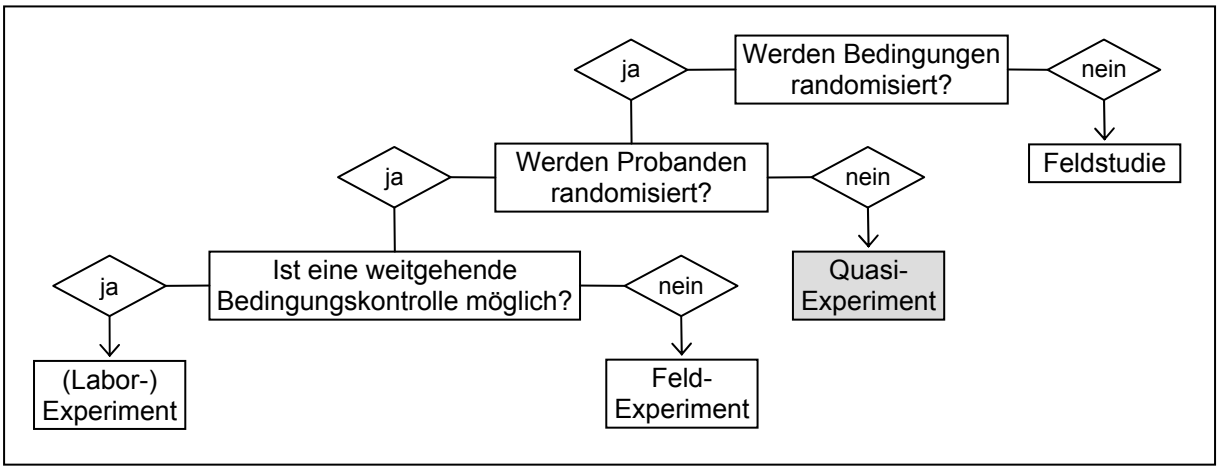


Abb.: 4.3 Übersicht experimenteller Versuchspläne (nach Musahl, Schwennen 2001)

Die hohe Bedingungskontrolle potentieller Störvariablen wie auch die Randomisierung der Bedingungen und Probanden lassen im Falle des Laborexperiments in ei-

nem höchstmöglichen Maße Veränderungen der abhängigen Variablen allein auf den Einfluss der unabhängigen Variabeln zurückführen, weshalb dieses Design die größte interne Validität aufweist. Laborexperimente werden deshalb in der Regel in speziellen, standardisierten Umgebungen durchgeführt, die frei von äußeren, nicht erwünschten Stimuli sind. Die artifizielle Laborsituation bedingt allerdings häufig eine Herabsetzung der Generalisierbarkeit und Übertragbarkeit der Forschungsergebnisse auf andere Situationen und damit auch der externen Validität. Dahingegen bieten quasi-experimentelle Felduntersuchungen auf Grund ihrer biotischen und nicht manipulierten Umgebungsbedingungen eine vergleichsweise hohe externe Validität (vgl. Bierhoff & Rudinger 1996) wie auch die folgende Abbildung verdeutlicht:

Abb.: 4.4: Auswirkung der Versuchsanordnung auf die interne bzw. externe Validität

Dieser Befund stellt die Versuchsplanung schulischer Interventionsstudien vor ein Dilemma, da im Rahmen des Regelunterrichts die Probanden auf Klassenebene zusammengefasst werden und eine natürlich gegebene, invariante Gruppe darstellen. Zielt die Forschungsfrage nun auf den Führungsstil des Lehrers oder, wie im vorliegenden Falle, auf einen Vergleich zweier Unterrichtsmethoden, so können die Probanden nur auf Klassenebene und damit nicht-randomisiert dem Treatment zugewiesen werden, um nicht ein technisch höchst fragwürdiges Design zu entwickeln, in dem ein Lehrer in einer Klasse zwei verschiedene Unterrichtsmethoden umsetzt. Daraus folgt allerdings, dass ein bereits im Vorfeld zufällig vorhandener Unterschied zwischen den Gruppen sich somit nicht durch zufällige Aufteilung der Probanden auf ein Treatment minimieren lässt, sondern zu einer Verzerrung der Daten führt. Daher ist bei quasi-experimentellen Versuchsplänen in besonderem Maße auf Bedingungskontrolle der moderierenden Variablen zu achten, um so auch unter nicht-randomisierten Bedingungen die interne Validität der Studie zu gewährleisten.

4.2 Forschungsfragen und Variablen der Untersuchung

4.2.1 Forschungsfragen

Ziel der Studie ist die Evaluation einer verschränkten Intervention im schulischen und häuslichen Bereich. Im Rahmen der schulischen Intervention erfolgt die Umsetzung einer nach dem Prinzip der *Group Investigation* (Gruppenrecherche, Sharan & Sharan 1994) konzipierten kooperativen Gruppenarbeit. Die häusliche Intervention sieht eine Beteiligung der Eltern bei der Hausaufgabenbearbeitung zur Vor- und Nachbereitung der Unterrichtsstunden vor. Kontrastiert werden die Interventionsgruppen durch Kontrollgruppen, die für den schulischen Bereich durch eine Problembearbeitung im lehrerzentrierten Unterrichtsgespräch gekennzeichnet sind und die Rolle der

Eltern bei der Hausaufgabenbearbeitung auf eine reine Kontrollfunktion beschränken. Die Wirksamkeit aller Treatmentvarianten wird in einem 2✶2 faktoriellen, quasi-experimentellen Design getestet.

Hieraus leiten sich mit Hinblick auf die zu evaluierende Unterrichtsintervention folgende Forschungsfragen ab:

1. Zeigen die unterrichtlich intervenierten Lernenden höhere bereichsspezifische Lernzuwächse als die Probanden der Kontrollgruppe?

2. Wird ein durch die Unterrichtsintervention erworbener Lernzuwachs durch gleichzeitige häusliche Intervention noch weiter vergrößert?

3. Lassen sich für die Phase der Problembearbeitung in den unterrichtlichen Treatments differentielle Muster identifizieren?

4.2.2 Untersuchungsvariablen

Der dieser Studie zugrundeliegende DFG-Antrag „Schulische und familiale Bedingungen des Lernens und Lernmotivation im Fach Chemie: Evaluation eines integrierten Interventionskonzeptes zur Säure-Base-Thematik“ (Sumfleth, Wild 2000) weist folgende unabhängige und abhängige Variablen aus:

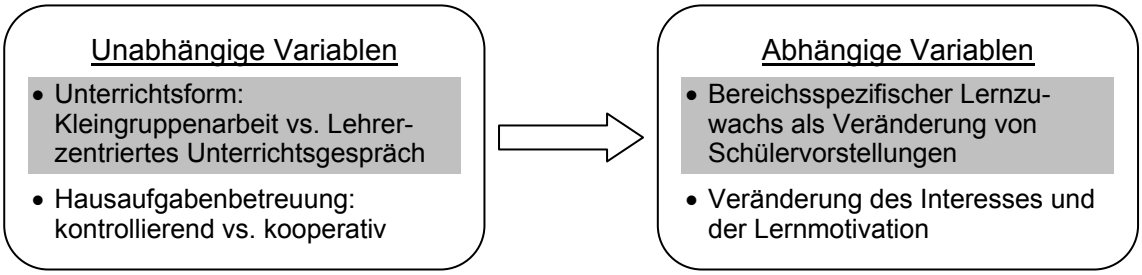


Abb. 4.5: Unabhängige und abhängige Variablen der Untersuchung

Im Rahmen dieser Arbeit erfolgt primär die Überprüfung der Lernwirksamkeit der Unterrichtsform auf die kognitiven Fähigkeiten (grau unterlegten Variablen) im Sinne einer Unterschiedshypothese: Analog zur ersten Fragestellung wird erwartet, dass sich der Lernzuwachs zwischen den beiden Unterrichtsformen signifikant zugunsten der kooperativ unterrichteten Probanden unterscheidet. Auf Grund des zweifaktoriellen Versuchsplans wird zusätzlich geprüft, ob die Treatmentkombination „schulische und häusliche“ Intervention zu einem Vorteil gegenüber den einfachintervenierten, kooperativ unterrichteten Probanden führt (Interaktionshypothese).

Ein ausführlicher Vergleich zur Überprüfung der differentiellen Wirkmächtigkeit der Hausaufgabenbetreuung (Haupteffekt) ist nicht Gegenstand dieser Arbeit, sondern wird im Rahmen der Hausaufgabenstudie bei Nicolai (2004) behandelt. Ebenso werden die von Sumfleth und Wild (2000) im Sinne einer multikriterialen Zielerreichung erwarteten motivationalen Effekte als weitere treatment-abhängige Variable, nur in sofern berücksichtigt, als dass sie in Kovariablenanalysen einfließen, um Aufklärung über den Einfluss unterschiedlicher Lehrpersonen auf die Entwicklung des Interesses und der Motivation der Lernenden zu geben. Detaillierte Analysen zum Einfluss der

Unterrichtsintervention auf Interesse, Motivation sowie weiterer affektiver Parameter sind bei Wild, Exeler und Gerber (2003) nachzulesen.

4.2.3 Personengebundene und untersuchungsbedingte Störvariablen

Gemäß Abbildung 4.1 wirken neben den unabhängigen Variablen noch weitere Randvariablen moderierend auf die abhängige Variable ein und erschweren so die Aussagekraft bezüglich des kausalen Wirkungseinflusses der unabhängigen Variablen auf die abhängige. Diese Problematik stellt sich insbesondere bei Felduntersuchungen, deren interne Validität durch den lebensnahen und daher zahlreichen Einfluss von Störvariablen gefährdet ist. Damit ein zu verzeichnender Unterschied zwischen Experimental- und Kontrollgruppe tatsächlich auf Treatmenteffekte zurückgeführt werden kann, ist daher sicherzustellen, dass die Stichproben vor der Untersuchung hinsichtlich untersuchungsrelevanter Merkmale vergleichbar bzw. äquivalent sind (Bortz, Döring 1995). Dies geschieht im Idealfall durch Konstanthalten oder, sofern dies nicht möglich ist, durch ein Kontrollieren (Aufnehmen) der betreffenden Variable. Im Folgenden wird die Handhabung potentieller Störvariablen näher dargestellt:

Schuleffekte

Um sicherzustellen, dass ein Treatmenteffekt nicht durch Spezifika einer einzelnen Schule alternativ erklärt werden könnte, wurden die Treatmentkombinationen paarweise den Schulen zugeordnet (Bildung von *matched samples*). Der Einfluss einer Schule ist daher nur zwischen den Schulen, nicht aber zwischen den Treatments nachweisbar.

Einfluss der Lehrerpersönlichkeit

Der Einfluss der Lehrerpersönlichkeit auf die Schulleistung ist seit langem bekannt. Daher wurden die an der Untersuchung teilnehmenden Lehrer über die Treatments gleichverteilt und somit an der Realisierung *aller* experimentellen Bedingungen beteiligt. Es wurde daher angestrebt, dass ein Lehrer den gesamten siebten Jahrgang einer Schule unterrichtet, sodass der Einfluss der Lehrerpersönlichkeit auf Schulebene als konstant angesehen werden kann. In einem Fall war dies aus stundenplantechnischen Gründen nicht möglich. Es gelang jedoch ein erfahrenes Lehrerehepaar auszuwählen, von dem davon ausgegangen werden kann, dass auf Grund jahrelanger gemeinsamer Planungserfahrung ein Höchstmaß an interpersoneller Übereinstimmung gewährleistet ist.

Schwierigkeitsgrad der Problemstellung

Der Schwierigkeitsgrad der Problemstellungen zwischen Experimental- und Kontrollgruppe kann als identisch und damit konstant angesehen werden. Beide Gruppen bekamen die identischen Arbeitsaufträge und die gleichen Medien. Lediglich die Größe der Experimentiermaterialien war in der Kontrollgruppe den Bedürfnissen von Demonstrati-

onsexperimenten angepasst. Ebenso wurden in beiden Gruppen die Versuchsergebnisse und Auswertungen in standardisierten Tafelbildern dokumentiert.

Time on task

Der enorme Einfluss der Unterrichtsquantität auf die Leistungsentwicklung ist durch verschiedene Metaanalysen hinlänglich gut dokumentiert. So weist die auf 31 Metaanalysen fußende „Megaanalyse“ von Walberg (1986) einen korrelativen Zusammenhang zwischen Unterrichtsquantität und Lernerfolg von 0.4 nach. Insbesondere im Fall einer Interventionsstudie besteht die Gefahr, der Experimentalgruppe auf Grund einer anspruchsvolleren Organisations- und Arbeitsphase gestützt von Innovationseffekten mehr Zeit für den gleichen Inhaltsbereich einzuräumen als der Kontrollgruppe. Zur Nivellierung unterrichtsmethodischer Innovationseffekte sowie zur Routinebildung der methodenbedingten organisatorischen Gruppenarbeitsprozesse wurde der Intervention eine nach dem gleichen Prinzip konzipierte Vorstudie aus dem Themenbereich „Verbrennungsreaktionen“ vorweggeschaltet. Zusätzlich wurde für jede Stunde ein standardisierter Verlaufsplan mit verbindlicher Zielvorgabe entwickelt. Experimental- und Kontrollgruppe wurden somit sowohl für die gesamte Unterrichtsreihe als auch innerhalb jeder einzelnen Unterrichtstunde zeitlich parallelisiert und die effektive Lernzeit konstant gehalten.

Vorwissen, soziale Herkunft, Intensität der elterlichen Hausaufgabenbetreuung, kindperzipiertes Instruktionsverhalten von Lehrenden und Eltern, Zeitpunkt der Unterrichtsstunde in der Stundentafel

Hierunter subsummiert sich eine ganze Palette potentieller Störvariablen, die nur äußerst schwierig eliminiert oder konstant gehalten werden können. Dennoch ist insbesondere der lernerfolgsterminierende Einfluss von Vorwissen und sozialer Herkunft gut erforscht (vgl. Kp. 2). Im Rahmen der Datenerhebung wurden daher die aufgeführten Variablen in Schüler- und Elternfragebögen erhoben, um sie einer statistischen Kontrolle zugänglich zu machen. Das bereichsspezifische Vorwissen wurde mit Hilfe von paper-pencil-Methoden, bestehend aus gebundenen Antwortformaten (Multiple-Choice-Test) und offenen Fragen, gestaffelt nach zunehmender Komplexität, erhoben (s.a. Kp. 4.3.2 Erhebungsinstrumentarium). Additiv wurden im Rahmen des Jugendlichenfragebogens die aktuellen Zeugnisnoten in den Hauptfächern und den Naturwissenschaften der Lernenden erfasst. Eine Kontrolle des sozialen Hintergrunds in den Treatments erfolgte mittels Items des Elternfragebogens, die neben der Schul- und Berufsausbildung Informationen auf einer elfstufigen Skala zum Status der gegenwärtigen Berufstätigkeit der Eltern enthielten.

Weiterhin wurden an Hand von Fragebogendaten Schülereinschätzungen bezüglich des Lehrverhaltens ihrer Lehrer, das traditionelle Verhalten der Eltern bei der Hausaufgabenbetreuung wie auch eine Einschätzung der Lehrer über das Verhalten der einzelnen Schüler gewonnen.

Schülerpersönlichkeit, Krankheitsausfälle, Hitzefrei, unvorhergesehne Tagesereignisse

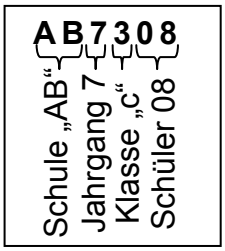
Auf Grund der unterrichtsnahen Bedingungen der Interventionsstudie ließen sich diese Bedingungen nicht kontrollieren und gingen als echte Störvariablen in die Erhebung ein. Besonders wiederholte Krankheitsausfälle, die ein Nachtesten einzelner Probanden verhinderten, führten zu einer Dezimierung des Datensatzes.

4.3 Untersuchungsdesign, Testinstrumentarium und Arbeitsplan

4.3.1 Stichprobe und Untersuchungsdesign

Die Erhebung fand in der siebten Jahrgangsstufe zweier Duisburger Gymnasien mit vergleichbarem Einzugsgebiet statt. Beide Gymnasien werden in der Unterstufe vierzügig geführt, so dass aus der Beteiligung beider gesamten siebten Jahrgänge insgesamt acht Untersuchungsklassen resultierten. Allen Untersuchungsklassen wurde ein Code zugewiesen, der sich aus einer Angabe über die Schule (KG bzw. FH), der Jahrgangsstufe (7) und einer numerischen Zuordnung der Klassen a-d (1-4) zusammensetzt. Innerhalb der Klasen wurden die alphabetischen Schülerlisten in einen zweistelligen Zahlencode überführt, so dass jeder Klasse und jedem Schüler ein eindeutiger, aber anonymisierter alphanumerischer Code zuzuordnen ist, der externen Personen keinen Personenrückschluss gestattet.

Codebeispiel:



Aus den acht Untersuchungsklassen resultierte zu Beginn der Studie eine Ausgangsstichprobe mit insgesamt 215 Probanden (davon 51,2% weiblich), deren Zahl sich jedoch versetzungsbedingt im Laufe der Studie (s.u.) dezimierte. Wie bereits unter 4.2.3 erwähnt, nahmen an der Untersuchung drei Lehrer teil.

Zur Überprüfung der Wirkungsmächtigkeit der schulischen (im Folgenden U+) und häuslichen (H+) Intervention wurde gemäß folgendem Schema ein 2*2-Design zur vollständigen Kombination aller Interventionsmaßnahmen realisiert:

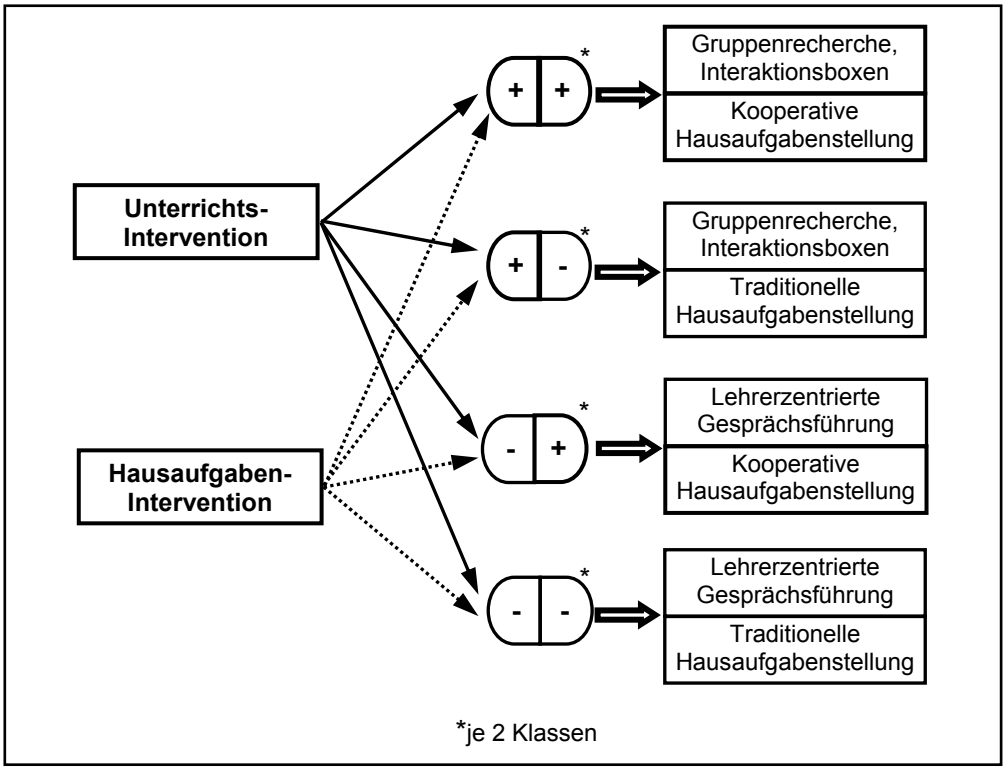


Abb. 4.6: 2*2-Design der Unterrichts- und Hausaufgabenintervention

Von besonderem wissenschaftlichen Interesse sind hierbei die Interventionsergebnisse der doppelt intervenierten „++“-Zelle, da hier, im Gegensatz zu den anderen Treatmentkombinationen, keine solitären Instruktionstechniken implementiert werden, sondern durch eine *Orchestrierung der Interventionsmaßnahmen* ein ausgewogenes Verhältnis von instruktionaler Unterstützung und selbstständiger Exploration gegeben ist und daher die stärksten und nachhaltigsten Effekte zu verzeichnen sein sollten.

Aus den oben genannten Gründen erfolgte durch Setzung an jeder Schule die Realisierung aller Treatmentvarianten. Die Zuordnung der Klassen auf das jeweilige Treatment erfolgte auf Schulebene hingegen randomisiert. Die vier aus dem Design resultierenden Treatmentkombinationen wurden mit je zwei Klassen besetzt und waren daher annähernd gleich stark repräsentiert. Die absoluten Probandenzahlen bezogen auf die jeweilige Untersuchungszelle betragen:

	U+/H+	U+/H-	U-/H+	U-/H-
Probanden	50	53	56	52

Tab. 4.1: Probandenverteilung auf die Untersuchungszellen

Die Klassengrößen umfassten ein Spektrum von 24 bis 28 Schülern, so dass pro Klasse jeweils 6-7 Kleingruppen gebildet wurden. In den überwiegenden Fällen setzte sich eine Kleingruppe aus vier Schülern zusammen; in Abhängigkeit von der Klassengröße musste in einigen Fällen auf Dreier- bzw. Fünfergruppen ausgewichen werden.

Die Zusammensetzung der Kleingruppen erfolgte nach einem mehrstufigen Prinzip: Zunächst wurde den Schülern die Gelegenheit gegeben, selbst Gruppen zu bilden, die dann in einem nächsten Schritt vom jeweiligen Fachlehrer mit Hinblick auf Arbeitsfähigkeit und Arbeitssicherheit geprüft wurden. Die auf dieser Basis gebildeten Kleingruppen wurden nach dem Prinzip der „mittleren Notendiskrepanz“ mit den erhobenen mittleren Schulnoten abgeglichen. Ziel hierbei war es, sicherzustellen, dass die mittlere Notendiskrepanz von +/- einer Schulnote innerhalb einer Kleingruppe nicht überschritten wurde. Auf diese Weise wurden Kleingruppen gebildet, die sich hinsichtlich ihrer Leistungsfähigkeit weder durch übermäßige Homogenität noch durch Heterogenität auszeichnen und gleichzeitig ein vertretbares Maß an Vergleichbarkeit zwischen den Gruppen bieten, da die intragruppalen und intergruppalen Varianzen nur minimal differieren. Darüber hinaus ist anzunehmen, dass die besten Testleistungsergebnisse von fortgeschritteneren Schülern erbracht werden, deren handlungsleitende Kognitionen im Vergleich zu schwächeren Schülern weniger stark auf alltagsweltlichen Vorstellungen sondern eher auf fachwissenschaftlich angemesseneren Konzepten fußen. Auch unter diesem Gesichtpunkt birgt die realisierte Gruppeneinteilung der „mittleren Diskrepanz“ Vorteile gegenüber einer leitungshomogenen Gruppeneinteilung.

4.3.2 Erhebungsinstrumentarium

In Anlehnung an die in der Evaluationsforschung gebräuchliche Bezeichnung der „summativen Evaluation“, als zusammenfassende Beurteilung der Wirksamkeit einer vorgegebenen Intervention (Bortz & Döring 1995), sollen aus Gründen der Übersichtlichkeit die im Rahmen von Vor- und Nachtests erhobenen Daten als *summative Daten* bezeichnet und den im Verlauf der Intervention erhobenen *Prozessdaten* gegenübergestellt werden. Diese Einteilung ist arbiträr und spiegelt nicht *disjunkte* Datenkategorien wieder, sondern charakterisiert vielmehr den *Zeitpunkt* der Datenerhebung (vor/nach oder während der Intervention). Die Leistungsfähigkeit summativer Daten und Prozessdaten zeigt sich über korrelative Zusammenhangsmaße in ihrem wechselseitigen Aufklärungspotenzial an der Fragestellung.

Summative Daten

Alle summativen Daten wurden mit Hilfe von paper-pencil-Verfahren erhoben. Den Schülern wurden im Rahmen eines Pre-Post-Follow-up-Designs zwei Fragebogeninstrumente zur Bearbeitung vorgelegt. Neben einem curricular validen, dreiteiligen Fachtest wurde ein komplexer Schülerfragebogen (Sumfleth, Wild 2001) zur Erfassung abhängiger und zu kontrollierender Variablen eingesetzt. Als zusätzliche Informationsquelle moderierender Variablen erhielten die Eltern einen Fragebogen (Pre-Post). Unmittelbar im Anschluss an die Intervention erhielten auch die beteiligten Lehrpersonen einen zweiteiligen Fragebogen, der sowohl Auskunft über eine allgemeine qualitative Bewertung der Interventionsmaßnahme geben sollte als auch eine Einschätzung der interventionsbegleitenden Entwicklungsverläufe der einzelnen Schüler vorsah. Die Kombination des Instrumentariums gibt somit Auskunft über die Leistungs- und Motivationsentwicklung der Schüler aus der Perspektive aller an der

Intervention beteiligten Personengruppen (Schüler-Eltern-Lehrer). Tabelle 4.2 gibt einen Überblick über den Einsatz der Fragebogenbogeninstrumente sowie eine selektive Auswahl der darin enthaltenen Skalen:

Test	**Messzeitpunkt**			**Testform/ -konstrukt**	**Erfassung von**	
	pre	post	foll' up		AV*	KV*
Fachtest	X	X	X	Mehrfachwahlaufgaben	X	
	X	X	X	Offene Fragen	X	
	X	X	X	Triaden-Test	X	
Schüler-fragebogen	X	X	X	Sach- und Fachinteresse	X	
				Schulnoten, Repetitantenstatus		X
				Motivationale Orientierung	X	
				Perzipiertes Instruktionsverhalten		X
Elternfragebogen	X	X	---	Chemische Vorbildung		X
				Erwerbstätigkeit & Sozioökonomischer Status		X
				Intensität der Hausaufgabenbetreuung		X
Lehrer-fragebogen	---	X	---	Beurteilung der Intervention		X
	---	X	---	Begabungs- und motivationsrelevante Schülermerkmale	X	

* AV = Abhängige Variable, KV = Kontrollvariable

Tabelle 4.2: Ausgewählte Skalen der eingesetzten Fragebögen

Der *Chemie-Fachtest* wurde von 98,6% der Schüler im Pretest und von 99,1% im Posttest beantwortet. Ausfälle ergaben sich nur durch Krankheit am Tag der Erhebung.

Die dreiteilige Gliederung des Chemie-Fachtests (siehe Anhang) in Multiple-Choice Aufgaben, offene Fragen und Triaden-Test wurde zur differenzierten Erfassung der Schülervorstellungen auf unterschiedlichen Verständnisebenen vorgenommen. Auf Grund umfangreicher Vorarbeiten zur Säure-Base-Thematik (Sumfleth 1987, Sumfleth, Ploschke u. Geisler 1999, Sumfleth u. Geisler 1999) liegen detaillierte Kenntnisse über häufig zu beobachtende Fehlkonzepte zu dieser Thematik vor, die bei der Konzeption der Items und Aufgaben wie auch der gesamten Konzeption der Unterrichtsreihe (vgl. Kp. 6.2) zum Tragen kamen. Der Multiple-Choice-Test zielt vornehmlich auf abrufbares, reproduktives Wissen, wohingegen die Beantwortung der offenen Fragen das selbstständige Formulieren der Antworten erfordert. Zusätzlich sind die offenen Fragen nach zunehmender Komplexität gestaffelt und repräsentie-

ren die schulischen Anforderungsbereiche „Reproduktion“, „Transfer“ und „problemlösendes Denken“.

Abschließend wird mit varianzanalytischen Methoden geprüft, ob sich die Schüler der verschiedenen Experimentalgruppen in der durchschnittlich erreichten Punktzahl unterscheiden. Da bislang kein valides und reliables Testinstrumentarium zum Themenbereich Säure-Base vorliegt, handelt es sich bei allen Tests um Eigenkonstruktionen. Die Tests sind curricular valide, beziehen sich direkt auf die Unterrichtsreihe und prüfen die Inhalte, die in den Experimentalgruppen erarbeitet worden sind. Die interne Konsistenz der Tests beträgt $.57 < \alpha < .79$ (Post- und Follow-up-Test) und kann als zufriedenstellend bezeichnet werden (vgl. Sumfleth, Wild 2001).

Der Triaden-Test ist auch bezüglich des Testformats eine vollständige Neukonstruktion und wurde in Anlehnung an die Verknüpfungstests (Sumfleth 1987) gemeinsam mit der BIQUA-Projektgruppe von Leutner (Leutner, Leopold 2001) entwickelt. Der Test dient der Erfassung von Wissen auf der Ebene des Verständnisses unter besonderer Berücksichtigung begrifflicher Zusammenhänge und besteht aus Items, welche die Beschreibung des Zusammenhangs von jeweils zwei Begriffen in einem Kontext verlangen, der durch einen dritten festgelegt wird. Triaden, die sich in einem Begriff unterscheiden und daher eine identische Paar-Relation enthalten, bieten somit die Möglichkeit, die Anwendung von Wissen in multiplen Kontexten prüfen zu können (s. Abbildung 4.7).

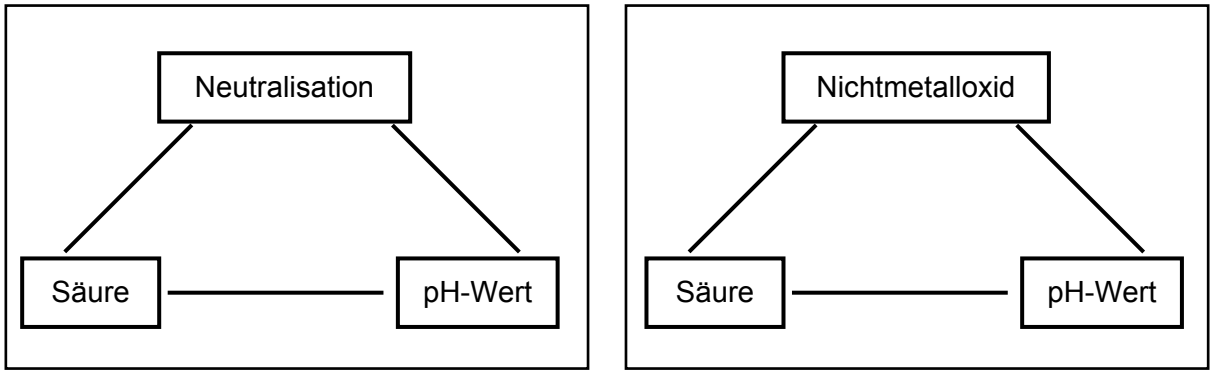


Abb. 4.7: Konsistenz der Proportion „Säure – pH-Wert“ in Abhängigkeit von einem dritten differierenden Begriff

Die Aufgabe der Probanden besteht darin, jede der drei Paar-Relationen zu erläutern. Wie die Ergebnisse einer Erfurter Pilotstudie zeigen, ist der Test als kriteriumsvalide anzusehen, da die Testergebnisse hoch mit einem zusätzlich eingesetzten Multiple-Choice-Test korrelieren. Zur Gewährleistung der Auswertungsobjektivität wurden zehn Prozent aller Triaden von geschulten Kodierern doppelt ausgewertet und die Beobachterübereinstimmung nach Cohen’s Kappa berechnet. Die gemittelten Kappa-Werte über alle fünfzehn Triaden liegen in einem Bereich von $0{,}65 < \kappa < 0{,}95$. Diese Werte können als überaus zufriedenstellend angesehen werden und belegen damit auch die Reliabilität des Instrumentariums.

Die Bearbeitungsquote des *Schülerfragebogens*[3] betrug 99,1% im Pre-Test bzw. 98,1% im Posttest. Der Schülerfragebogen wurde sowohl zur Erfassung abhängiger Variablen als auch zur Kontrolle moderierender Variablen eingesetzt. Hinsichtlich der abhängigen Variablen enthält er Skalen zur motivationalen Orientierung und dem Chemieinteresse. Weitere Skalen, die das kindperzipierte Instruktionsverhalten der Lehrer sowie elterliche Erziehungsverhalten erfassen, dienen der Kontrolle etwaiger moderierender Effekte auf die Interventionen. Die internen Konsistenzen der Skalen des Schülerfragebogens sind ebenfalls zufriedenstellend ($.57 < \alpha < .90$).

Eine weitere Kontrolle von potentiellen Störvariablen erfolgte mit Hilfe des *Elternfragebogens*[3]. Die Skalen des Fragebogens zielten mit Hinblick auf die Hausaufgabenstudie sowohl auf unmittelbar treatmentmoderierende Konstrukte wie zum Beispiel das kindperzipierte elterliche Erziehungsverhalten, Angaben zur Qualität und Intensität der Hausaufgabenbetreuung sowie der chemisch-fachlichen Vorbildung der Eltern, diente aber darüber hinaus im Rahmen der Evaluation der Unterrichtsstudie auch als Informationsquelle über den sozio-ökonomischen Status der Familien.

Im *Lehrerfragebogen*, den die beteiligten Lehrpersonen unmittelbar im Anschluss an die Intervention erhielten, wurden zum einen Fragen bezüglich einer allgemeinen Einschätzung der Intervention, der angewandten Arbeitsweisen und Lernstrategien gestellt. Zum anderen sollten die Lehrpersonen für jeden einzelnen Schüler Fragen zu begabungsrelevanten Verhaltensmerkmalen und zu wahrgenommenen motivationalen Veränderungen der Schüler beantworten.

Prozessdaten

Unmittelbar im Anschluss an jede Unterrichtsstunde wurde ein treatmentbegleitender Kurzfragebogen eingesetzt, in dem die Schüler gebeten wurden, das perzipierte Lehrerverhalten, ihr aktualisiertes Interesse sowie, mittels einer Adjektivliste, ihre affektiven Befindlichkeiten und Emotionen während des Unterrichts anzugeben. Ein analoger Kurzfragebogen wurde, bei entsprechender Umformulierung der Items, ebenfalls im Anschluss an jede Hausaufgabensituation eingesetzt.

Zusätzlich zu den eingesetzten Stundenfragebögen wurden alle Unterrichtsstunden in der Interventionsgruppe und der Kontrollgruppe videographiert und volltranskribiert. Die Videos dienten zum einen als Treatmentcheck, um zu dokumentieren, ob und inwieweit die Lehrer den präskriptiven Instruktionen gefolgt sind. Darüber hinaus dient die Videographierung aber auch der kategoriengeleiteten Analyse des Unterrichts mit Hinblick auf einen Vergleich der Experimentierphasen von Interventions- und Kontrollgruppe als auch einer differenzierten Analyse der Kleingruppenarbeitsphasen. Eine Darstellung des entwickelten Kategoriensystems findet sich in Kapitel 7.

[3] Schüler-, Eltern- und Lehrerfragebogen wurden von der Arbeitseinheit pädagogische Psychologie der Universität Bielefeld, Prof. Dr. Elke Wild, zusammengestellt und konstruiert. Eine ausführliche Skalendokumentation ist unter Wild, Exeler, und Gerber (2003) nachzulesen.

Aus technischer Sicht liegen seit der TIMSS-Videostudie gute Standards für eine erfolgreiche Videographierung von Unterrichtssequenzen vor (Stigler u. Fernandez 1995, Stigler et al. 1999). Im Gegensatz zur TIMSS-Videostudie wurde in dieser Untersuchung jedoch nicht mit einer Kamera, sondern analog zur Kieler BIQUA-Studie (Seidel, Dalehefte u. Meyer 2001) mit zwei Kameras gearbeitet. Eine Kamera wurde in beiden Unterrichtsituationen starr fixiert und diente zur Aufnahme der Totalen (T-Kamera). Die zweite Kamera wurde während des Unterrichts manuell gesteuert (Aktions-/ A-Kamera) und erfasste in der Kontrollgruppe den jeweils aktiv sprechenden Schüler bzw. den Lehrer. Die folgende Abbildung gibt einen Überblick über die Kamerapositionierung einer Interventionsklasse:

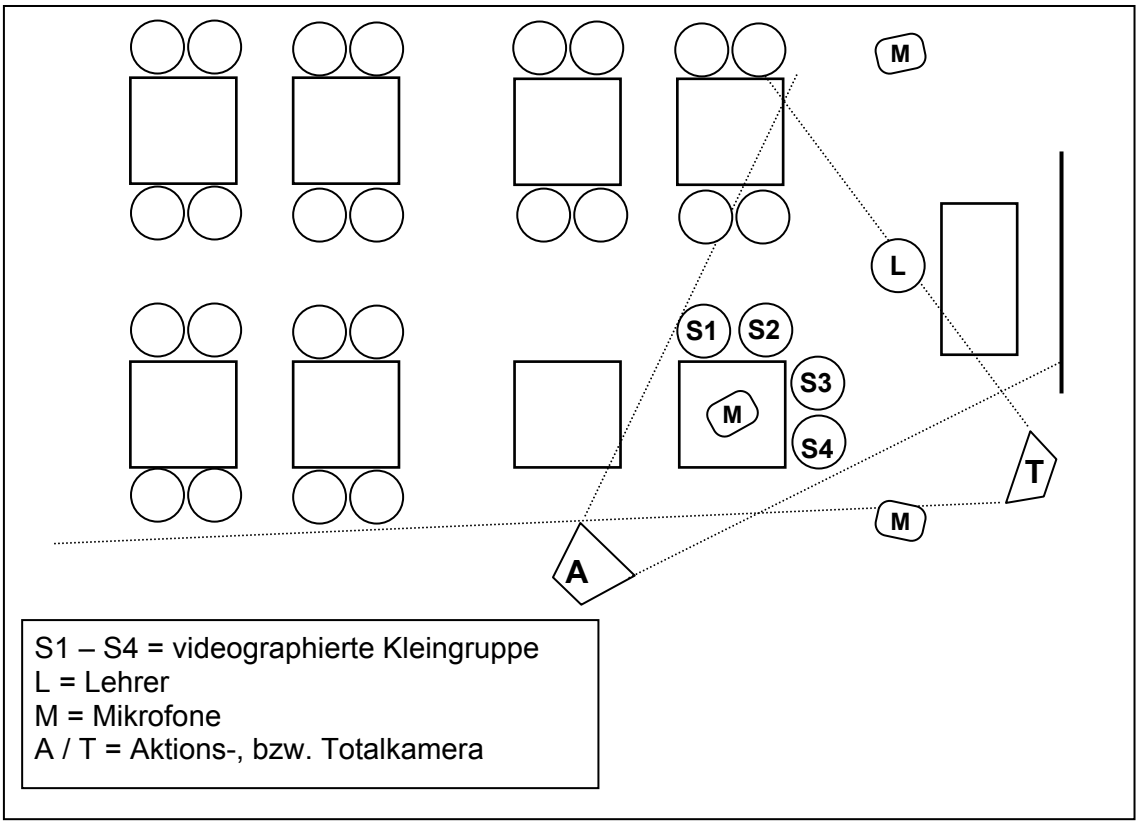


Abb. 4.8: Kamerapositionen während der Unterrichtsintervention

In den unterrichtlich intervenierten Klassen wurde während der Kleingruppenarbeitsphase die Aktionskamera gänzlich auf eine ausgewählte Kleingruppe gerichtet, die längsschnittlich über die gesamte Intervention videographiert wurde. Auswahlkriterium, welche Gruppe hierzu herangezogen wurde, war die Bereitschaft der jeweiligen Eltern der Kleingruppenschüler auch einer häuslichen Videographierung bei der Hausaufgabensituation zuzustimmen, so dass von einer möglichst großen Anzahl an Schülern häusliche und schulische Videodokumentationen erstellt werden konnten.

Nach Abschluss der Videoaufnahmen wurden die Filme im Labor digitalisiert, in ein MPG-Format komprimiert und anschließend auf CD gesichert. Bei insgesamt acht Klassen über sechs Unterrichtsstunden (exklusive der Vorstunde) ergaben sich somit insgesamt 48 videographierte Unterrichtssituationen, jeweils 24 aus beiden Unterrichtsexperimentalgruppen.

4.3.3 Arbeitsplan

Der Erhebungszeitraum der Untersuchung erfolgte von Januar bis Dezember 2001. Der Intervention lief ein Lehrertraining voraus, in dem den Lehrern sowohl das didaktisch-methodische Konzept der Unterrichtsintervention als auch die ausgearbeiteten Stundenverlaufspläne vorgestellt wurden und die notwenigen Verhaltensmaßnahmen zur Wahrung der Durchführungsobjektivität (time on task, identische Tafelbilder in beiden Gruppen etc.) vermittelt wurden. Um mögliche Versuchsleitereffekte zu vermeiden wurde keine der beiden Unterrichtsvarianten als die überlegenere dargestellt.

Aus organisatorischen Gründen verlief die Datenerhebung in vier Wellen, die sich wie folgt auf die Untersuchungsklassen verteilten (U^+= Unterrichtsintervention, H^+= Hausaufgabenintervention):

	1. Welle		2. Welle		3. Welle		4. Welle	
Klasse	KG73	KG74	FH72	FH73	KG71	KG72	FH71	FH74
Treatment	U^+ / H^+	U^- / H^-	U^- / H^+	U^+ / H^-	U^+ / H^-	U^- / H^+	U^+ / H^+	U^- / H^-

Tabelle 4.3: Verteilung der Klassen auf die Treatments und Erhebungszeiträume

Vor der Realisierung der Unterrichtsintervention wurde in allen Klassen eine Vorstunde aus dem Themenbereich „Verbrennungsreaktionen“ durchgeführt. Diese Vorstunde diente vor allem dazu, alle Beteiligten - Lehrer und Schüler - sowohl mit der spezifischen Methodik der Unterrichtsintervention und den Interaktionsboxen als auch mit der durch die Videoaufnahmen bedingte Bild- und Tontechnik vertraut zu machen, um während der Haupterhebung interventionsevozierte Innovationseffekte zu minimieren.

Für die daran anschließende Haupterhebung stellt sich der Arbeitsplan wie folgt dar:

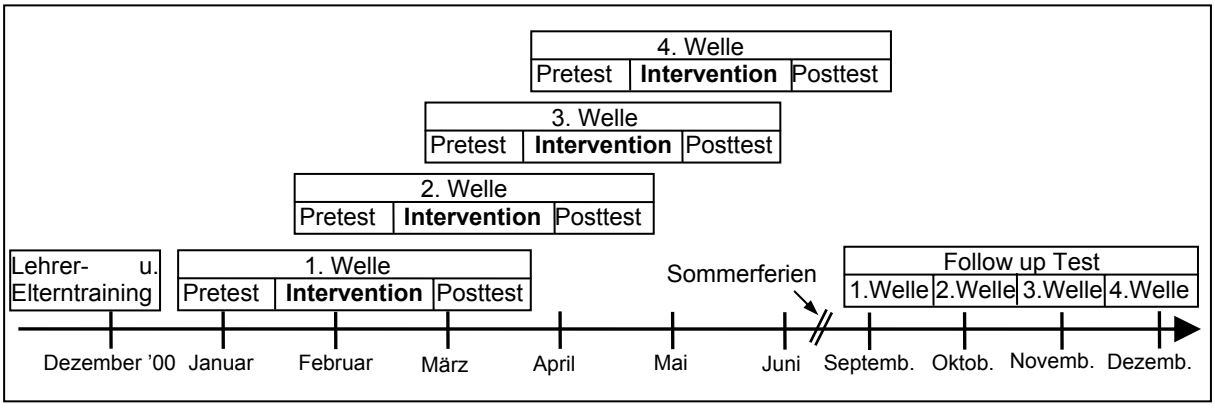


Abb. 4.9: Arbeitsplan während der Intervention

Wie der Abbildung zu entnehmen ist, wurden die Pre-Tests einen Monat vor der Intervention durchgeführt, um Auswirkungen des Tests auf die Intervention zu verhindern. Die Post-Tests fanden dagegen unmittelbar im Anschluss an die Intervention statt, in der Regel in der Folgestunde. Die sechs Monate nach Abschluss der Intervention durchgeführten Follow-up-Tests dienten vor allem zur Erhebung der Nachhaltigkeit der Intervention. Nur durch diese Maßnahme lässt sich herausfinden, ob durch die Intervention auch so genannte längerfristige Kompetenzen, gekennzeich-

net durch die Fähigkeit für einen *zeitlichen Transfer* („L-Kompetenzen“, Hager u. Hasselhorn, 2000), vermittelt werden konnten.

5. Darstellung der Unterrichtsreihe

5.1 Schülerexperimente mit Interaktionsboxen

Kern der schulbasierten Intervention ist eine spezielle Form der Gruppenarbeit, die so genante Gruppenrecherche oder „Group Investigation“ (Sharan 1994, s.a. Kp.3.2.3), die eine methodische Variation des im Chemieunterricht tradierten Schülerexperiments darstellt. Die Gruppenarbeitsphase ist in den Regelunterricht implementiert und stellt sich in einem groben Verlaufsschema wie folgt dar:

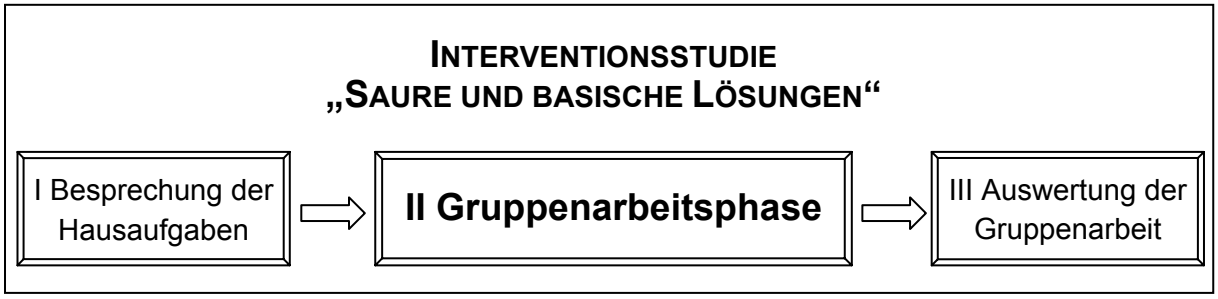


Abb. 5.1: Verlaufsschema der Unterrichtsintervention

Damit folgt die Intervention in ihrer konzeptionellen Umsetzung einer Forderung von Renkl (1996), wonach innovative Modelle situierten Lernens kombinierbar mit konventionellen Unterrichtsformen sein sollten.

Die Kleingruppen erhalten speziell konzipierte Interaktionsboxen, mit deren Hilfe eine diskursiv-konstruierende Erarbeitung des Problemraumes ermöglicht wird. Ziel der Gruppenarbeit ist damit die gemeinsame Konzeption eines Problemlösungsweges. Neben dem Arbeitsauftrag enthalten die Interaktionsboxen zahlreiche Experimentiergegenstände sowie instruktionale Hilfen zur Problemlösung. Durch die zu leistende Auswahl und Kombination der Experimentiergegenstände müssen die Schüler untereinander in eine sachbezogene Diskussion treten, die sich im Gegensatz zum konventionellen Unterricht dadurch auszeichnet, dass keiner der Gesprächsteilnehmer einen Expertenstatus besitzt, wodurch eine gleichberechtigte, symmetrische Kommunikation ermöglicht werden soll.

Im Gegensatz zu der im traditionellen Chemieunterricht zu beobachtenden Praxis wird in dieser Unterrichtsform nicht ein rein handwerkliches „Nachkochen“ einer vom Lehrer vorbereiteten Experimentiervorschrift erwartet. Sofern es der Stundeninhalt zulässt, verfolgt die Aufgabenstellung eine explorative Zielsetzung. Leutner und Fischer (2002) wie auch Seidel et al. (2002) verweisen darauf, dass die traditionelle Aufgabenstellung von Schülerexperimenten im naturwissenschaftlichen Unterricht durch die Vorgabe von situationalen und operationalen Zielen gekennzeichnet ist, das heißt, es wird sowohl eine Vorgabe hinsichtlich des zu erreichenden Ergebnisses als auch darüber, wie dieses Ergebnis zu erreichen ist, gemacht. Leutner und Fischer (2002) führen kritisch an, dass ein elaboriertes Verständnis der operationalen Vorgaben keine notwendige Voraussetzung für ein erwartungskonformes Erreichen einer situationalen Zielsetzung ist. Demgegenüber steht im Falle einer explorativen Zielsetzung die Entwicklung von Hypothesen, das Entwickeln und Durchführen eines

Experiments zur Hypothesenprüfung sowie die hypothesenreflektierende Auswertung des Experiments im Vordergrund. Aus der Motivationsforschung ist aber bekannt, dass, sofern den Lernenden das notwendige Vorwissen zur Analyse des Problemraums fehlt, bei derart offenen Aufgabenstellungen negative motivationale Effekte zu erwarten sind (Fischer & Horstendahl 1997).

Weinert (1997) verweist in diesem Zusammenhang darauf, dass Lehr- Lernprozesse suboptimal verlaufen, wenn sie ausschließlich auf rezeptives *oder* selbstorganisiertes Lernen setzen. Gesucht wird daher eine Balance zwischen selbstständiger Exploration und instruktionaler Unterstützung. In dieser Überlegung gründet sich die Funktion der instruktionalen Hilfen, die so konzipiert wurden, dass sie dem Schüler eine sinnvolle Vorstrukturierung des Problemraums bieten.

Darüber hinaus erlaubt die Materialien- und Chemikalienausstattung der Experimentierboxen die praktische Umsetzung verschiedener Lösungsansätze, explizit auch bekannter Schülervorstellungen, die im Experiment repräsentiert und falsifiziert werden können. Fehler sind in dieser Experimentierphase kein auszumerzendes Übel, sondern ein intendiertes Zwischenziel, welches dazu dient, über die ausgelöste Irritation an der Alltagsvorstellung neue Denkwege zu erschließen und ein „conceptual growth“ zu ermöglichen. Viele der im Kontext der PISA und TIMS-Studie publizierten Arbeiten gehen dieses Defizit (z. B. Müller & Nieswand 1999) an und weisen gerade auf eine den deutschen Lehrenden eigene „Fehlerphobie“ hin, die das wahre Entwicklungspotenzial von Fehlern verkennt: „Fehlerverbote sind Entwicklungsverbote“, so Kahl (2002). Gegenbeispiel zu der kritisierten Praxis beschreibt Schümers (1998) Analyse des japanischen Mathematikunterrichts, der in vielen Parametern den zentralen Prinzipien des Unterrichtens mit Interaktionsboxen entspricht.

Bei der Bemühung um eine Verbesserung der schülerexperimentellen Praxis steht der hier vorgestellte Weg keineswegs allein da: So stellt Lehmann (2002) mit der Leitkarten (Leika-) Methode eine Form des autonomen Wissenserwerbs vor, in dem die Lernenden ebenfalls funktional differenzierte Informationskarten (Leit-, Geräte-, Lexikonkarte) erhalten, wenngleich die Handlungsziele, ähnlich dem eines Lernzirkels, häufig operational formuliert sind.

Eine größere Parallele weisen die in dieser Studie entwickelten Aufgaben mit den aus dem anglo-amerikanischen Bereich stammenden „egg race“-Aufgaben auf (Borrows 1989), in denen mit Hilfe einer Materialauswahl eine bestimmte Aufgabe gelöst werden muss.[4] Inzwischen liegen weitere Arbeiten vor, die diese Idee für den Chemieunterricht adaptiert haben (Bouma & Mast 1988, Scharf & Gärtner 1994 sowie Gärtner & Scharf 1997, 2001, Gärtner & von Borstel 2003).

Die im Rahmen der Interaktionsboxen entwickelten Aufgaben weichen jedoch insofern von der ursprünglichen Tradition der egg-race-Aufgaben ab, als dass im Sinne einer explorativen Zielsetzung ein stärker hypothesengeleitetes Arbeiten ermöglicht

[4] Zur Namensgebung: Bei der ersten „egg-race“-Aufgabe Mitte der 70er Jahre wurde gefordert, mit Hilfe eines Gummibandes ein Gefährt zu bauen, welches ein Hühnerei soweit wie möglich transportiert.

wird. Darüber hinaus wird in dieser Arbeit das Konzept der Interaktionsboxen konsistent über eine ganze Unterrichtsreihe entwickelt, wohingegen egg-race-Aufgaben in aller Regel nicht mehr als eine Unterrichtsstunde umfassen.

5.2 Die Unterrichtsreihe „saure und basische Lösungen“

Die Richtlinien des Landes Nordrhein-Westfahlen (MSWF NRW 1993) weisen die Einführung der Begriffe „saure Lösung“ und „saurer Regen“ innerhalb der Unterrichtsreihe „Luftverschmutzung und Maßnahmen zur Luftreinhaltung“ als obligatorisches Thema für die Jahrgangsstufe sieben aus. Die gleichzeitige Einführung des Begriffs der „basischen Lösung“ ist dahingegen nicht verbindlich vorgeschrieben. Aktuelle, zur Einführung für das Land NRW genehmigte Gymnasialschulbücher belegen allerdings die Möglichkeit, das Thema „saure und basische Lösungen“ im fortgeschrittenen Anfangsunterricht der Jahrgangsstufe sieben zu unterrichten (Asselborn, Jäckel, Risch 2001, Arnold & Franik 2000, Eisner et al. 1994, Tausch, von Wachtendonk 1996). Die Durchsicht der Schulbuchliteratur verdeutlicht die Möglichkeit, das Thema gleichermaßen innerhalb der Unterrichtsreihe „Luft und Verbrennung“ wie auch in der Unterrichtsreihe „Wasser als Oxid“ zu behandeln. Auf der Inhaltsebene umfassen alle in der Literatur vorgeschlagenen Unterrichtsgänge die Einführung der Aggregationsklassen der sauren und basischen Lösungen, deren Unterscheidung mit Hilfe von verschiedenen Indikatoren, den pH-Wert zur genaueren Kennzeichnung saurer und basischer Lösungen sowie das Verhalten von Nichtmetall- und Metalloxiden gegenüber Wasser.

Wenngleich die Schüler bereits durch die modellhafte Deutung der Aggregatzustände über ein erstes Teilchenkonzept verfügen, so bedarf ein vertiefendes Verständnis der Säure-Base-Theorie der Kenntnis von Ionen und Salzen sowie höheren mathematischen Voraussetzungen (Logarithmus) zur Behandlung des pH-Werts. Alle diese Lernvoraussetzungen sind am Ende der Jahrgangstufe sieben nicht gegeben, weshalb die Behandlung des Themas nur phänomenologisch, auf der Kontinuumsebene möglich ist.

Die unterrichtliche Aufbereitung des Themas kann unter sehr unterschiedlichen Aspekten erfolgen. Abbildung 5.2 gibt einen Überblick, über fachsystematische, alltagsweltliche und historische Aspekte der Säure-Base-Thematik:

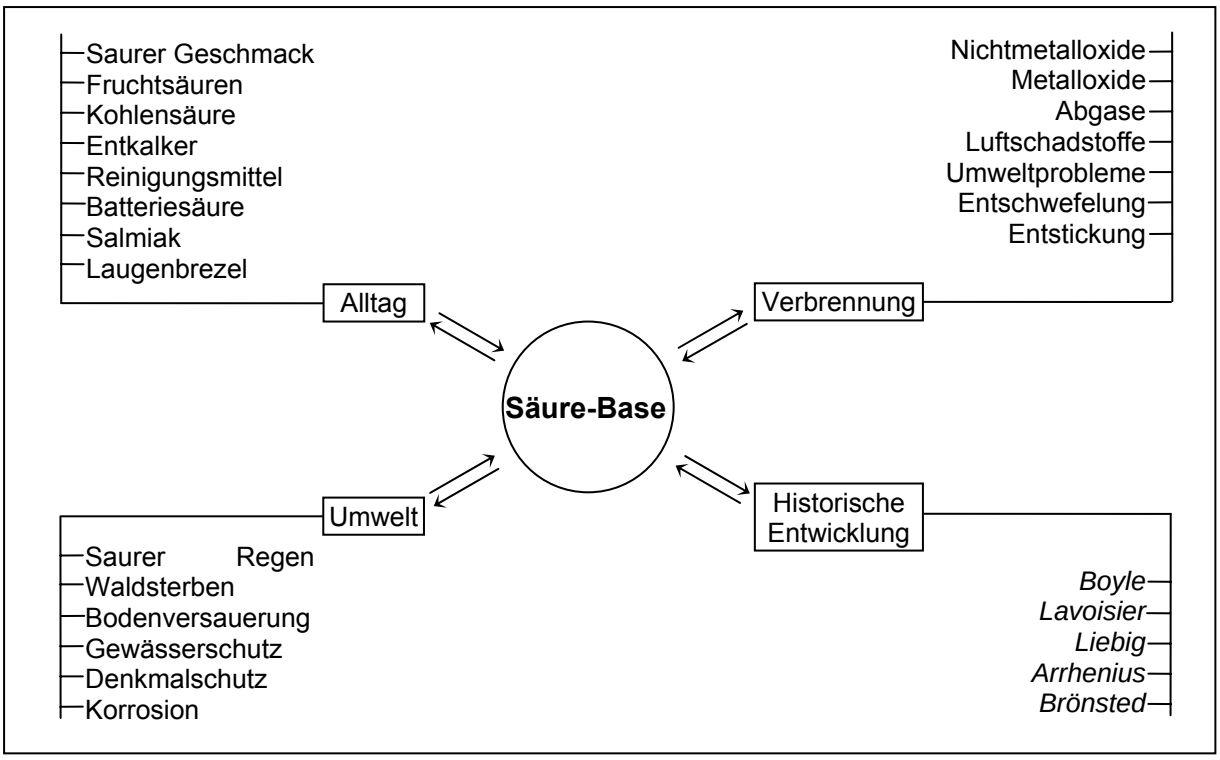


Abb. 5.2: Wege zum Säure-Base Begriff (Häusler, Pavenzinger1992)

Aufgrund der Vielschichtigkeit und des hohen Alltagsbezugs des Themas ist die Anzahl der bereichsspezifischen Publikationen, in denen Vorschläge für Unterrichtsgänge zum Themenbereich Säure-Base vorgelegt werden, enorm groß. Es bedarf daher der Betonung, dass bei der inhaltlichen Darstellung der Unterrichtsreihe ein Verweis auf die fachdidaktische Literaturlage nur selektiv und ohne Anspruch auf Vollständigkeit erfolgen kann. Darüber hinaus liegt die Schwerpunktsetzung dieser Arbeit auf der Evaluation der Unterrichts*methode* „Gruppenrecherche mit Interaktionsboxen“; ein Innovationsanspruch bezüglich der *inhaltlichen* Gestaltung der einzelnen Unterrichtsstunden wird damit ausdrücklich nicht erhoben. Vielmehr wird der Argumentation von Gramm (2000) gefolgt und die zu verzeichnende Dominanz der fachdidaktischen Publikationen kritisch bewertet, die empirisch nicht gesicherte Experimente und „Musterstunden“ von immer neuen Variationen bekannter Inhaltsbereiche beinhalten.

Auf der inhaltlichen Ebene gestaltet sich die die Unterrichtsreihe „saure und basische Lösungen“ wie folgt:

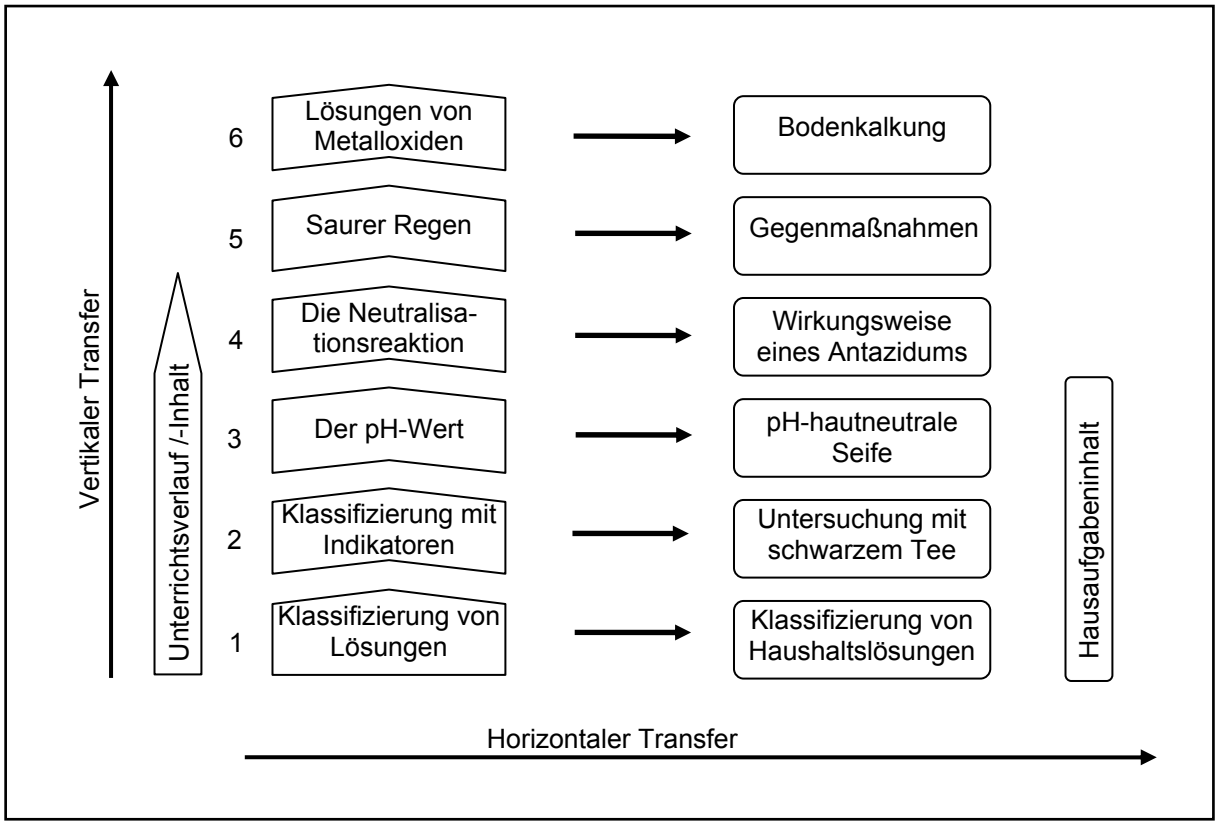


Abb. 5.3: Inhaltlicher Verlauf der Unterrichtsintervention

Die Unterrichtsreihe umfasst sechs Unterrichtsstunden, wobei für jede Stunde eine korrespondierende Hausaufgabenstellung entwickelt wurde.

Ausgangspunkt der Unterrichtsreihe ist eine Auswahl von zwölf Lösungen, die weitestgehend dem häuslichen Bereich entnommen wurden und die unter fachsystematischen Gesichtpunkten mit Hilfe von Indikatoren (Rotkohlsaft und Bromthymolblau) als saure, neutrale und basische Lösungen kategorisiert werden können (vgl. Hirsch & Horlacher 1987, Häusler & Worofka 1987, Kogelnik 1982, 1983). Zur weiteren Differenzierung saurer und basischer Lösungen wird in der dritten Stunde der pH-Wert eingeführt (vgl. Rampf 1987). Hieran schließt in der vierte Stunde eine phänomenologische Einführung der Neutralisation an. In den letzten beiden Stunden der Unterrichtsreihe wird am Beispiel des sauren Regens sowie der forstwirtschaftlichen Gegenmaßnahmen zum sauren Regen das Verhalten von wässrigen Metall- und Nichtmetalloxidlösungen erörtert (vgl. Dämmgen & Frühauf 1985a, 1985b, Goss & Eddleton 2003)

Wie Abb. 5.3 zu entnehmen ist, sind die Stunden so aufeinander abgestimmt, dass über den zeitlichen Verlauf der Unterrichtsreihe ein vertikaler Transfer zum Tragen kommt, während die Hausaufgabenstellung bezogen auf das Lernziel der jeweiligen Unterrichtsstunde einen horizontalen Transfer erfordert. Ein vertikaler Transfer, so Hasselhorn und Mähler (2000), liegt vor, wenn eine zu erwerbende Fertigkeit direkt zum Erwerb einer übergeordneten Fertigkeit beiträgt, während ein horizontaler oder auch *lateraler* Transfer Situationen gleicher Komplexität bezeichnet.

Allen Unterrichtsstunden, sowohl der Interventionsgruppe als auch der Kontrollgruppe, liegt ein einheitliches Artikulationsschema zur Stundenphasierung zu Grunde (Abb. 5.4). Das Schema ist bewusst nach pragmatischen und nicht lernpsychologischen Gesichtspunkten artikuliert und weicht damit von problemorientierten Unterrichtsverfahren (z. B. Schmidkunz, Lindemann 1995) ab, da die zentrale Funktion der eingesetzten Verlaufspläne hier darin liegt, die unterrichtenden Lehrer über die Gemeinsamkeiten und Unterschiede zwischen Interventions- und Kontrollgruppe zu informieren. Dennoch wurde, mit Ausnahme der ersten Stunde, eine problemorientierte Stundenkonzeption realisiert.

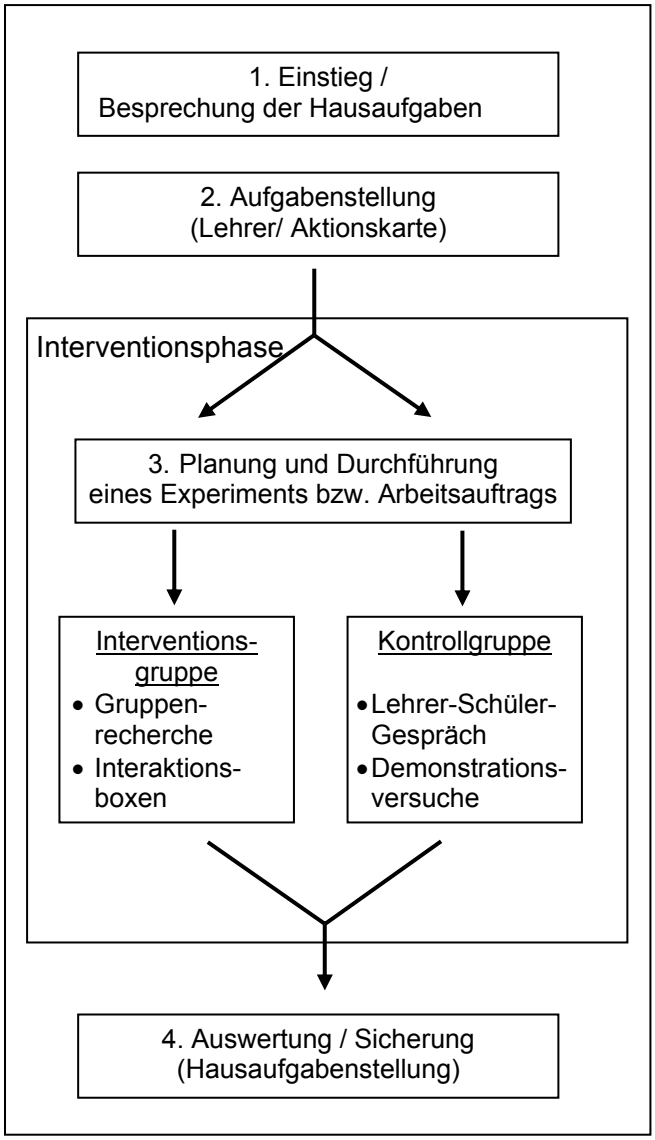


Abb. 5.4: Stundenphasierung der Unterrichtsreihe

Das Schema verdeutlicht, dass hinsichtlich der ersten beiden wie auch der letzten Phase, die Interventions- und Kontrollgruppe identisch verlaufen. Lediglich während der Planungs- und Experimentierphase (knapp 50% der Unterrichtszeit) differiert der methodische Ablauf der beiden Gruppen.

Sämtliche Arbeitsmaterialien und instruktionalen Hilfen der Interventionsgruppe liegen auch der Kontrollgruppe vor. Einzig die Größe der Medien und Experimentiermaterialien sind den Bedürfnissen im Plenum angepasst. Die Experimente werden in der Kontrollgruppe als Lehrer- oder Schülerdemonstrationsversuch durchgeführt.

Im Folgenden wird der Stundenverlauf der intervenierten Unterrichtsreihe ausführlich dargestellt. Auf eine separate Darstellung der Experimentierphase in der Kontrollgruppe wird hierbei verzichtet.

5.2.1. Die Vorstunde

Vor Durchführung der Unterrichtsreihe „saure und basische Lösungen“ wurde eine Vorstunde aus dem Themenbereich „Luft und Verbrennung – Oxidation“ vorweggeschaltet, um Innovationseffekte bei der Datenerhebung zu vermeiden (siehe 5.3.3). Aus diesem Grund wurden alle bei der Durchführung der Intervention notwendigen videotechnischen und unterrichtsmethodischen Anforderungen mit Ausnahme des treatmentbegleitenden Testinstrumentariums auch bereits in dieser Stunde realisiert.

Thema der Stunde

Die Funktion des Dochts einer Petroleumlampe

Einstieg

Zum Stundenbeginn präsentiert der Lehrer den Schülern eine brennende Petroleumlampe und fordert die Schüler auf, die Stoffeigenschaften von Petroleum[5] zu benennen. Die Schüler geben hierauf den Aggregatzustand bei Raumtemperatur mit „flüssig“ sowie als weitere – evidente – Stoffeigenschaft die *Brennbarkeit* an. (Darüber hinaus sind in der Praxis häufig auch die aus dem alltagsweltlichen Umgang mit aromatisierten Lampenöl induzierten Nennungen „farbig“ und „duftend“ festzustellen.)

Zur Überprüfung der angegebenen Stoffeigenschaft „Brennbarkeit“ wird in einem Demonstrationsversuch etwas Petroleum in eine Abdampfschale gegeben, und es wird versucht, dieses zu entzünden. Entgegen der Vermutung der Schüler ist dies nicht möglich, stattdessen gelingt es sogar, das brennende Streichholz im Petroleum zu löschen. Gemäß gängigen Klassifikationsschemata kann dieses Experiment funktional als „Problemexperiment“ (Pfeifer, Häusler, Lutz 1997) bzw. als „Experiment zum Wecken einer Fragehaltung“ (Hermanns, Wambach 1984) spezifiziert werden, da es die Frage aufwirft, „warum brennt das Petroleum in der Lampe, nicht aber in der Abdampfschale?“, wodurch zur eigentlichen Problemstellung übergeleitet wird.

Problemstellung

Schüler versuchen die obige Frage sehr häufig zu beantworten, indem sie darauf hinweisen, dass in der Versuchsanordnung ein Docht fehlt und unterfüttern diese Behauptung mit zwei Erklärungsansätzen:

1. Der Docht selbst verbrennt
2. Das Petroleum benötigt den Docht, um brennen zu können

Die erste Vermutung ist leicht zu entkräften, indem ein ca. 10 cm langer, ungewachster Kerzendocht entzündet wird. Es ist eindeutig ersichtlich, dass der reine Kerzendocht zwar gut, aber ebenso schnell verbrennt, was zu einem raschen Erlischen der Petroleumlampe nach dem Abbrennen des Dochtes führen müsste.

[5] An dieser Stelle wird darauf verzichtet, Petroleum explizit als Stoffgemisch des Mitteldestillats (Siedebereich 150 – 250°C) der Rohöldestillation zu bezeichnen.

Interventionsphase

An dieser Stelle des Unterrichtsgangs werden die Interaktionsboxen ausgeteilt und die Schüler in die arbeitsgleiche Gruppenarbeit entlassen, verbunden mit der Aufgabe, mit Hilfe der Bestandteile der Box eine improvisierte, aber funktionsfähige Petroleumlampe zu bauen.

Bestandteile dieser Box sind:

- Teelichtunterschale
- Büroklammer
- Kupferdraht
- Gummiband
- Heftzwecke
- Kreidestück
- Lavastein
- Stück Alufolie
- Magnesiarinne
- Streichhölzer

Die Komponenten der Box wurden so ausgewählt, dass sowohl nützliche als auch nutzlose Gegenstände vorhanden sind. Da unter den Bestandteilen kein „klassischer“ Kerzendocht vorhanden ist, müssen die Schüler überlegen, welche Gegenstände sie für einen improvisierten Docht benutzen können. Hierzu eignet sich neben dem Stück Tafelkreide auch ein Blumensteinchen aus Vulkanstein und die Magnesiarinne.

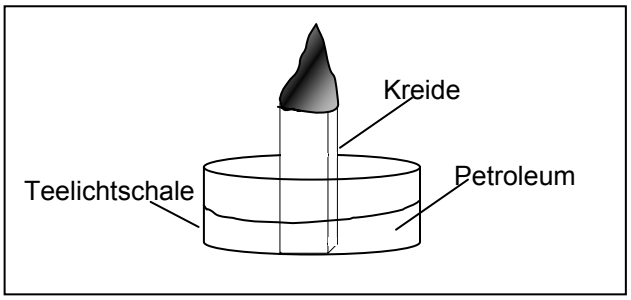


Abb. 5.5: Improvisierte Petroleumlampe

Ein häufig realisierter Versuchsaufbau ist in der nebenstehenden Abbildung dargestellt.

Als Zusatzaufgabe wird gefordert, die Größe der Flamme zu regulieren, was sich durch Umwickeln der Kreide mit der Aluminiumfolie erwirken lässt (vgl. Blume et al. 1994). Eine nähere Untersuchung aller als Dochtersatz geeigneter Gegenstände zeigt, dass der Docht selbst nicht brennbar sein, wohl aber über eine poröse, saugfähige Oberfläche verfügen muss.

Auswertung

In der abschließenden Auswertungsphase wird die Frage aufgeworfen, welche Funktion die beobachtete, poröse Beschaffenheit aller geeigneten Dochtmaterialien für die Entzündung des Petroleums hat. Da die Interpretation dieses Aspektes erfahrungsgemäß als sehr anspruchsvoll eingestuft werden kann und das phänomenologisch beobachtete „Aufsaugen“ des Petroleums durch den Docht kein hinreichendes Indiz für ein daraus abzuleitendes Verdunsten des Petroleums für die Schüler darstellt, wird noch einmal der zum Stundenbeginn durchgeführte Versuch in modifizierter Fassung durchgeführt: Das Petroleum wird nun in der Abdampfschale für einige Minuten mit dem Bunsenbrenner erhitzt, so dass der Flammpunkt (60°C) überschritten wird. Nach Abschalten des Brenners wird nun wiederum langsam ein brennendes Streichholz an das Petroleum geführt, wobei es noch *vor* dem Kontakt des Streich-

holzes mit der flüssigen Phase des Petroleums zu einer Flammentwicklung kommt. Es ist nun ersichtlich, dass nicht die Flüssigkeit, sondern die Gasphase des Petroleums brennt und das Vorliegen einer hinreichend großen Konzentration Petroleums in der Gasphase eine Voraussetzung für ein Entzünden ist. Die Schüler können dann abschließend schlussfolgern, dass Materialien, die über eine poröse, saugfähige Oberfläche verfügen, sich deshalb gut als Dochtersatz eignen, weil diese Materialien ein rasches Verdampfen des Petroleums auch bei Raumtemperatur ermöglichen.

In der Hausaufgabenstellung sollen die Schüler an einem nahen Transfer den Unterrichtsinhalt reorganisieren und die Vorgänge beim Anzünden einer Kerze als eine Abfolge von Schmelzen des Wachses, Aufsteigen, Sieden und Verdampfen am Docht und Entzünden des Wachsdampfes beschreiben.

5.2.2 Die erste Unterrichtsstunde

Thema der Stunde

Klassifizierungsmöglichkeiten verschiedener Lösungen

Einstieg

Der Einstieg in die Unterrichtsreihe erfolgt in Form eines Rückgriffs auf die Unterrichtseinheit „Lösevorgänge“, um die notwendigen Fachbegriffe „Lösungen“, “Lösemittel“ sowie „Löslichkeit“ zu wiederholen. Diese Wiederholung erfolgt mit Hinblick auf die in der Unterrichtsreihe umzusetzende Sprachregelung, wonach generell von sauren und basischen *Lösungen* zu reden ist, nicht jedoch von Säuren und Basen, um die fachsprachlichen Fundamente für eine Erarbeitung der Arrheniustheorie in der Jahrgangsstufe zehn zu legen. Ebenso wird auf das beschreibende Adjektiv „alkalisch“ gänzlich zugunsten von „basisch“ verzichtet, um die verwirrende Begriffsvielfalt auf Seite der Basen einzugrenzen.

Aufgabenstellung

Der anschließende Arbeitsauftrag fordert die Schüler auf, eine Palette von Lösungen, die gleichermaßen der Lebenswelt der Schüler wie auch dem fachwissenschaftlichen Kontext entnommen sind, nach einem von ihnen zu entwickelnden, plausiblen Kategoriensystem zu gruppieren. Dieser Auftrag stellt keine Problemstellung im Sinne Dörners (1976) dar, da zwischen kognitivem Ausgangs- und Zielzustand keine Diskrepanz besteht und daher auch keine kognitive Barriere zum Erreichen des Zielzustands überwunden werden muss. Neben der Einübung kommunikativer Techniken zur Bildung eines Gruppenkonsenses sollen vielmehr die Schüler mit diesem Arbeitsauftrag ihre impliziten Alltagsvorstellungen zur Kategorisierung von Lösungen explizit artikulieren.

Interventionsphase

Die Schüler erhalten ein Tablett mit folgenden Lösungen (bzw. Reinstoff):

- Destilliertes Wasser
- Feinwaschmittellösung
- Fensterreiniger
- Kalkwasser
- Limonade
- Mineralwasser
- Natriumhydroxidlösung
- Rohrreinigerlösung
- Speiseessig
- WC-Reiniger
- Zitronensaft
- Zuckerwasser

Der Arbeitsauftrag, der gleichermaßen den Kleingruppen wie auch dem Plenum der Kontrollgruppe gestellt wird, lautet: „Teilt die zwölf Flüssigkeiten in Gruppen ein, so wie es euch sinnvoll erscheint. Gebt dann den Gruppen einen Namen, der die Gruppe treffend kennzeichnet.“ Kalkwasser und Natriumhydroxidlösungen werden hierbei vom Lehrer als Laborchemikalien vorgestellt.

In Anlehnung an die Ergebnisse von Geisler (1999) ist davon auszugehen, dass statt der im Chemieunterricht anzustrebenden Klassifizierung in saure, basische und neutrale Lösungen zu diesem Ausbildungszeitpunkt alltagsweltlich geprägte Konzepte dominieren. Derartige Ordnungsklassen zeigen häufig eine pragmatisch-funktionale Komponente und lassen ein Gruppierungsschema erkennen, das an alltäglichen Verwendungsmöglichkeiten orientiert ist. So werden von vielen Schülern die Gruppen „Reinigungsmittel“ und „Lebensmittel“ oder eine Zusammenfassung beider Gruppierungen zu „Haushaltsmitteln“ vorgeschlagen. Einer ähnlich alltagsweltlich geprägten Einteilung entspricht eine Unterscheidung der dichotomen Klassen „giftig vs. ungiftig“. Darüber hinaus versuchen aber bereits auch einzelne Schüler, welche die Kategorie „Säure“ einführen, eine an der Fachwissenschaft orientierte Systematik zu generieren. Diese Systematisierungsversuche scheitern in der Regel daran, dass die Schüler noch nicht über das Basekonzept verfügen und der Aggregationsklasse „Säure“ kein entsprechendes Analogon gegenüberstellen können. Dies stimmt mit dem in der Literatur beschriebenem Sachverhalt überein, wonach der Bekanntheitsgrad der Fachbegriffe „Säure“ und „saure Lösung“ sehr viel größer ist als „Base“ bzw. „basische Lösung“ (Nay, Glatzel 1978). Schüler, die die Kategorie „Säure“ verwenden, zeichnen sich zu diesem Unterrichtszeitpunkt daher häufig durch die Verwendung von Mischkonzepten aus, in denen der Stoffgruppe der Säuren anderen alltagsweltlichen Gruppierungen gegenübergestellt wird (z. B. Getränke – Säuren – Reinigungsmittel). Niedderer (1996) bezeichnet diese Mischkonzepte, in denen alltagsweltliche und fachwissenschaftliche Konzepte nebeneinander existieren, als kognitive Zwischenvorstellungen (*intermediate conceptions*). Die Existenz von Zwischenvorstellungen verdeutlicht sich ferner, wenn etwa einerseits die Kategorie „Säure“ verwendet wird, Säuren aber aus einer alltagsweltlichen Annahme allgemein als ätzend und daher gefährlich attribuiert werden (vgl. Hand 1989, Hand & Treagust 1988) und gleichermaßen saure und basische, aggressive Reinigungsmittel, wie WC- und Rohrreiniger - auf Grund der Warnhinweise - als Säure eingruppiert werden. Es ist ein konzeptioneller Grundgedanke der Unterrichtsreihe, dass derartige Zwischenvorstellungen mit Hilfe der Aufgabenstellung provoziert und artikuliert werden können, um sie im weitern Verlauf des Unterrichts thematisieren zu können.

Auswertung und Hausaufgabenstellung

Nachdem sich die Gruppen auf ein Ordnungsschema geeinigt haben, werden die Gruppenergebnisse im Plenum vorgestellt. Aufgabe der anschließenden Diskussion ist die verbindliche Einigung der Klasse auf ein gemeinsames System. Dadurch werden spätesten zu diesem Zeitpunkt durch Auflistung aller Gruppenvorschläge die oben erwähnten Zwischenvorstellungen einer breiteren Schülerschaft bekannt gemacht und zur Diskussion gestellt. Diese Zwischenvorstellungen werden von der Mehrheit der Schüler wegen der angeführten Widersprüche bzw. Unvollkommenheiten als nicht tragfähig angesehen, weshalb sich das Plenum in der Regel für rein alltagsweltlich terminierte Ordnungsschemata entscheidet, die von fachwissenschaftlichem Standpunkt zwar ebenso abzulehnen sind, sich jedoch intern als ausgesprochen konsistent erweisen.

Für den weiteren Verlauf der Unterrichtsreihe ist allein aus organisatorischen Gründen die Einigung auf ein beliebiges, gemeinsames System von Belang, damit eine einheitliche Kontrastierung gegenüber dem in der folgenden Stunde zu entwickelnden System gewährleistet ist. Aus den oben genannten Gründen gibt es seitens der Schüler allerdings eine klare Präferenz für das folgende Ordnungsschema:

Lebensmittel	Reinigungsmittel	Chemikalien
- Limonade - Zitronensaft - Mineralwasser - Speiseessig - Zuckerwasser	- WC-Reiniger - Feinwaschmittellösung - Fensterreiniger - Destilliertes Wasser - Rohrreinigerlösung	- Kalkwasser - Natriumhydroxidlösung

Tab. 5.1: Lebensweltliche Kategorisierung der eingesetzten Lösungen

Die abschließende Hausaufgabenstellung fordert die Schüler dazu auf, in ihrem Haushalt nach weiteren Lösungen zu suchen und sie in das gemeinsame Klassifikationsschema einzutragen. Bezogen auf die Progression der Unterrichtsreihe verfolgt die Hausaufgabenstellung das Ziel, den Pool der Lösungen zu erweitern, um sowohl in der folgenden Stunde als auch in der folgenden Hausaufgabenstellung das gewählte Kategoriensystem weiter diskutieren zu können.

5.2.3 Die zweite Unterrichtsstunde

Thema der Stunde

Unterscheidung saurer und basischer Lösungen mit Hilfe von Rotkohlsaft

Einstieg

Zum Stundeneinstieg stellen die Schüler ihre Hausaufgaben vor und erweitern das in der letzten Stunde erstellte Ordnungsschema um die in ihrem Haushalt vorgefundenen Lösungen.

Aufgabenstellung

Nachdem die Schüler darüber informiert wurden, dass in der heutigen Stunde mit Rotkohlsaft eine weitere Lösung zur Verfügung steht, erhalten sie den zweiteiligen Auftrag, mit Hilfe der zusätzlichen Lösung die Stringenz des Kategoriensystems der letzten Stunde zu überprüfen und als Ableitung aus den Experimenten eine weitere Gruppierungsmöglichkeit vorzuschlagen.

Interventionsphase

Zur Initiierung der geforderten Experimentierreihe müssen die Schüler Kenntnis von dem beigelegten Farbverlaufsschema von Rotkohlsaft nehmen und den Plan entwickeln, die zwölf Lösungen sukzessive mit Rotkohlsaft zu versetzen. Mit Hinblick auf den Erwerb wissenschaftlich kontrollierter Arbeitsweisen ist es von Bedeutung, dass die Schüler immer nur eine Lösung mit Rotkohlsaft versetzen. Weiterhin sollte das jeweils eingesetzte Volumen an Rotkohlsaft bei allen Mischungsversuchen vergleichbar sein.

Die Mischversuche zeigen bei allen zwölf Lösungen mit Rotkohlsaft eine unmittelbare, intensive Farbreaktion. Da es sowohl im neutralen pH-Bereich als auch im stark basischen Bereich zu einer um einige Minuten verzögerten Farbreaktion von violett nach blau bzw. im basischen Bereich von grün nach gelb kommen kann, ist der Box eine Farbabbildung beigelegt, der die Farbabfolge von Rotkohlsaft gegenüber anderen Lösungen zu entnehmen ist, ohne jedoch die entsprechenden Fachbegriffe vorwegzunehmen. Die Schüler stellen somit zunächst fest, dass sich die Lösungen auch hinsichtlich ihres Farbumschlages nach Zugabe von Rotkohlsaft einteilen lassen. Eine dementsprechend vorgenommene Gruppierung kommt zu folgendem Ergebnis:

rot	blau-violett	grün-gelb
- WC-Reiniger - Speiseessig - Limonade - Zitronensaft - (Mineralwasser)	- Destilliertes Wasser - Zuckerwasser - Feinwaschmittellösung - (Mineralwasser)	- Fensterreiniger - Natriumhydroxidlösung - Kalkwasser - Rohrreinigerlösung

Tab. 5.2: Farbreaktion der eingesetzten Lösungen mit Rotkohlsaft

Der Vergleich dieses Ordnungsschemas mit der alltagsweltlichen Klassifizierung der letzten Stunde zeigt, dass beide Systeme nicht miteinander kongruieren, da beispielsweise Lösungen der Gruppe „Lebensmittel“ sowohl eine rote wie auch blauviolette (Zuckerwasser) Färbung mit Rotkohlsaft hervorrufen können. Noch inkonsistenter verhält sich die Gruppe der Reinigungsmittel, deren Vertreter sich in allen drei Gruppen der obigen Einteilung wieder finden.

Wenngleich alltagsweltliche Klassifizierungen unter pragmatischen Gesichtspunkten ihre Berechtigung haben, so veranlasst die obige Farbgruppierung die Schüler aber dazu, über weitere Klassifikationsmöglichkeiten nachzudenken. Aus dem vorange-

gangen Chemieunterricht verfügen die Schüler zum einen über eine erste Vorstellung von einer chemischen Reaktion im Sinne einer Stoffumwandlung, wissen aber auch, dass sich Stoffe an ihren chemischen Eigenschaften erkennen lassen, so dass im vorliegenden Fall die gleiche Farbreaktion gegenüber Rotkohlsaft auf die gleiche zugrunde liegende Reaktion zurückgeführt werden kann. Das gemeinsame Kriterium aller Stoffe einer Kategorie ist somit nicht funktional, sondern durch eine chemische Eigenschaft begründet.

Zusatzinformation:

Das Gemüse „Rotkohl" trägt nur in Norddeutschland diesen Namen. In Süddeutschland kennt man es unter dem Namen „Blaukraut".
Während die süddeutsche Variante (*Blau*kraut) nur mit Wasser gekocht wird, gibt man bei der sauren, norddeutschen Art (*Rot*kohl) Essig hinzu....

Abb. 5.6: Instruktionale Hilfe zur zweiten Stunde

Um den Schülern einen Ansatzpunkt für ihre Überlegungen zur Benennung der Kategorien zu geben, ist der Box eine instruktionale Hilfe Abb. 5.6) beigelegt, die über die regional unterschiedliche Bezeichnung von Rotkohl bzw. Blaukraut informiert und dabei auf die unterschiedliche Zubereitungsart hinweist. Nutzen die Schüler diese Hilfe, so werden sie mit hoher Wahrscheinlichkeit Lösungen, die mit Rotkohlsaft eine rote Färbung ergeben als „saure Lösungen" bezeichnen, wenn auch die fachwissenschaftlich korrekte Bezeichnung nicht die zentrale Zielsetzung der Stunde darstellt.

Auswertung und Hausaufgabenstellung

Nachdem das auf die Farbumschlagsreaktionen basierende Kategoriensystem an der Tafel fixiert wurde, werden die Vorschläge zur näheren Charakterisierung der Kategorien diskutiert. Aus den genannten Gründen wird dabei von den Schülern in der Regel die Kategorie „saure Lösungen" vorgeschlagen. Die korrespondierende Bezeichnung „basische Lösungen", für die Stoffgruppe, die mit Rotkohlsaft eine grüngelbe Färbung hervorrufen, ist dahingegen in der Regel vom Lehrer als neuer Fachbegriff einzuführen. Sofern der Vorschlag nicht seitens der Schüler eingebracht wird, ist es an dieser Stelle nicht notwendig, die verbleibende dritte Gruppe als „neutrale Gruppe" zu spezifizieren, da diese Begrifflichkeit leichter mit Hilfe des pH-Werts in der folgenden Stunde eingeführt werden kann.

In der Hausaufgabenstellung sollen die Schüler Lösungen aus ihrem Haushalt mit schwarzem Tee untersuchen, ihre Beobachtungen notieren und deuten. Die Schüler lernen mit schwarzem Tee einen weiteren Indikator kennen, mit dessen Hilfe ebenfalls saure von basischen Lösungen unterschieden werden können und müssen so die im Unterricht erworbenen Fähigkeiten und Fertigkeiten im Klärungsgespräch mit ihren Eltern reorganisieren.

5.2.4 Die dritte Unterrichtsstunde

Thema der Stunde

Einführung des pH-Werts zur Charakterisierung von sauren und basischen Lösungen

Einstieg

Die Besprechung der Hausaufgaben fokussiert zum einen auf allgemeine Anforderungen beim Experimentieren. Da der Indikatorumschlag bei schwarzem Tee von gelb-orange (sauer) nach dunkelbraun (basisch) weniger eindrucksvoll und offenkundig erfolgt als beim Rotkohlsaft, werden die Schüler zum einen zum sorgfältigen Beobachten aufgefordert. Zum andern soll dadurch aber auch die stereotype Vorstellung vermieden werden, dass saure bzw. basische Lösungen immer auf die gleiche Art mit Indikatoren reagieren. Gefordert wird ferner eine planvolle Auswahl der zu untersuchende Lösungen in dem Sinne, dass *klare* Lösungen eingesetzt werden sollen, bei denen gemäß der vorhergegangenen Stunde eine begründete Vermutung hinsichtlich der Eingruppierung in saure/basische Lösungen besteht.

Abschließend wird in dieser Phase die Funktion von Rotkohlsaft und schwarzem Tee dahingehend abstrahiert, dass beide zur Unterscheidung von sauren und basischen Lösungen herangezogen werden können. Damit verbunden wird der Begriff Indikator eingeführt.

Aufgabenstellung

Thematischer Schwerpunkt der Stunde ist die Einführung des pH-Werts als eine abstrakte Größe, die angibt wie sauer oder basisch eine Lösung ist. Da für ein differenziertes Verständnis vom pH-Wert die mathematischen Voraussetzungen (Logarithmus) fehlen, kann zu diesem Zeitpunkt die pH-Skala nur phänomenologisch thematisiert werden. Der Fachbegriff „pH-Wert“ muss durch den Lehrer bzw. über die instruktionalen Hilfen der Interaktionsboxen dogmatisch-informativ eingeführt werden. Die zweiteilige Aufgabenstellung der Stunde erfordert von den Schülern eine Abstufung der bekannten Lösungen der ersten Stunde innerhalb der drei gebildeten Kategorien sowie eine Auflistung allgemeiner Eigenschaften zur Beschreibung von sauren und basischen Lösungen.

Interventionsphase

Zur Einführung des pH-Werts wurde eine Definition ausgewählt, die den pH-Wert als „Maß für den sauren Charakter einer Lösung“ beschreibt. Zur Vermeidung späterer Fehlvorstellungen wurde darauf Wert gelegt, auch bei umschreibenden Definitionen im Unterrichtsgespräch, den pH-Wert nicht - wie häufig in der Schulbuchliteratur – als Angabe darüber, „wie stark sauer oder alkalisch eine Lösung ist“ (Asselborn, Jäckel, Risch 2001) zu definieren, um irreführende Anklänge in Bezug auf den in der Sekundarstufe II einzuführenden Begriff der „Säurestärke“ zu vermeiden.

Im Verlauf der Gruppenarbeitsphase sollen die Schüler mit Hilfe der pH-Teststäbchen den pH-Wert der zwölf Lösungen messen und die Lösungen auf dieser

Basis untereinander abstufen. Die verhältnismäßig trivialen quantitativen Messungen dienen dazu, mit Hilfe der pH-Skala wissenschaftlich angemessene, allgemeine Eigenschaften von sauren und basischen Lösungen zusammenzutragen und alltagsweltliche Annahmen stärker zu differenzieren oder zu relativieren. Dabei sollen folgende Aspekte herausgearbeitet werden:

- Lösungen, deren pH-Wert kleiner ist als sieben, bezeichnet man als saure Lösungen.
- Lösungen, deren pH-Wert größer ist als sieben, bezeichnet man als basische Lösungen.
- Bei einem pH-Wert von sieben spricht man von einer neutralen Lösung.

Einer instruktionalen Hilfe, welche das Verpackungsetikett eines Rohrreinigers wiedergibt, ist zu entnehmen, dass vor der Gefahr schwerer Verätzungen gewarnt wird.

Aus der Gebrauchsanweisung eines Rohrreinigers:

Ätzend

Verursacht schwere Verätzungen. Unter Verschluss und für Kinder unzugänglich aufbewahren. Behälter dicht geschlossen halten. Bei Berührung mit den Augen sofort gründlich mit Wasser spülen und Arzt konsultieren. Beschmutzte Kleidung sofort ausziehen. Bei der Arbeit geeignete Schutzhandschuhe und Schutzbrille tragen.

Abb. 5.7: Instruktionale Hilfe zur dritten Stunde

Da von den Schülern die Rohrreinigerlösung mit einem pH-Wert von 13 gemessen wurde folgt hieraus, dass allgemein Lösungen mit einem besonders kleinen oder besonders großen pH-Wert ätzende Eigenschaften aufweisen. Damit wird eine häufig zu beobachtenden Schülervorstellung, wonach allein saure Lösungen (diese aber prinzipiell) als gefährlich bezeichnet werden, aufgegriffen und widerlegt, da die Interpretation der pH-Messungen eine Relativierung des eindimensionalen Bildes erfordert: Die Gefährlichkeit einer Lösung lässt sich nicht allein aus der Zugehörigkeit zu einer bestimmten Stoffklasse ableiten, da einerseits Lebensmittel wie Zitronensaft und Limonade einen sauren pH-Wert aufweisen, andererseits aber auch basische Lösungen mit einem hohem pH-Wert ätzend reagieren können. Dadurch soll auch eine adäquate Grundlage für ein Neutralisationskonzept gelegt werden, das nicht durch die prekonzeptuelle Annahme der Neutralisation als die „Zerstörung“ einer Säure geprägt wird (Hand & Treagust 1988).

Auswertung und Hausaufgabenstellung

An Hand einer Tabelle werden die innerhalb der Interventionsphase oben dargestellten Ergebnisse diskutiert und mit folgendem Merksatz gesichert: „Der pH-Wert gibt das Ausmaß des sauren bzw. basischen Charakters einer Lösung an. Je kleiner der pH-Wert desto saurer, je größer der pH-Wert desto basischer ist die Lösung. Lösungen mit einem pH-Wert von 7 bezeichnet man als neutrale Lösungen.“

Die gemessenen pH-Werte der ausgesuchten Lösungen betragen:

Lösung	pH-Wert	Lösung	pH-Wert
WC-Reiniger	1	Destilliertes Wasser	7
Speiseessig	2	Feinwaschmittellösung	7-8
Limonade	2	Fensterreiniger	9
Zitronensaft	3	Kalkwasser	11
Mineralwasser[6]	6	Natriumhydroxidlösung	12
Zuckerwasser	7	Rohrreinigerlösung	13

Tabelle 5.3: pH-Werte der eingesetzten Lösungen

In der Hausaufgabenstellung sollen die Schüler zunächst den pH-Wert von Handseife aus ihrem Haushalt messen. In der anschließenden, zweiteiligen Transferaufgabe erhalten die Schüler zunächst die Information über so genannte pH-hautneutrale Seifen, deren pH-Wert 5,6 beträgt. Mit Rückgriff auf die Ergebnisse der Unterrichtsstunde soll erklärt werden, welche Aussagen man mittels dieser Information über die menschliche Haut treffen kann und warum der Begriff pH-hautneutral missverständlich sein könnte.

5.2.5 Die vierte Unterrichtsstunde

Thema der Stunde

Neutralisation einer sauren Lösung mit einer basischen Lösung

Einstieg

Der Vergleich der von den Schülern gemessenen pH-Werte von Handseife reicht erfahrungsgemäß von leicht sauren Werten (um 5) bis hin zu deutlich basischen (um 11). Die weitere Besprechung der Hausaufgaben lässt deutlich werden, dass das normale Milieu der menschlichen Haut einen leicht sauren Charakter aufweist und die Hautverträglichkeit deutlich basischer Seifen aus diesem Grund kritisch zu bewerten ist. In diesem Zusammenhang wird der aus der Lebenswelt bekannte Begriff des „Säureschutzmantels“ herangezogen, um den Nutzen hautneutraler Seifen zu erörtern und - im Sinne der Zielsetzung der letzten Stunde - mit der desinfizierenden Funktion des Säureschutzmantels gegenüber Keimen und Krankheitserregern auch positive Eigenschaften von sauren Lösungen herauszustellen. An dieser Stelle kann seitens des Lehrers auch auf die Verwendung von Ascorbinsäure als Antioxidationsmittel zur Konservierung von Lebensmitteln hingewiesen werden.

[6] Ein saurer pH-Wert des Mineralwassers setzt einen hohen Kohlensäureanteil bei gleichzeitig niedriger Hydrogencarbonatkonzentration voraus. Der Einsatz eines derartigen Mineralwassers ist anzustreben, da sich dadurch eine weitere Gelegenheit bietet, eine alltagsvorstellungsbedingte, universelle Gefährlichkeit von sauren Lösungen kritisch zu thematisieren.

Aufgabenstellung

Zur Vorbereitung der Aufgabenstellung wird den Schülern eine weitere Lösung in Form von Salzsäure vorgestellt. Die Gruppen bekommen sodann den Auftrag, die saure Eigenschaft Salzsäure aufzuheben und eine neutrale Lösung herzustellen. Sprachlich wird an dieser Stelle bewusst der Begriff „Neutralisation“ vermieden, da dieser erst am Ende der Stunde im Zuge der Auswertung eingeführt werden kann. Neben den bekannten Materialien und Lösungen erhalten die Interaktionsboxen dieser Stunde noch Pasteurpipetten, um ein tropfengenaues Zugegeben der Lösungen zu ermöglichen.

Interventionsphase

Die Aufgabenstellung dieser Stunde hebt sich hinsichtlich der Komplexität deutlich von den Anforderungen der ersten Stunden ab, da zur Lösung des Arbeitsauftrags alle bislang gewonnen Erkenntnisse integriert werden müssen, wodurch sich auf der Planungsebene die Progression des vertikalen Transfers ausdrückt.

Zum planvollen Lösen der Aufgabe müssen die Schüler zunächst mit Hilfe der Teststäbchen den pH-Wert der Salzsäure (pH 2) bestimmen. Neben einer Übersicht der gemessenen pH-Werte der letzten Stunde enthalten die Interaktionsboxen einen Zahlenstrahl der pH-Skala (Abb. 5.8), mit dessen Hilfe verschiedene experimentelle Ansätze, die auf unterschiedlichen Schülervorstellungen zur Neutralisation fußen, realisiert werden können. Unter den Prekonzepten existieren sowohl solche, die deutliche Anklänge an Alltagsvorstellungen erkennen lassen als auch solche, die bereits eine arithmetische Systematik aufweisen. Die häufigsten Vorschläge lauten:

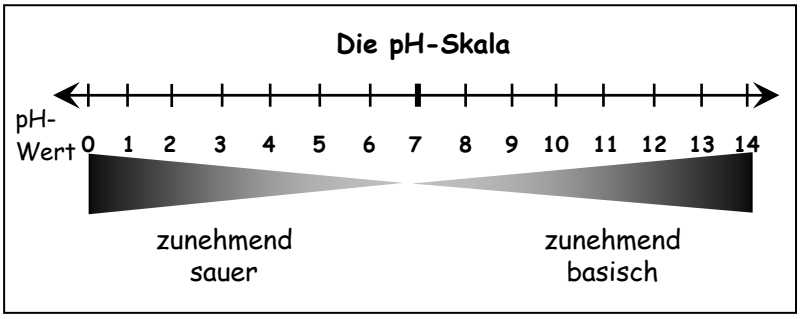


Abb. 5.8: Zahlenstrahl der pH-Skala

1. Verwendung einer neutralen Lösung
   Dieser Ansatz geht davon aus, dass man eine neutrale Lösung benötigt, um eine neutrale Lösung zu erhalten. In diesem Fall wird destilliertes Wasser als geeignetes Mittel zur Neutralisation vorgeschlagen. Um die bei hinreichend großer Verdünnung tatsächliche eintretende Annäherung an den neutralen pH-Wert zu verhindern, sind die Volumina der eingesetzten Salzsäure und verwendeten Glasgeräte so bemessen, dass das benötigte Volumen an destilliertem Wasser technisch nicht eingesetzt werden kann.

2. Verwendung einer süßen Lösung
   Dieser Ansatz ist klar lebensweltlich geprägt und geht von der Erfahrung aus, wonach „süß“ die gegenteilige Geschmackswahrnehmung zu „sauer“ darstellt. Eine Überprüfung dieses Vorschlages ermöglicht die Zuckerlösung.

3. Verwendung einer basischen Lösung mit dem pH-Wert 9
   Hinter diesem Konzept verbirgt sich eine erste systematische Herangehensweise. Schüler begründen diesen Vorschlag mit der Subtraktionsrechnung “neun minus zwei ergibt sieben“. Zur Überprüfung der Hypothese wurde eine Lösung mit einem pH-Wert 9 aufgenommen (Glasreiniger). Aufgrund des logarithmischen Charakters des pH-Werts würde jedoch das tausendfache Volumen der vorgelegten Salzsäure benötigt, um äquimolare Konzentrationsverhältnisse zu erhalten, was wiederum aus technisch-apparativen Gründen nicht möglich ist.
4. Man benötigt eine Lösung mit den größten basischen Eigenschaften
   Diese Überlegung ist bei den vorgegebenen Konzentrationsverhältnissen prinzipiell zielführend und wird von Schülern als eine der ersten Herangehensweisen präferiert. Sofern jedoch nicht mit der Pasteurpipette tropfenweise Rohrreinigerlösung zugegeben wird, führt dieser Ansatz in der Regel zu einem Übertitrieren der Salzsäure.
5. Man benötigt bei gleichen Volumina eine basische Lösung, die genauso viele pH-Einheiten vom Neutralpunkt entfernt ist, wie die Salzsäure mit einem pH-Wert von zwei
   Dieser Ansatz stellt unter den gegebenen Bedingungen die Optimallösung dar und führt zum Einsatz von Natriumhydroxidlösung. Dennoch muss auch in diesem Fall möglichst tropfengenau mit den Pasteurpipetten gearbeitet werden, da sich ansonsten schnell geringfügig saure bzw. basische pH-Werte einstellen.

Darüber hinaus zeigen sich vereinzelt noch weitere Lösungsansätze, die aber im Allgemeinen experimentell falsifiziert werden können, wie zum Beispiel der Einsatz einer geringfügig sauren Lösung mit einem pH-Wert von fünf, da „fünf plus zwei sieben ergibt“.

Auswertung und Hausaufgabenstellung

Die von den Gruppen umgesetzten Lösungsvorschläge werden mit Hilfe des pH-Zahlenstrahls auf einer Overheadfolie besprochen. Die im 5. Lösungsansatz dargelegte theoretische Überlegung wird zur Sicherung im Tafelbild festgehalten, der Fachbegriff „Neutralisation“ eingeführt und als „die Aufhebung der sauren Eigenschaften einer Lösung mit einer basischen Lösung (oder umgekehrt)“ definiert.

Die abschließende Hausaufgabenstellung zielt mit dem Gegenstand „Magensäure und Sodbrennen“ erneut auf einen alltagsweltlichen Inhaltsbereich. Die Schüler sollen zunächst Überlegungen zur funktionellen Bedeutung der Magensäure anstellen und im zweiten Teil der Hausaufgabe versuchen, die Wirkungsweise eines Antazidums zu erklären. Dabei werden die tatsächlich zugrunde liegenden Pufferaspekte im Sinne einer sektoralen didaktischen Reduktion, gemäß den Anforderungen der Jahrgangsstufe sieben, nicht thematisiert.

5.2.6 Die fünfte Unterrichtsstunde

Thema der Stunde

Reaktion von Nichtmetalloxiden mit Wasser am Beispiel des sauren Regens

Einstieg

Zum Einstieg in die fünfte Unterrichtsstunde erhalten die Schüler zur integrierten Wiederholung zunächst ein Puzzle, welches die zentralen Lernziele der ersten vier Stunden beinhaltet (s. Anhang).

Die anschließende Hausaufgabenbesprechung greift mit der desinfizierenden Wirkung der Magensäure (neben der biochemischen Funktion bei der Eiweißverdauung) einen Säureaspekt auf, der bereits in Zusammenhang mit dem Säureschutzmantel der Haut thematisiert wurde. Im zweiten Teil der Hausaufgabe wird von den Schülern erwartet, dass sie die Eigenschaften eines Antazidums auf Grund der Verwendung bei einer Übersäuerung des Magens als basisch beschreiben, es jedoch aus zwei Gründen nicht die intendierte Wirkung sein kann, die Magensäure zu neutralisieren. Zum einen, das sollte auch aus der ersten Aufgabenstellung hervorgehen, kommt der Magensäure eine wichtige Funktion im Organismus zu, weshalb der natürliche pH-Wert des Magens erhalten bleiben sollte. Zum anderen würde eine hypothetische Neutralisation der Magensäure auf Grund des geringen pH-Werts entweder ein immenses Volumen eines Antazidums mittleren pH-Werts, oder aber den Einsatz eines gesundheitsgefährdenden, weil sehr basischen Antazidums erfordern.

Das Thema der fünften Stunde behandelt die Reaktion von Nichtmetalloxiden mit Wasser. Damit wird die Stringenz des Unterrichtsgangs zunächst unterbrochen und mit der Stoffklasse der Oxide auf einen Gegenstand der Unterrichtsreihe „Luft und Verbrennung“ rekurriert, mittelbar wird mit dem Verhalten wässriger Nichtmetalloxidlösungen jedoch der Themenbereich saure und basische Lösungen wieder aufgegriffen.

Aufgabenstellung

Die fünfte Stunde beinhaltet die einzige nicht experimentell zu bearbeitende Aufgabenstellung, da mit der zu thematisierenden Stoffklasse der Nichtmetalloxide bei Raumtemperatur gasige Verbindungen vorliegen, deren Einsatzmöglichkeit im Schülerexperiment zum Zeitpunkt der Intervention aus technischen Gründen nicht möglich war. Für die Zukunft wird geplant, auch diese Stoffe durch Abfüllen in Einwegspritzen mit Einweghähnen einem schülerexperimentellen Arbeiten zugänglich zu machen (Walpuski 2003). Die Aufgabenstellung dieser Stunde ist daher an einen Sachtext (Abb. 5.9) gebunden, an den eine dreiteilige Aufgabenstellung geknüpft ist, die von den Schülern in kooperativer Kleingruppenarbeit zu bearbeiten ist.

Interventionsphase

Die Schüler sollen zunächst die Auswirkungen des sauren Regens auf die Umwelt zusammenstellen und dadurch im schülerinternen Gespräch wechselseitig ihr Textverständnis dokumentieren. Die zweite Aufgabe fokussiert auf die Entstehungsbedingungen des sauren Regens. Hierzu ist es notwendig, dass die Schüler die relevanten Gase dem Text entnehmen und die Löslichkeit dieser Gase in Regenwasser als entscheidende Voraussetzung für die Bildung des sauren Regens anführen.[7] Die letzte Aufgabe fragt nach der Gemeinsamkeit aller am sauren Regen beteiligter Gase und stellt damit für die Schüler die größte zu erbringende Transferleistung da. Ein Verweis darauf, dass alle beteiligten Verbindungen Gase oder alternativ Oxide seien, ist nicht hinreichend, da den Schülern aus ihrem Chemieunterricht auch andere Gase (Sauerstoff, Stickstoff) und auch weitere Oxide (Calciumoxid, Magnesiumoxid) bekannt sind, welche aber offenkundig nicht an der Entstehung des sauren Regens beteiligt sind. Stattdessen wird erwartet, dass auf die aus dem Anfangsunterricht bekannte Unterscheidung der Elemente in Metalle und Nichtmetalle zurückgegriffen wird und alle am sauren Regen beteiligten Gase als Nichtmetalloxide bezeichnet werden können.

Waldschäden durch sauren Regen

Unsere Wälder sind krank. Besonders betroffen sind die Nadelbäume; aber auch an den Laubbäumen werden zunehmend mehr Schäden beobachtet.
Wissenschaftler vermuten, dass eine der Hauptursachen für die Waldschäden die Luftverschmutzung ist. Kraftwerke, Industrieanlagen und Kraftfahrzeuge blasen jährlich Millionen Tonnen schädlicher Abgase in die Luft.
Der Ausstoß von Schwefeldioxid ist der Hauptverursacher des „sauren Regens". Hinzu kommen weitere Gase wie Stickstoffoxide und Kohlenstoffoxide. Beim Lösen dieser Gase im Regenwasser bilden sich saure Lösungen. Diese sind für die Ausbildung des sauren Regens verantwortlich.
Die Folge des „sauren Regens" ist eine Schädigung der Blätter und Wurzeln der Bäume und eine Herabsetzung ihrer natürlichen Widerstandskraft gegenüber tierischen und pflanzlichen Schädlingen. Außerdem sind auch Seen und Gewässer von den Auswirkungen des „sauren Regens"

Abb. 5.9: Informationstext zum sauren Regen

Auswertung und Hausaufgabenstellung

Die Ergebnisse der Gruppenarbeit werden zur Auswertung vergleichend vorgestellt und mit dem oben skizzierten Erwartungshorizont der Aufgabenstellungen abgeglichen. Darüber hinaus ist allgemein festzuhalten, dass beim Lösen von Nichtmetalloxiden in Wasser saure Lösungen entstehen. Ausgehend von dieser Erkenntnis wird es den Lernenden nun möglich, den zunächst unerwartet sauren pH-Wert des Mineralwassers (3.Std.) auf die Lösung des Nichtmetalloxids Kohlenstoffdioxid in Wasser zurückzuführen.

Die zweiteilige Hausaufgabenstellung fordert die Schüler zunächst dazu auf, ihren Eltern die Voraussetzungen und die Entstehung des sauren Regens zu erklären und

[7] Als binnendifferentielle Maßnahme kann an dieser Stelle der Frage nachgegangen werden, ob und wie sich der pH-Wert während eines Regenschauers verändert.

den durchschnittlichen pH-Wert von saurem Regen zu ermitteln. Im Anschluss sollen Überlegungen angestellt werden, wie auf der Verhaltensebene dazu beigetragen werden kann, die verursachenden Faktoren des sauren Regens zu reduzieren.

5.2.7 Die sechste Unterrichtsstunde

Thema der Stunde

Reaktion von Metalloxiden in wässriger Lösung

Einstieg

Bei der Besprechung der Hausaufgaben sollte deutlich werden, dass die für den sauren Regen verantwortlichen Gase nicht allein auf industrielle Emissionen zurückzuführen sind, sondern auch durch Kaminbrand, offenes Feuer und PKW-Verkehr freigesetzt werden, also bei nahezu jeder Verbrennungsreaktion. Damit sollte auch herausgestellt werden, dass sich der Schadstoffausstoß auch durch Beschränkungen im privaten Bereich reduzieren lässt.

Eine kurze Wiederholung zur Theorie der Entstehung des sauren Regens erfolgt im Schülervortrag. Parallel hierzu bereitet der Lehrer ein Modellexperiment vor, in welchem eine Schwefelportion in einem wassergefüllten Standzylinder verbrannt wird. Nach erfolgter Verbrennung sind die gasigen Schwefeldämpfe für die Schüler deutlich sichtbar über der wässrigen Phase im Standzylinder zu erkennen. Durch Schütteln des Standzylinders entsteht schwefelige Säure, wodurch modellhaft der sauren Regen simuliert wird. Gleichzeitig lässt sich die Löslichkeit von Schwefeldioxid in Wasser veranschaulichen, da nach dem Schütteln die Gasphase ungetrübt ist. Legt man im Standzylinder 150 – 200 mL Wasser vor, so beträgt der pH-Wert der Lösung nach dem Schütteln 3-4, was ein realistischer Bereich für den natürlichen sauren Regen ist. Anschließend wird der hergestellte „saure Regen“ in Probenflaschen abgefüllt und auf die Schülergruppen verteilt.

Aufgabenstellung

Die Schüler erhalten die Aufgabe, die erhaltene Probe des sauren Regens so gut wie möglich zu neutralisieren und anzugeben, welche Gemeinsamkeit alle Stoffe besitzen, die sich zur Neutralisation eignen.

Im Unterschied zur geforderten Neutralisation der vierten Stunde erhalten die Schüler nicht die zwölf Lösungen mit bekanntem pH-Wert, sondern lediglich vier Feststoffproben (Calciumoxid, Zucker, Magnesiumoxid und Kochsalz). Erschwerend wird auch die Anzahl der pH-Teststäbchen auf fünf reduziert, um zu verhindern, dass durch unplanmäßiges Experimentieren, solange verschiedene Lösungen zusammengeschüttet werden, bis ein neutraler pH-Wert gemessen werden kann.

Interventionsphase

Analog zur vierten Unterrichtsstunde sollten die Schüler zunächst den pH-Wert ihrer sauren Lösung bestimmen. Um den Verbrauch an pH-Teststäbchen einschränken zu können, wurde der Interaktionsbox mit Bromthymolblau ein weiterer Indikator beige-

geben. Die Schüler sollten diesen Indikator nutzen, um richtige Tendenzen in Bezug auf eine Annäherung an einen neutralen pH-Wert während der Neutralisationsversuche zu ermitteln.

Ein planvolles Vorgehen erwartet ferner, dass die Schüler im nächsten Schritt Lösungen aller Feststoffe herstellen und deren pH-Wert bestimmen. Dabei ist festzustellen, dass trotz der geringen Löslichkeit von Calcium- und Magnesiumoxid deren Überstand deutlich basisch reagiert, wohingegen die wässrigen Lösungen von Kochsalz und Rohrzucker neutral reagieren. Auch hier bietet sich in Anlehnung an die dargestellten Lösungsvorschläge der vierten Stunde die Möglichkeit, Schülerkonzepte, wie das Neutralisieren mit einer süßen oder einer neutralen Lösung, experimentell umzusetzen und zu falsifizieren. Zielführend ist dagegen die tropfenweise Zugabe der wässrigen Lösungen von Calciumoxid oder Magnesiumoxid, wobei Magnesiumoxid aufgrund der geringeren Basizität praktikabler in der Handhabung ist. Um den zweiten Teil der Aufgabenstellung beantworten zu können, enthalten die Interaktionsboxen eine instruktionale Hilfe, auf der Informationen zu den vier Feststoffen enthalten sind (Abb. 5.10). Dieser Hilfe können die Schüler entnehmen, dass die beiden Stoffe, die sich zur Neutralisation eignen, Metalloxide sind. Es sollte die verallgemeinernde Aussage getroffen werden, dass wässrige Lösungen von Metalloxiden basisch reagieren. Die pauschale Angabe „alle Oxide können verwendet werden“, ist nicht hinreichend, da in der letzten Stunde die Bildung saurer Lösungen aus Nichtmetalloxiden thematisiert wurde.

Calciumoxid:	**Hierbei handelt es sich um ein Metalloxid, welches das Metall Calcium mit Sauerstoff bildet.**
Zucker:	**Zucker ist eine pflanzliche Verbindung, die aus den Elementen Kohlenstoff, Wasserstoff und Sauerstoff besteht.**
Kochsalz:	**Der chemische Name dieser Verbindung lautet Natriumchlorid. Dieses Salz lässt sich aus den Elementen Natrium und Chlor herstellen.**
Magnesiumoxid:	**Diese Verbindung gehört zu der Gruppe der Metalloxide. Magnesiumoxid entsteht bei der Reaktion des Metalls Magnesium mit Sauerstoff.**

Abb. 5.10: Hinweise zu den Feststoffen der sechsten Stunde

Auswertung und Hausaufgabenstellung

Die zentralen Ergebnisse der Gruppenarbeit sind inklusive der verallgemeinernden Aussage über die Basizität metalloxidischer Lösungen im Tafelbild zu sichern.

In der Hausaufgabenstellung sollen die Schüler zunächst reproduktiv die zentralen Ziele der fünften und sechsten Stunde vergleichend darstellen und angeben, dass wässrige Lösungen von Nichtmetalloxiden sauer und wässrige Lösungen von Metalloxiden basisch reagieren. Der zweite Teil der Hausaufgabenstellung zielt auf den Einsatz von gebranntem Kalk als Gegenmaßnahme für die in Folge des sauren Regens zu verzeichnende Versauerung von Waldböden. Mit Hilfe eine beigelegten Tex-

tes (s. Anhang) sollen sich die Schüler zunächst über die Entstehung von Calciumcarbonat informieren und angegeben können, dass gebrannter Kalk der Trivialname für Calciumoxid ist, welches durch Brennen von Calciumcarbonat bei 1000°C unter Abspaltung von Kohlenstoffdioxid gebildet wird.

Abschließend sollen die Schüler unter Verwendung aller innerhalb der Unterrichtsreihe erworbener Fachbegriffe die Funktion des Ausstreuens von gebranntem Kalk angeben. Dabei soll deutlich werden, dass das reine Aufstreuen von gebranntem Kalk noch nicht zur Absenkung des sauren pH-Werts führen kann, sondern erst in Kontakt mit (Regen)Wasser eine basische Lösung gebildet wird.

6. Auswertung der Schulleistungstests

Der Gesamtstudie liegt ein quasi-experimentelles Design mit zwei zweifach gestuften unabhängigen Variablen zugrunde (vgl. Kp. 4.3), wodurch sich nach vollständiger Kombination beider Faktoren vier Untersuchungszellen (im Folgenden U^+/H^+, U^+/H^-, U^-/H^+, U^-/H^-) ergeben. Mit Hilfe multivariater, mehrfaktorieller Varianzanalysen mit Messwiederholung wurde der genuine Haupteffekt und der Interaktionseffekt (Gruppe*Zeit) auf die Lernleistung und die Lernmotivation geprüft (Sumfleth et al. 2002). Im Rahmen dieser Arbeit erfolgt eine Beschränkung der Datenauswertung hinsichtlich des Einflusses der Unterrichtsintervention auf die Lernleistung. Die inferenzstatistische Wirkungskontrolle der Intervention über alle drei Messzeitpunkte erfolgte ebenfalls durch Varianzanalysen mit Messwiederholung. Darüber hinaus kommen zur statistischen Überprüfung differentieller Lernzuwächse zwischen den Gruppen einfaktorielle Varianzanalysen ohne Messwiederholung (ANOVA) zum Einsatz[8].

6.1 Eingangsbedingungen und Treatment-Check

Um eine Vergleichbarkeit der Untersuchungsgruppen zu gewährleisten, wurden diese in Bezug auf verschiedene Variablen wie Alter und Geschlecht der Schüler, Zahl der Repetenten sowie weiterer Variablen untersucht. Da aus kognitionspsychologischer Sicht schon seit langem das Vorwissen der Schüler als ein entscheidender Prädiktor der Lernleistung identifiziert wurde und dieser Sachverhalt auch durch umfangreiche fachdidaktische Forschungsergebnisse konkretisiert werden konnte (Sumfleth 1988, Sumfleth 1996, Sumfleth & Telgenbüscher 2000), wurde ca. drei Wochen vor der jeweiligen Intervention ein Vortest durchgeführt, in dem das fachspezifische Vorwissen erhoben wurde.

Nach Codierung der Variablen ergibt sich für die acht Untersuchungsklassen folgende deskriptive Datenlage bezüglich der offenen Fragen sowie des Multiple-Choice-Tests.

			Offene Fragen		Multiple-Choice	
Schule	Klasse	N $_{(\Sigma = 211)}$	M	SD	M	SD
KG	1	26	2,5	2,2	25,5	5,2
	2	28	1,8	1,1	24,8	7,5
	3	24	2,2	1,4	23,7	8,5
	4	26	2,7	1,8	28,1	7,5
FH	1	26	2,8	1,4	26,5	7,6
	2	28	2,9	1,7	27,9	4,8
	3	27	2,3	1,7	26,7	6,6
	4	26	2,6	1,7	25,7	5,8

Tabb. 6.1: Deskriptive Statistiken der Daten des Eingangstests
M= Mittelwert, SD= Standardabweichung

[8] Die statistische Auswertung wurde mit dem Softwareprogrammpaket SPSS 11.0.1 durchgeführt.

Eine auf der Basis dieses Datensatzes gerechnete Varianzanalyse (ANOVA) ergibt für beide Testformen keinen signifikanten Unterschied zwischen den acht Untersuchungsklassen ($F_{(7,210)}$ = 1.14; n.s. (offene Fragen) und $F_{(7,210)}$ = 1.28; n.s. (Multiple-Choice)). Eine varianzanalytische Untersuchung der Eingangsleistungen, differenziert nach den beiden Gymnasien, zeigte ebenfalls keinen signifikanten Unterschied ($F_{(1,210)}$ = 2.42; n.s. bzw. $F_{(1,210)}$ = 1.58; n.s.).
Daher können bezüglich des Vorwissens sowohl Klassen- als auch Schuleffekte ausgeschlossen und die acht Klassen dem Arbeitsplan entsprechend auf die vier Untersuchungszellen aufgeteilt werden (vgl. Kp. 4).

Die erreichten Punktzahlen der Probanden der vier Treatments im Eingangstest verteilen sich wie folgt:

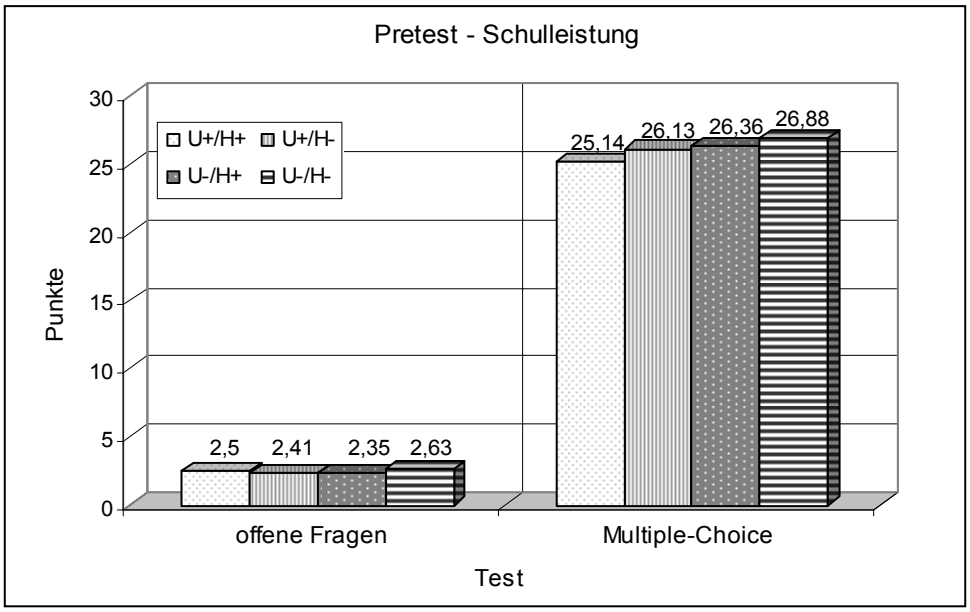


Abb. 6.1: Vergleich der Leistungseingangsdaten

Der Skalenbereich des Tests mit offenem Antwortformat verläuft von 0 bis 20 Punkte, der des Multiple-Choice-Tests von 0 bis 56 Punkte. Entsprechend der unterschiedlichen Antwortformate unterscheiden sich die im Mittel erreichten Punktwerte beider Testformen sowohl relativ zueinander als auch absolut zum jeweiligen Maximalwert. Während die im Mittel nur zu gut 12 Prozent richtig beantworteten offenen Fragen vollständig auf Vorwissen zurückzuführen sind, erklärt sich die richtige Beantwortungsquote von über 46 Prozent im Multiple-Choice-Test durch die Ratewahrscheinlichkeit, welche bei der vorliegenden dichotomen Item-Antwortmöglichkeit 50 Prozent beträgt.

Nach inferenzstatistischer Analyse zeigen die vier Treatments erwartungsgemäß keine statistisch bedeutsamen überzufälligen Differenzen ($F_{(3,210)}$ = 0.28; n.s. (offene Fragen) bzw. $F_{(3,210)}$ = 0.59; n.s. (Multiple Choice)), ein Gruppeneffekt zum Zeitpunkt t_0 kann somit ausgeschlossen werden. Damit sind die vier aggregierten Untersuchungszellen einer weiteren Analyse von Haupteffekten und treatmentinduzierten Interaktionseffekten zugänglich.

6.2 Inferenzstatistische Auswertung der offenen Aufgaben

Eingangs des Fachtests wurden den Schülern fünf Aufgaben im freien Antwortformat vorgelegt. Lienert & Raatz (1998) unterscheiden Aufgaben mit freiem Antwortformat hinsichtlich so genannter *Ergänzungsaufgaben* von den in dieser Studie gewählten *Kurzaufsätzen*, bei denen die Probanden selbstständig einen kurzen Fließtext zur Beantwortung der Aufgabenstellung formulieren müssen. Während Ergänzungsaufgaben, die durch die wortweise Ergänzung einer Aussage gekennzeichnet sind (z. B. Lückentext), noch ein vertretbares Maß an Auswertungsobjektivität zugesprochen wird, stellt dieser Aspekt eine Schwierigkeit bei Kurzaufsätzen dar, weshalb deren Standardisierbarkeit häufig in Frage gestellt und dieser Aufgabentypus in aktuellen Tests nur noch selten verwendet wird (Lienert & Raatz 1998). Für den Einsatz von Kurzaufsätzen spricht hingegen ihre hohe ökologische Validität, da dies der im Kontext des Fachunterrichts vorherrschende Aufgabentyp ist, wohingegen Aufgaben mit einem gebundenen Antwortformat nur sehr selten verwendet werden. Da der zentrale Kritikpunkt gegenüber Kurzaufsätzen auf ihre mangelnde Standardisierbarkeit abhebt, wurde zur Gewährleistung der Auswertungsobjektivität für jede Aufgabe ein Punkteschlüssel erstellt, dem dezidierte Angaben über die Vergabe von (Teil-) Punkten zu entnehmen sind. Jede Aufgabe wurde mit maximal vier Punkten bewertet. Die Güte des Punkteschlüssels wurde durch die Berechung der Auswertungsübereinstimmung zweier unabhängiger Auswerter an Hand einer Stichprobe im Umfang von zehn Prozent der Gesamtprobandenzahl hinreichend bestätigt ($\kappa > .8$). Darüber hinaus zeigen die offenen Fragen eine zufrieden stellende Korrelation ($r = .59$) mit dem standardisierten Multiple-Choice-Test. Aus den dargestellten Gründen erscheint der Einsatz einer begrenzten Anzahl von offenen Aufgaben im Verbund mit anderen standardisierten Testverfahren zur Leistungserhebung gerechtfertigt. Im Folgenden werden zunächst die Haupt- und dann die Interaktionseffekte an Hand der Datenauswertung der offenen Fragen dargestellt.

6.2.1 Offene Fragen - Haupteffekt >Zeit<

Nach Auswertung und Codierung der fünf offenen Fragen ergibt sich für die vier Experimentalgruppen zu den Messzeitpunkten im Pre- und Post-Test folgende Leistungsverteilung:

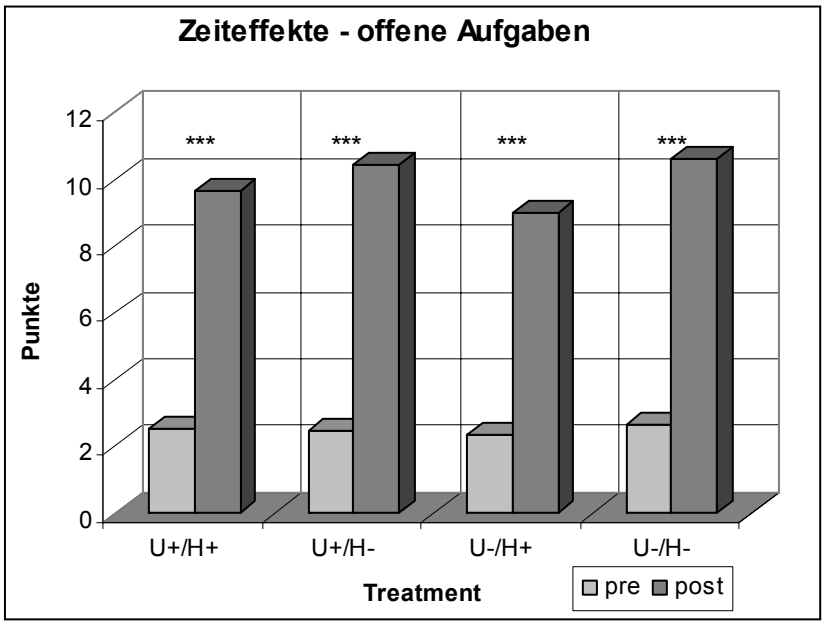


Abb. 6.2: Haupteffekt >Zeit< der vier Treatments

Wie die Grafik verdeutlicht, ist im Vergleich der Pre-Post-Daten eine große Leistungszunahme über die Unterrichtsstunden zu erkennen. Dieser Unterschied ist für alle Treatments hoch signifikant ($p<.001$). Zusätzlich wurden für den Haupteffekt >Zeit< aller Treatments die separaten partiellen Effektgrößen (η^2) berechnet, die in einem Bereich zwischen .75 (U^-/H^+) und .84 (U^+/H^-) liegen, was auf einen hohen Effekt des erteilten Unterrichts und somit auch der problemorientierten Unterrichtsskripte (vgl. Kp. 5) schließen lässt.

6.2.2 Interaktionseffekt

Während durch die Haupteffekte ausschließlich differentielle Aussagen über die Treatments zu einem bestimmten Zeitpunkt (Haupteffekt >Gruppe<) bzw. eines einzelnen Treatments über die Zeit (Haupteffekt >Zeit<) gemacht werden können, kennzeichnet die Interaktion einen darüber hinausgehenden Effekt. Dieser ist dadurch zu erklären, dass mit der Kombination beider Faktoren eine abweichende Wirkung oder ein eigenständiger Effekt verbunden ist (Bortz 1999: 285). Das bedeutet allerdings nicht, dass im vorliegenden Fall ein Interaktionseffekt lediglich einen gemeinsamen Effekt der Faktoren >Gruppe< und >Zeit< beschreibt, da ein rein additives Zusammenwirken beider Faktoren den erwartungsgemäßen „Normalfall“ definiert. *Überzufällige Abweichungen* von der Additivität hingegen lassen sich nur durch die Wirkung eines zusätzlichen (Interaktions-)Effekts, hervorgerufen durch Wechselwirkung der Faktoren, erklären (Bortz & Döring 1995). Dadurch allerdings wird es möglich, Erkenntnisse über die *differentielle* Leistungs*entwicklung* der Treatmentgruppen über die Zeit zu gewinnen. Weiterhin wirkt sich die spezielle Art einer identifizierten Interaktion - ordinal, hybrid oder disordinal (s.u.) - auf die Interpretation der Haupteffekte aus, die dann gegebenenfalls zu modifizieren ist.

Wie bei der Darstellung des Testinstrumentariums (Kp. 4) ausgeführt, wurden zur Erhebung interventionsüberdauernder Effekte alle Leistungstests sechs Monate nach der Intervention im Rahmen einer Follow-up-Erhebung ein weiteres Mal durchge-

führt, um so Informationen über die Nachhaltigkeit der Leistungsveränderungen zu erhalten. Mit Hinblick auf das auszuwählende varianzanalytische Verfahren resultiert hieraus ein Untersuchungsplan, der oberflächlich betrachtet einem zweifaktoriellen Plan (Gruppe und Zeit) entspricht, was eine zweifaktorielle ANOVA als statistisches Auswerteverfahren erfordern würde. Tatsächlich ist jedoch ein zweifaktorieller Untersuchungsplan dadurch gekennzeichnet, dass für jede Faktorenkombination eine andere Stichprobe untersucht wird, wohingegen im vorliegenden Fall, bedingt durch die Messwiederholung, die Messwerte eines Treatments mehrmals untersucht werden und damit voneinander abhängen bzw. korrelieren (Bortz & Döring 1995). Aus diesen Randbedingungen resultiert ein Messverfahren, welches als *Varianzanalyse mit Messwiederholung* bezeichnet wird und das im Gegensatz zu einer einfachen zweifaktoriellen Varianzanalyse durch zusätzliche Kontrollberechnungen und Korrekturfaktoren gekennzeichnet ist.

Als eine notwendige Voraussetzung der Varianzanalyse wird eine Homogenität sowohl in den beobachteten Fehlervarianzen der einzelnen Faktorenstufenkombinationen als auch in den Korrelationen zwischen den Faktorstufen gefordert (Forderung nach *Zirkularität* bzw. *Sphärizität*, Bortz 1999). In der Praxis zeigt sich jedoch häufig, dass diese Forderung verletzt wird, was zu einer progressiven Entscheidung zugunsten der H_1 führt. Daher sollte bei Versuchsplänen mit mehr als zwei Messwiederholungen geprüft werden, ob die erste Messung mit der zweiten Messung in dem gleichen Maße korreliert wie mit der dritten Messung. Als eine Prüfmethode bietet sich hierzu der Mauchly-Test auf Sphärizität an, dem, sofern die Sphärizität verletzt wird, Korrekturfaktoren zur so genannten Epsilonadjustierung zu entnehmen sind. Mit Hilfe dieses Faktors (<1) werden die Zähler- und Nennerfreiheitsgrade für den kritischen F-Wert reduziert, wodurch letztendlich die Wahrscheinlichkeit einer progressiven Entscheidung zugunsten der H_1 verringert wird (Bortz 1999).

Im Folgenden wird eine Varianzanalyse der Leistungsergebnisse der offenen Aufgaben über alle drei Messzeitpunkte durchgeführt. Auf Grund der zwischen den Messzeitpunkten t_2 und t_3 liegenden Versetzungszeugnisse reduziert sich in diesem Fall der Datensatz auf n= 169.

Treatment	Pre-Test		Post-Test		Follow-up-Test		N (Σ=169)
	M	SD	M	SD	M	SD	
U^+/H^+	**2,6**	1,5	**9,9**	4,5	**8,4**	4,1	41
U^+/H^-	**2,6**	2,0	**11,0**	3,3	**9,9**	3,8	44
U^-/H^+	**2,4**	1,7	**9,1**	4,1	**7,7**	4,1	44
U^-/H^-	**2,8**	1,9	**10,9**	3,9	**8,0**	4,0	40

Tab. 6.2: Offene Fragen - Deskriptive Statistiken der Varianzanalyse mit Messwiederholung

Erwartungsgemäß zeigen alle Treatments ihr Leistungsmaximum zum Zeitpunkt des Post-Tests (vgl. Abb. 6.3), wohingegen sechs Monate später ein für alle Treatments

signifikanter Rückgang der mittleren Punktzahl zu verzeichnen ist. Bei der weiteren deskriptiven Analyse der Testleitungen treten drei Aspekte in den Vordergrund:

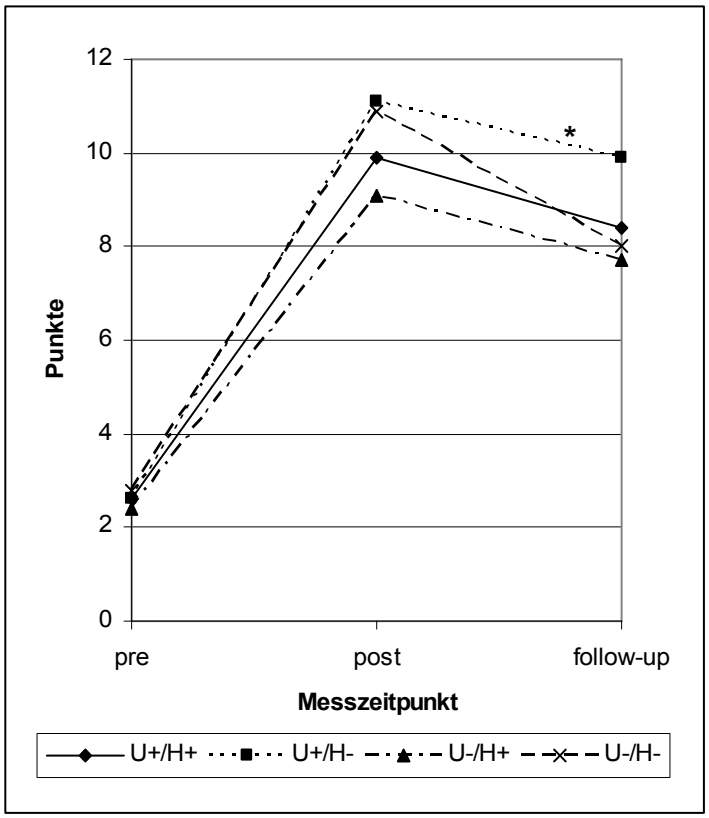


Abb. 6.3: Leistungsentwicklung „offene Fragen“

- Sowohl zum zweiten als auch zum dritten Messzeitpunkt kann das halbintervenierte Treatment U^+/H^- den höchsten absoluten Lernzuwachs verzeichnen.
- Ebenso erzielt das Treatment mit der umgekehrten Faktorenkombination (U^-/H^+) zu allen Messzeitpunkten das schlechteste Testergebnis.
- Erwartungswidriger Weise schneidet das zweifach intervenierte Treatment U^+/H^+ zum zweiten Messzeitpunkt nicht am besten ab. Die reine Kontrollgruppe U^-/H^- erreicht hingegen eine Punktzahl, die nur marginal (nicht signifikant) schlechter ist, als die der Spitzengruppe. Dieser Befund kehrt sich jedoch zum Follow-up-Test um, da das Treatment U^-/H^- den stärksten Leistungsabfall verzeichnet, was auf eine nur geringe Nachhaltigkeit des erworbenen Wissens schließen lässt.

Die inferenzstatistische Auswertung über alle drei Messzeitpunkte zeigt nach dem Mauchly-Test eine Verletzung der Zirkularitätsannahme ($p>0.05$). Als Faktor zur Epsilonadjustierung wurde die Berechnung nach Greenhouse-Geiser gewählt, da es sich hierbei um ein konservatives Verfahren handelt, das auch bei kleineren Stichprobenumfängen herangezogen werden kann (Diehl & Staufenbiel 2002). Nach anschließender Korrektur der Freiheitsgrade zeigt sich über alle drei Messzeitpunkte ein signifikanter Interaktionseffekt ($F_{(5,8,\ 323,9)} = 2,16$; $p < .05$) bei einer Effektgröße von $\eta^2=0.04$ (= 4% geklärte Varianz), was einer Effektstärke von d= 0,20 entspricht (schwach bis mittlerer Effekt nach Bortz 1997). Tendenziell signifikant ($p < .1$) ist darüber hinaus der überdurchschnittliche Leistungsabfall der Gruppe U^-/H^- vom zweiten zum dritten Messzeitpunkt, was die gute Post-Testleistung dieses Treatments mit Hinblick auf einen langfristigen, interventionsüberdauernden Lernerfolg relativiert. Dieser Befund bestätigt sich auch, wenn man zur Messung des langfristigen Lernerfolgs direkt die Leistungsentwicklung vom Pre- zum Follow-up-Test ermittelt.

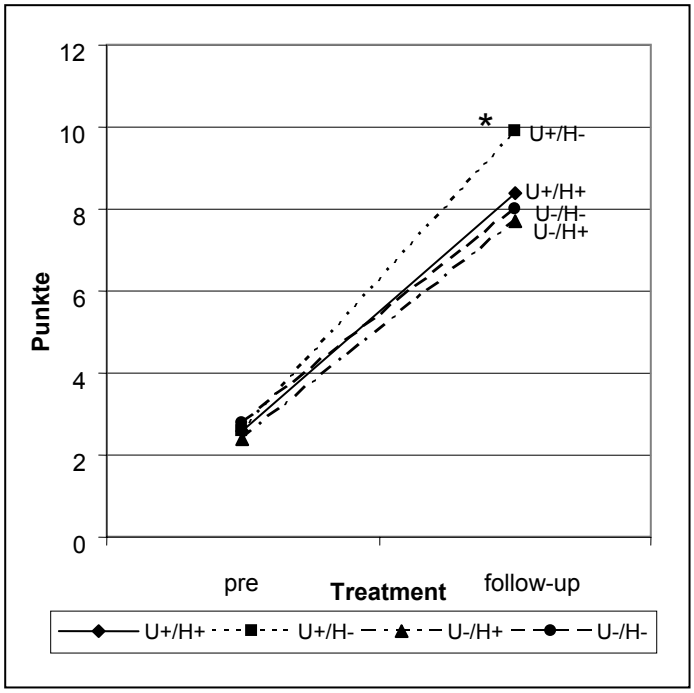


Abb. 6.4: Leistungsentwicklung „offene Fragen“ t_0 – t_2

Neben einem bereits in der Grafik (Abb. 6.4) deutlich sichtbaren Leistungsvorteil des U^+/H^--Treatments ($F_{(3,165)}$ = 2,8; $p < .05$) zeigt nun das U^+/H^+-Treatment geringfügig bessere, wenngleich nicht signifikante, Leistungsergebnisse als das U^-/H^--Treatment. Ebenso ist eine Zunahme der Effektgröße in der Interaktion zu verzeichnen (η^2=0.05; d= 0.23).

Statistisch signifikante Leistungsergebnisse sind hingegen nicht zu verzeichnen, wenn man statt der Pre- und Follow-up-Testergebnisse die Leistungsentwicklung unmittelbar nach Beendigung der Intervention vergleicht und die Pre- und Post-Testergebnisse einer Varianzanalyse unterzieht. Dies gilt unabhängig davon, ob die Berechnung auf Basis des versetzungsbedingt probandenreduzierten Datensatzes ($F_{(3,165)}$= 1,6; n.s.) oder an Hand des Ausgangsdatensatzes erfolgt ($F_{(3,206)}$= 1,5; n.s.).

In einer Zwischenbilanz zur Auswertung der Fachaufgaben im offenen Antwortformat lässt sich somit feststellen, dass ein lernförderliches Potenzial der Intervention erst nach einem größeren Zeitraum zur Geltung kommt, wohingegen direkt nach Abschluss der Intervention Treatment- und Kontrollgruppe keinen signifikant unterschiedlichen Leistungszuwachs aufweisen. Ebenso zeigt sich der größte Interaktionseffekt bei direkter Kontrastierung der Messzeitpunkte t_1 gegen t_3.

6.2.2.1 Interaktionsdiagramme

Der Nachweis eines signifikanten Interaktionseffekts (Abb. 6.3) weist nicht nur auf eine Interaktion der beiden festen Faktoren >Gruppe< und >Zeit< hin, sondern ist darüber hinaus bedeutsam als Beleg für die Zulässigkeit einer globalen Interpretation signifikanter Haupteffekte (Bortz & Döring 1995). Zur genauen Bestimmung der Art der Interaktion wird zunächst für jeden der beiden festen Faktoren ein so genanntes *Interaktionsdiagramm* erstellt, bei dem auf der Ordinate die abhängige Variable (Punktzahl) und auf der Abszisse die Faktorstufen eines Hauptfaktors aufgetragen werden. In das Diagramm wird für jede Stufe des jeweils anderen Faktors ein Linienzug eingefügt, der die Mittelwerte der Faktorstufenkombination verbindet. Daraus ergeben sich für die Faktoren >Gruppe< und >Zeit< die folgenden Interaktionsdiagramme:

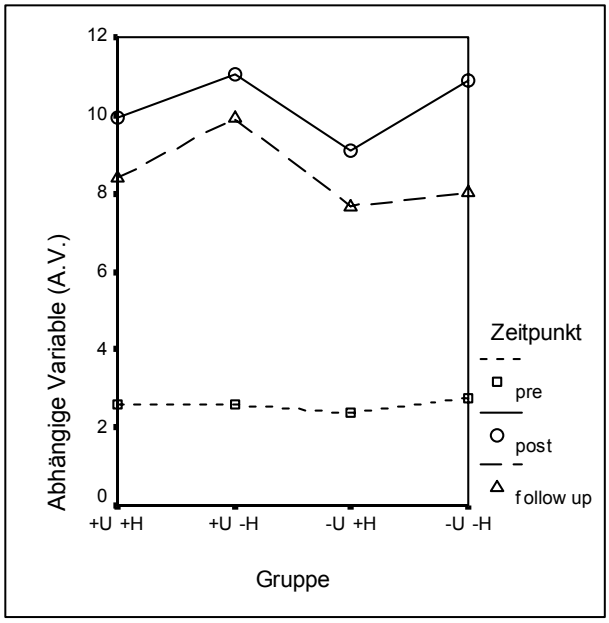


Abb. 6.5: Interaktionsdiagramm des Faktors >Gruppe<

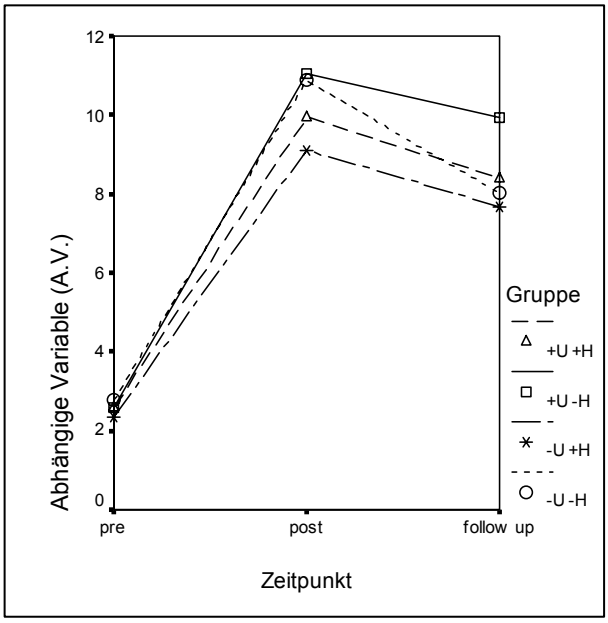


Abb. 6.6: Interaktionsdiagramm des Faktors >Zeit<

Den Diagrammen ist zu entnehmen, dass nur für den Faktor >Zeit< alle Graphen gleichsinnig (Anstieg von pre nach post, Abfall von post nach follow-up) verlaufen, wohingegen für den Faktor >Gruppe< die drei Graphen ungleichsinnig verlaufen, woraus sich als Interaktionstypus eine *hybride Interaktion* ableiten lässt, bei der nur einer der beiden Haupteffekte global interpretierbar ist. Da globale Interpretationen der Haupteffekte nur bei einem stetigen Verlauf der Graphen zulässig sind, folgt hieraus zum einen, dass der Haupteffekt >Gruppe< nicht solitär interpretiert werden darf, was unter anderem den bereits im Rahmen des Treatment-Checks nachgewiesenen Ausschluss eines Gruppeneffekts zum Zeitpunkt t_0 bestätigt. Zum anderen gestattet der nahezu waagerechte Verlauf des Graphen zum Zeitpunkt „Pre“ in Abb. 6.5 eine globale Interpretation des Haupteffekts >Zeit<, wodurch die in Kp. 6.2.1 Berechnungen legitimiert werden.

6.2.2.2 Post-hoc-Analysen

Der oben beschriebene signifikante Vorteil des U+/H- Treatments ist nicht unmittelbar den Ergebnissen der Varianzanalyse zu entnehmen, da im Rahmen des *Allgemeinen Linearen Modells* (ALM) lediglich eine signifikante Interaktion als solche identifiziert werden kann, ohne jedoch Angaben darüber zu machen, zwischen welchen Gruppen signifikante Unterschiede bestehen. Zwar legt ein Blick auf die deskriptiven Daten der Post- und Follow-up-Testergebnisse bereits einen Vorteil des U^+/H^--Treatments nahe, verlässliche Auskunft hierüber erhält man jedoch erst mit Hilfe des multiplen Mittelwertvergleichs einer nachgeschalteten *post-hoc-Analyse*. Die Vielzahl der zur Verfügung stehenden Verfahren basieren nahezu alle auf dem paarweisen Vergleich sämtlicher Mittelwerte, unterscheiden sich aber in ihrer Konservativität, das heißt in ihrer Tendenz, die H_0 im Zweifel beizubehalten. Darüber hinaus setzt die Mehrzahl der Verfahren eine Varianzhomogenität der Gruppen (die Fehler müssen in den verschiedenen Gruppen ein vergleichbare Varianz haben, was auch eine Rand-

bedingung für die Varianzanalyse ist) voraus, wohingegen einige Tests nicht auf diese Annahme basieren (z.B. Tamhame-T2 oder Dunnet T3).

Alle post-hoc-Analysen belegen, dass mit einer mittleren Differenz von ca. 1,5 Punkten der größte mittlere Leistungsunterschied zwischen den Untersuchungsgruppen U^+/H^- und U^-/H^+ zu verzeichnen ist. Aus den unterschiedlichen Berechnungsvorschriften der verschiedenen post-hoc-Testverfahren resultieren für eine Varianzanalyse der offenen Aufgaben über alle drei Messzeitpunkte gleichermaßen unterschiedliche Signifikanzen, wie der folgende Vergleich dreier populärer post-hoc-Tests, nämlich des LSD-Tests (*l*east *s*ignificant *d*ifference), des Tukey-HSD-Tests (honestly difference significance) und des Scheffé-Tests, für den paarweisen Vergleich der Untersuchungsgruppen U^+/H^- gegen U^-/H^+ verdeutlicht:

Mittlere Differenz (U^+/H^- gegen U^-/H^+)	Standardfehler	Signifikanz		
		LSD	Tukey-HSD	Scheffé
1,47	0.56	0.009**	0.047*	0.081^+

Tab. 6.3: Paarweiser Vergleich der Untersuchungsbedingungen U^+/H^- gegen U^-/H^+

Während die paarweisen Vergleiche aller anderen Kombinationen von Untersuchungsgruppen in keinem anderen Testverfahren signifikante Unterschiede zeigen, ist die mittlere Differenz zwischen den Untersuchungsgruppen U^+/H^- gegen U^-/H^+ wahlweise als hoch signifikant (LSD-Test) oder nur als tendenziell signifikant (Scheffé-Test) zu bewerten. Der Grund für die unterschiedlichen Irrtumswahrscheinlichkeiten ist im Wesentlichen darin zu sehen, ob und in wie weit die Verfahren eine Korrektur der Signifikanzgrenzen, der so genannten α-Adjustie-rung, vornehmen: Durch die Vielzahl der durchzuführenden Mittelwertvergleiche steigt die Wahrscheinlichkeit, rein zufällig einen signifikanten Unterschied zu finden; man spricht von einer α–Fehler-Kumulierung (oder auch >Inflation des 03B1–Niveaus<). Für den konkreten Fall bedeutet dies, dass die Wahrscheinlichkeit, fälschlicherweise die H_0 zu verwerfen, um mehr als das Vierfache überhöht ist, sofern keine α-Adjustierung vorgenommen wird. Somit ist der LSD-Test auf Grund des Fehlens einer α-Fehlerkorrektur der Test, der vergleichsweise progressive Entscheidungen zu Gunsten der H_1 herbeiführt, wohingegen es sich bei dem Scheffé-Test um ein konservatives Verfahren handelt, welches im einzelnen Vergleich eher für eine Beibehaltung der H_0 entscheidet (Nachtigall & Wirtz, 1998). Als Empfehlung für die Praxis verweisen Diehl und Staufenbiel (2001) auf die Ergebnisse zahlreicher Monte-Carlo-Studien, die unter dem Gesichtspunkt der α–Adjustierung das Tukey-Verfahren als eine gute Wahl erscheinen lassen, weshalb im Folgenden allein dieses Verfahren herangezogen wird.

Zusammenfassend belegen die dargelegten Analysen die Wirkmächtigkeit der verschiedenen Treatments auf die gesamte untersuchte Zielpopulation hinsichtlich eines Lerneffekts über die Zeit. Zur Prüfung der Unterschiedhypothese – *ist das unterrichtsintervenierte Treatment der Kontrollgruppe überlegen?* – werden zum Ziel der Absicherung und Differenzierung des Interaktionseffekts im Folgenden weitere Analysen durchgeführt.

6.2.3 Analyse ausgewählter offener Transferaufgaben

Während der Multiple-Choice-Test (s. Kp. 6.3) überwiegend deklaratives (nominales) Wissen bis hin zu funktionalem naturwissenschaftlichem Wissen erfasst, wird – gemäß ihrem Anforderungscharakter – mit Hilfe der offenen Transferaufgaben getestet, inwieweit das erworbene Wissen auch auf andere Zusammenhänge angewendet werden kann. Die fünf Transferaufgaben wurden so konzipiert, dass das Ausmaß der zu erbringenden Transferleistung von einer einfachen Aufgabe mit hohen reproduktiven Anteilen bis hin zu einer komplexen Aufgabe, welche auf die gesamten Lernziele der Unterrichtsreihe fokussiert, progressiv zunimmt. Auf Grund der hieraus resultierenden unterschiedlichen Aussagekraft der offenen Transferaufgaben sollen diese im Folgenden auch differenziert ausgewertet werden.

Treatment	Pre-Test					Post-Test					Follow-up-Test				
	A1	A2	A3	A4	A5	A1	A2	A3	A4	A5	A1	A2	A3	A4	A5
U^+/H^+	**0,74**	**0,66**	**0,3**	**0,86**	**0,61**	**1,89**	**2,56**	**1,15**	**2,36**	**1,97**	**1,78**	**2,33**	**0,86**	**2,05**	**1,65**
U^+/H^-	**0,12**	**0,69**	**0,22**	**0,81**	**0,58**	**2,63**	**2,39**	**1,12**	**2,33**	**1,96**	**2,45**	**2,48**	**1,23**	**1,88**	**1,94**
U^-/H^+	0,18	0,61	0,13	1,03	0,41	1,6	2,4	1,0	2,12	1,61	1,11	2,13	1,18	1,83	1,3
U^-/H^-	0,21	0,75	0,25	0,77	0,65	2,88	3	1,11	1,97	1,64	1,84	2,43	0,68	1,5	1,53

Tab. 6.4: Mittlere Leistungsergebnisse >offene Aufgaben<, differenziert nach Aufgabe und Messzeitpunkt

Die Testleistungen der Probanden zeigen auf Grund des fehlenden Vorwissens im Pre-Test über alle Treatments erwartungsgemäß durchgängig geringe Werte.

Darüber hinaus fällt bei einer ersten, oberflächlichen Interpretation der Daten das durchweg schlechte Abschneiden aller Probanden bei der Beantwortung der dritten Aufgabe auf. Wenngleich auch bei dieser Aufgabe einzelne Probanden in Post- und Follow-up-Test die maximale Punktzahl (4) erreichen, bestätigt der im Rahmen der Itemanalyse berechnete Schwierigkeitsindex der Aufgabe von .28 die Annahme, dass es sich hierbei um ein sehr schwieriges Item handelt. Vergleicht man diesen Wert mit den zufrieden stellenden Schwierigkeitsindices der übrigen Aufgaben ($.45<P<.65$), so muss bei der Auswertung dieser Aufgabe mit Hinblick auf die deskriptive Datenlage von einem Bodeneffekt ausgegangen werden.

Ursächlich für die wider Erwarten schlechte Testleistung bei diesem Item könnte eine ungenügende horizontale Vernetzung einzelner naturwissenschaftlicher Inhaltsbereiche vorangegangener Jahrgangsstufen sein, wie eine qualitative Bewertung der dargebotenen Schülerantworten zeigt: Ausgehend von der Vorgabe, dass im Magen des Menschen ein saures Milieu vorherrscht (vgl. 3./4. Unterrichtsstunde), der Stuhl hingegen einen neutralen bis basischen pH-Wert aufweist, wurden die Schüler befragt, welche Rückschlüsse sich hieraus auf die pH-Verhältnisse im Darm sowie die damit

verbundene chemische Reaktion ziehen lassen. Ein überwiegender Teil der Schülerantworten ließ hierbei Defizite beim deklarativen Wissen von der Anordnung und Reihung der Organe im Körper erkennen (Jahrgangsstufe 5/6, Biologie), ohne das eine Lösung der Aufgabe nicht möglich ist. Aus den dargelegten Gründen wird daher von einer weiteren Auswertung dieses Items abgesehen.

Aufschlussreich für eine differenzierte Einschätzung der Performanz der Treatments in Abhängigkeit vom Anforderungsniveau der Aufgaben – über die letztlich unterschiedliche Kompetenzen vermittelt werden sollen – liefern mit Aufgabe eins und fünf die Extrempole des Tests.

Mit der richtigen Zuordnung verschiedener Lösungen in die selbstständig zu nennenden Substanzklassen „saure“, „basische“ und „neutrale Lösungen“ wurde im Rahmen der ersten Aufgabe eine rein reproduktive Lernleistung gefordert (s. 1./2. Stunde). Die grafische Auswertung dieser Aufgabe (Abb. 6.7) zeigt über die drei Messzeitpunkte einen diffusen Verlauf. Unmittelbar nach Abschluss der Unterrichtsreihe lassen sich zwei Leistungsgruppen ausmachen: Das einfachintervenierte Treatment U^+/H^- sowie die reine Kontrollgruppe U^-/H^- unterscheiden sich in der Testleistung untereinander nicht signifikant, jedoch in jeder paarweisen Kombination gegenüber dem schlechter abschneidenden einfachintervenierten U^-/H^+ bzw. dem doppelt intervenierten Treatment U^+/H^+, wie post-hoc-Analysen belegen ($p < .05$).

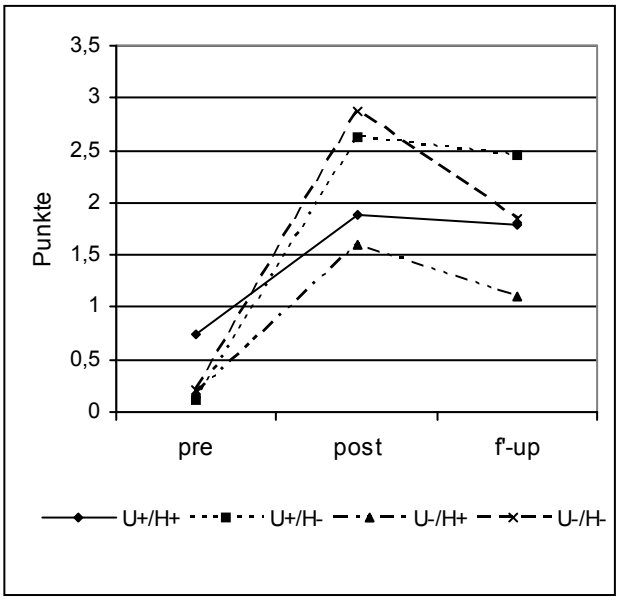


Abb. 6.7: Leistungsentwicklung Aufgabe 1

Es lässt sich daher nicht schlüssig begründen, inwieweit der Leistungserfolg bei dieser Aufgabe zum Zeitpunkt des Post-Tests auf eine der beiden – oder auch auf die Interaktion beider – Interventionen zurückgeführt werden kann. Vielmehr lässt sich vermuten, dass sowohl das Anforderungsniveau dieser Aufgabe als auch der Testzeitpunkt in diesem Fall das Leistungsergebnis überdurchschnittlich moderieren, da die reine Wiedergabe von Unterrichtsinhalten unmittelbar im Anschluss an die unterrichtete Einheit sehr stark den Lernerfolgskontrollen im traditionellen naturwissenschaftlichen Unterricht entsprechen. Treatmentinduzierte Effekte könnten daher von partiell bedeutsamen Klassen- oder Lehrereffekten (s. Kp. 6.4.4) sowie von individuellen routinierten Lernstrategien überlagert werden.

Dahingegen zeigt die Leistungsentwicklung zum Follow-up-Test einen leichter zu interpretierenden Trend: Während die Schülerinnen und Schüler der beiden unterrichtsintervenierten Treatments nach diesem Zeitraum nur geringfügig schlechtere Testleistungen zeigen, nimmt die Leistung der Lernenden, die in traditioneller Weise unterrichtet wurden drastisch ab; im Fall des Treatments U^-/H^- ist eine Leistungsabnahme von 37% zu verzeichnen, im Fall des U^-/H^+-Treatments beträgt die Leistungsabnahme 30%. Im Vergleich dazu beträgt die Leistungsabnahme der unterrichtsin-

tervenierten Treatments lediglich 6,8% (U^+/H^-) bzw. sogar nur 5,8% für das U^+/H^+-Treatment. Der Unterschied in der Leistungsabnahme beider unterrichtsintervenierten Treatments verhält sich gegenüber den beiden Treatments, in denen ein frontalunterrichtliches Arrangement umgesetzt wurde, signifikant besser, wodurch die Wirkmächtigkeit der Intervention auch in Hinblick auf die Nachhaltigkeit einfacher reproduktiver Lernziele belegt wird.

Die Leistungsergebnisse der zweiten Aufgaben zeigen in weiten Zügen ein ähnliches Bild – wenngleich auch hier die Gruppe U^-/H^- die besten Ergebnisse im Post-Test aufweist, folgt hierauf ebenso die stärkste Leistungsabnahme zum Follow-up-Test. Dies erklärt sich durch das im Vergleich zur ersten Aufgabe vergleichbare bzw. nur geringfügig gesteigerte Anforderungsniveau und ist somit indirekt auch ein Indiz für die Reliabilität beider Aufgaben.

Du bekommst die Aufgabe, mit einem Experiment herauszufinden, ob es sich bei einem dir unbekannten Pulver um ein Metall oder um ein Nichtmetall handelt. Als Materialien erhältst du: Einen Bunsenbrenner, einen Verbrennungslöffel, ein Becherglas, einen Standzylinder mit Glasplatte, Wasser und ein pH-Teststäbchen.

Anmerkung: Erhitzt man das Pulver an der Luft, so kommt es zu einer heftigen Reaktion.

Abb. 6.8: Offene Fragen, Aufgabe 5

Es steht zu vermuten, dass aus den gleichen Gründen die Ergebnisse der komplexeren Aufgaben vier und fünf in eine einheitliche Richtung weisen, weshalb im Folgenden allein eine Interpretation der fünften Aufgabe erfolgen soll, die wie folgt lautet:

Zur Lösung dieser Aufgaben müssen die Schüler tiefer gehende, elaborierte Konzepte anwenden, da zunächst einmal die Oxidbildung der unbekannten Substanz angegeben werden muss. Im Anschluss daran lässt sich an Hand der Reaktion des gebildeten Oxids gegenüber Wasser eine allgemeingültige Aussage über die Charakteristik des zugrunde liegende Elements treffen. Da der zusätzliche Arbeitsschritt – Bildung des Oxids – nicht explizit im Text erwähnt wird, sondern als notwendige, zusätzliche Information erst von den Schülern einzuholen ist und darüber hinaus die apparativ-technische Lösung zum Auffangen des gebildeten Oxids in Abhängigkeit vom zu erwartenden Aggregatzustand des Stoffes entwickelt werden muss, erfordert diese Aufgabe von den Schülerinnen und Schülern Kompetenzen, die gemäß der Typologisierung nach PISA, als *konzeptuell und prozedural* (Stufe IV) zu bezeichnen sind. Als weitere prozedurale Kompetenz muss von den Lernenden der Aggregatzustand des gebildeten Oxids ins Kalkül gezogen werden, da

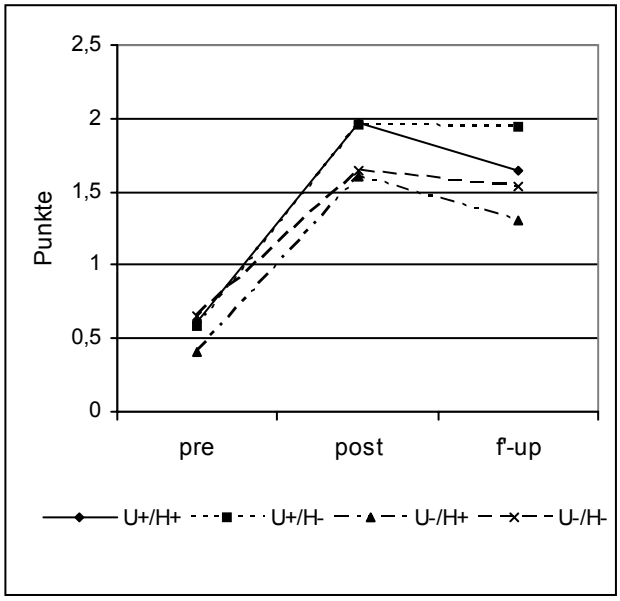


Abb. 6.9: Leistungsentwicklung Aufgabe 5

die überwiegend gasförmigen Nichtmetalloxide eine andere Versuchsapparatur (pneumatische Wanne) bedingen, als die bei Raumtemperatur festen Metalloxide.

Die grafische Darstellung der Leistungsergebnisse dieser Aufgabe zeigt einen im Vergleich zur ersten Aufgabe leichter zu interpretierenden Verlauf (Abb. 6.9). Gemäß dem größeren Schwierigkeitsgrad der Aufgabe fallen zunächst die geringeren Mittelwerte auf. Zum Zeitpunkt des PostTests lassen sich praktisch nur zwei Leistungsgruppen ausmachen, da sich die Werte beider unterrichtsintervenierten Treatments sowie der nichtintervenierten Treatments untereinander nur marginal unterscheiden. Kontrastiert man hingegen die Formen der Unterrichtsintervention gegeneinander, so ist (bei gegebener Homogenität der Varianzen) ein signifikanter Unterschied nachzuweisen. Dieser Effekt verliert sich allerdings auf Grund der relativ starken Leistungsabnahme des zweifach intervenierten Treatments zum Follow-up-Test. Zu diesem Zeitpunkt ist lediglich ein signifikanter Leistungsvorteil des einfach intervenierten U^+/H^--Treatments gegenüber allen anderen Gruppen zu verzeichnen. Bedingt durch die hohe Leistungsstabilität dieses Treatments über die Zeit ist allerdings die Differenz gegenüber den beiden nicht intervenierten Unterrichtsformen sogar hoch signifikant.

Vergleicht man die differenzierten Analysen der offenen Aufgaben miteinander, so lässt sich dahingehend ein einheitlicher Trend ausmachen, dass unabhängig vom Testzeitpunkt die Leistungsvorteile überwiegend bei den unterrichtlich intervenierten Treatments zu verzeichnen sind, die darüber hinaus durch eine größere Nachhaltigkeit in der Behaltensleistung des erworbenen Wissens gekennzeichnet sind. Die Ergebnisse der Kontrastberechnungen, in denen die jeweils gleichen unterrichtlichen Treatments zusammengefasst und gegeneinander kontrastiert werden, eröffnen als weitere Möglichkeit der Datenauswertung den Einfluss der Hausaufgabenintervention auf die Treatments herauszupartialisieren und lediglich von einem festen Faktor (Unterrichtsintervention) auszugehen. Dieses Vorgehen sollte dann zu klareren und konsistenteren Ergebnissen zwischen den Gruppen führen (s. Kp. 6.4.1).

6.3 Inferenzstatistische Auswertung des Multiple-Choice-Tests

Der Multiple-Choice–Test setzt sich aus insgesamt vierzehn Items mit je vier Attraktoren bzw. Distraktoren zusammen. Da die Anzahl der Attraktoren und Distraktoren nahezu gleichverteilt sind (27 zu 29), kann die Ratewahrscheinlichkeit sowohl auf der Ebene der einzelnen Antwortmöglichkeit als auch für den gesamten Test mit 50% angesetzt werden, weshalb eine Korrektur der Rohdaten entfällt. Punkte wurden sowohl für das Ankreuzen eines Attraktors als auch für das Auslassen der Distraktoren vergeben, so dass maximal 56 Punkte zu erlangen waren. Items, bei denen keine der vier Antwortmöglichkeiten angekreuzt wurden, galten als nicht bearbeitet und wurden daher mit null Punkten gewertet. Dieses Auswertungsverfahren bedingt, dass bei vollständiger Nichtbeantwortung eines Items auch die richtigerweise nicht angekreuzten Distraktoren nicht gewertet werden („harte Auswertung“). Für den letztgenannten Aspekt wurde eine zweite Auswertungsvariante diskutiert, bei der auch bei

vollständiger Nichtbeantwortung eines Items die richtigerweise nicht angekreuzten Distraktoren bepunktet wurden („weiche Auswertung“).

Vergleicht man die mittleren Gesamtpunktzahlen beider Auswertungsvarianten, so zeigt sich erwartungsgemäß die größte Differenz im Pre-Test (M_1= 26,1; M_2= 29,2), da, bedingt durch das fehlende Vorwissen, bei diesem Test besonders viele Items nicht beantwortet wurden. Der Korrelationskoeffizient nach Pearson weist mit r=. 69 auf eine mittelhohe Korrelation hin ($p<0.01$). Die Differenz in den beiden Auswertungsvarianten nivelliert zum Post-Test deutlich (M_1=42,2; M_2=42,5). Der Korrelationskoeffizient deutet mit r=.98 auf eine hochbedeutsame und hoch signifikante ($p<0.01$) Korrelation hin. Die Korrelation beider Auswertungsvarianten hinsichtlich des Lernzuwachses fällt mit r=.84 entsprechend geringer aus, ist aber immer noch als sehr hoch zu betrachten. Die Entscheidung zu Gunsten der „härteren“ Auswertungsvariante führt somit zu keinen statistisch bedeutsamen Konsequenzen, bildet aber die plausibleren absoluten Werte in den Differenzen bzw. Lernzuwächsen zwischen den Messzeitpunkten ab und ist damit das im Zweifel konservativere Verfahren.

6.3.1 Multiple-Choice-Test – Haupteffekt >Zeit<

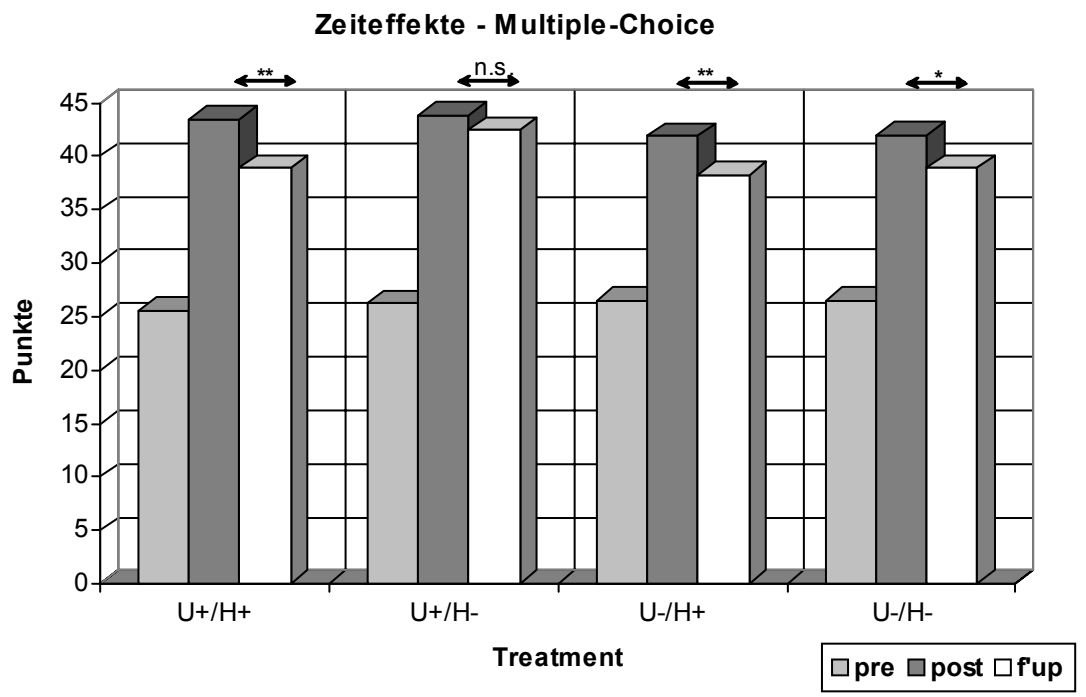
Zur Auswertung des Multiple-Choice-Tests wurden auch hier zunächst Mittelwerte und Standardabweichungen der vier Treatments berechnet, die sich über die drei Messzeitpunkte wie folgt verteilen.

Treatment	Pre-Test		Post-Test		Follow-up-Test		N
	M	SD	M	SD	M	SD	(Σ=169)
U^+/H^+	**25,6**	7,9	**43,5**	5,7	**39,1**	7,6	41
U^+/H^-	**26,2**	6,2	**43,8**	4,9	**42,4**	6,0	44
U^-/H^+	**26,5**	6,4	**41,9**	7,1	**38,3**	7,7	44
U^-/H^-	**26,5**	7,1	**42,0**	6,7	**38,9**	6,9	40

Tab. 6.5: Multiple-Choice - Deskriptive Statistiken der Varianz-Analyse mit Messwiederholung

Die hohen Eingangswerte sind auf die testbedingt hohe Wahrscheinlichkeit, mit der die richtigen Antworten auch geraten werden können, zurückzuführen. Geringfügige Abweichungen der Eingangswerte von den dokumentierten Daten des Treatmentchecks erklären sich durch die versetzungszeugnisbedingte Reduzierung des Datensatzes. Neben dem erwarteten intragruppalen Leistungsverlauf über die Zeit ist auf deskriptiver Ebene eine Reduzierung der Standardabweichung beider unterrichtsintervenierten Treatments im Post-Test auffällig.

Analog zur Auswertung der offenen Aufgaben weisen alle Treatments hoch signifikante Zeiteffekte auf ($p<.001$). Diese Zeiteffekte sind auch für den Vergleich Pre-Test vs. Follow-up-Test zu verzeichnen und damit als über die Zeit stabil zu betrachten.



Anm.: Pre-Post- bzw. Pre-Follow-up-Effekte sind in allen Fällen hoch signifikant (p<.001).

Abb. 6.10 Haupteffekt >Zeit< der vier Treatments im Multiple-Choice-Test

Betrachtet man die Zeiteffekte, wie sie sich im Vergleich von Post-Test zu Follow-up-Test ergeben, so ist Folgendes zu verzeichnen: Drei Treatments weisen signifikante (U^-/H^-) bzw. hoch signifikante (U^+/H^+ und U^-/H^+) Leistungsabnahmen auf, was über eine Latenzzeit von sechs Monaten zu erwarten ist. Dahingegen ist für das Treatment U^+/H^- keine signifikante Leistungsabnahme festzustellen. Der mittlere Leistungswert dieses Treatments ist sechs Monate nach Beendigung der Intervention sogar höher als bei den unterrichtlich nicht intervenierten Treatments unmittelbar nach Abschluss der Intervention. Dies bestätigt die im Rahmen der Analyse der offenen Aufgaben (vgl. Kp. 6.2.2) bereits diskutierte Überlegenheit dieses Treatments.

6.3.2 Interaktionseffekt

Zur Berechnung eines Interaktionseffekts >Treatment*Zeit< wurde eine univariate Varianzanalyse mit Messwiederholung durchgeführt. Bezogen auf eine gleichzeitige Erfassung aller drei Messzeitpunkte zeigt diese Analyse keine Interaktion zwischen dem jeweiligen Treatment und dem Testzeitzeitpunkt ($F_{(1,165)} = 1{,}34$: n.s.). Dieser Befund verändert sich auch dann nicht, wenn man, entsprechend den Auswertungsergebnissen der offenen Aufgaben, die Varianzanalyse auf die Messzeitpunkte Pre-Test und Follow-up-Test beschränkt, um so stärker die interventionsüberdauernden Lerneffekte herausstellen zu können. Auch in diesem Fall ergibt sich bei der Berücksichtigung von zwei festen Faktoren keine statistisch bedeutsame Interaktion ($F_{(1,163)} = 1{,}17$; n.s.). Anders verhält es sich, wenn man auf der Basis der Leistungsdifferenzen zwischen Pre-Test und Follow-up-Test eine One-Way-Anova durchführt. In die-

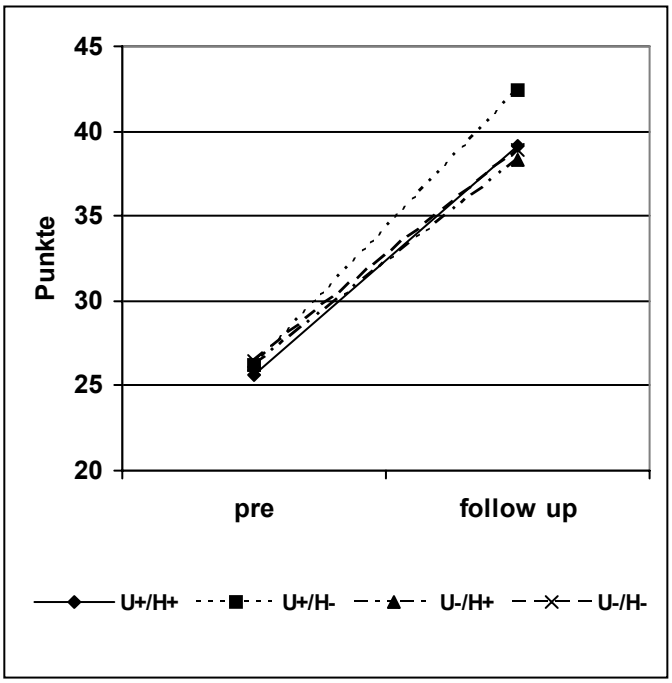


Abb. 6.11: Diagramm zur One-Way-ANOVA; Ergebnisse des Multiple-Choice-Tests

sem Fall ist zumindest ein tendenziell signifikanter Unterschied zwischen den Mittelwerten der Leitungsdifferenzen auszumachen ($F_{(1,165)} = 2,24$; $p<0.1$).

Auf Grund des Verlaufs des nebenstehenden Diagramms lässt sich weiterhin vermuten, dass sich diese tendenzielle Signifikanz im Wesentlichen auf den Einfluss des Treatments U^+/H^- zurückführen lässt. Eine zusätzlich durchgeführte Kontrastberechnung, in der die Leistungsergebnisse des Treatments U^+/H^- gegen die zusammengefassten Ergebnisse der übrigen drei Treatments kontrastiert wurden, bestätigt diese Vermutung ($p<0.05$). Überdies verfehlt der Levene-Test zur Prüfung der Homogenität der Varianzen in diesem Fall die Signifikanzgrenze mit $p= 0.061$ nur knapp. Geht man bei der Berechnung der Kontraste daher alternativ von inhomogenen Varianzen aus, so erreicht der Leistungsunterschied des U^+/H^--Treatments gegenüber den anderen Treatments sogar ein hoch signifikantes Niveau ($p<.001$), wodurch ein weiteres Mal die herausgehobene Lernleistung des Treatments unter Beweis gestellt wird. Dieser Befund wirft allerdings auch die Frage auf, in wie weit die dem Untersuchungsdesign zugrunde liegende Annahme, wonach die Leistungsergebnisse durch zwei festen Faktoren (UV) terminiert werden, gerechtfertigt ist.

6.4 Gruppenvergleiche

Die im Rahmen der Berechnung der Interaktionseffekte dargestellten Ergebnisse (6.2.2 und 6.3.2) werfen auf Grund der überraschend guten signifikanten Leistungen des einfach intervenierten Treatments U^+/H^- sowie der durchgängig schlechten Leistungen des ebenfalls einfach intervenierten Pendants der umgekehrten Faktorenkombination U^-/H^+ generelle Fragen bezüglich der Lernwirksamkeit der Hausaufgabenintervention auf. Daher wurden zur allgemeinen Kontrolle des Einflusses der Hausaufgabenintervention auf die Lernleistung die beiden hausaufgabenintervenierten Treatments sowie die Treatments, in denen eine traditionelle Hausaufgabenbearbeitung realisiert wurde, aggregiert und an Hand der Leistungsergebnisse der offenen Aufgaben varianzanalytisch untersucht. Hierbei ergibt die Varianzanalyse mit Messzeitwiederholung über alle drei Messzeitpunkte (Unterrichtsintervention als Kovariate) keine signifikante Interaktion zwischen den Faktoren >Zeit< und >Hausaufgabenintervention< ($F_{(2,332)} = 2,3$; n.s.). Ebenso zeigt eine zur stärkeren Akzentuie-

rung langfristiger Effekte durchgeführte univariate Varianzanalyse des Lernzuwachses zwischen Pre- und Follow-up-Test keine überzufälligen Interaktionen ($F_{(1,166)}$ = 1,5; n.s.).

Es kann daher davon ausgegangen werden, dass die Art der Hausaufgabenintervention für das Leistungsergebnis im Fachtest nicht von statistisch überzufälliger Bedeutsamkeit ist. Dieser Befund bestätigt sich auch in den Interaktionsdiagrammen der Haupteffekte, wie die folgende Abbildung verdeutlicht:

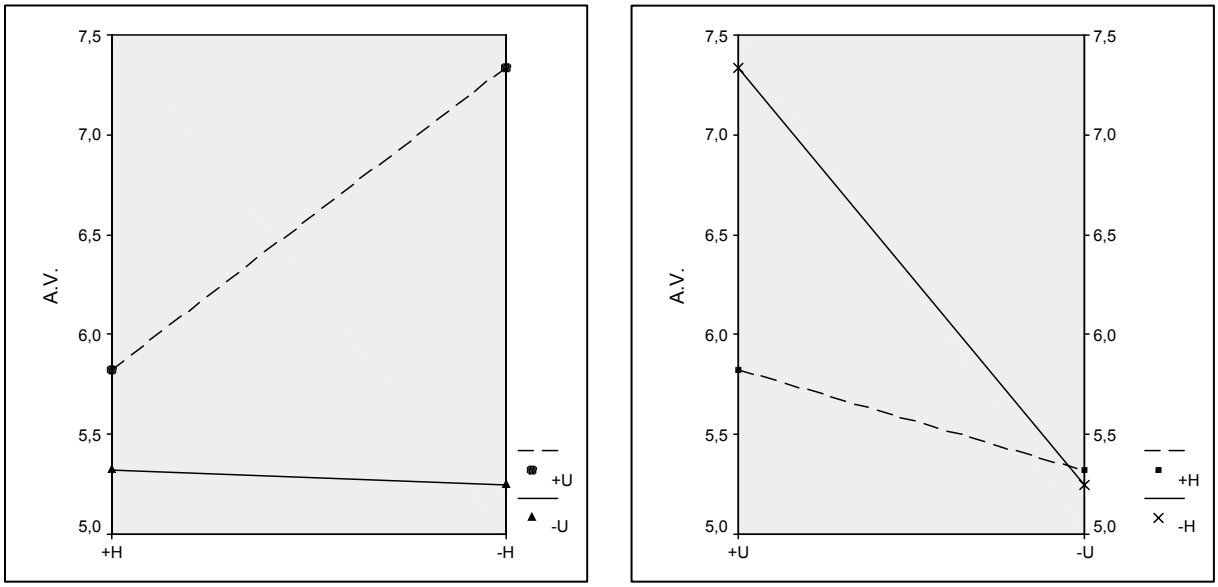


Abb. 6.12: Interaktionsdiagramme der Haupteffekte >Hausaufgabenintervention< (links) und >Unterrichtsintervention< (rechts).

Da im Interaktionsdiagramm des Haupteffekts >Hausaufgabenintervention< die Linienzüge einen gegenläufigen Trend aufweisen, müssen sich die Linienzüge im rechten Diagramm überschneiden, wenngleich der Trend in diesem Fall gleichsinnig ist (vgl. Bortz 1999). Während die Rangfolge der Mittelwerte des Haupteffekts >Unterrichtsintervention< (rechtes Diagramm) für beide Stufen des Faktors >Hausaufgabenintervention< in der Form $U^+ > U^-$ gilt, ist dies umgekehrt für die Rangfolge des Haupteffekts >Hausaufgabenintervention< (linkes Diagramm) nicht festzustellen: Die Aussage $H^+ > H^-$ gilt nur (marginal) für die Linie U^-, wohingegen für U^+ dieser Trend sogar in deutlich umgekehrter Richtung verläuft. Hieraus resultiert eine hybride Interaktion, der zur Folge der Haupteffekt >Hausaufgabenintervention< nicht interpretiert werden kann.

Zusammenfassend belegen die Kovarianzanalyse wie auch die Interpretation der Interaktionsdiagramme, dass eine Lernwirksamkeit der Hausaufgabenintervention nicht nachgewiesen werden kann. Zur besseren Abschätzung des Effekts der Unterrichtsintervention sollen daher im Folgenden die jeweiligen Treatments der gleichen Unterrichtsintervention aggregiert werden und von einer einfaktoriellen Abhängigkeit der Leistungsergebnisse ausgegangen werden.

6.4.1 Kooperative Kleingruppenarbeit vs. lehrerzentrierte Gesprächsführung

Nach Aggregation der Daten der offenen Aufgaben zeigt die Varianzanalyse mit Messwiederholung über alle drei Messzeitpunkte zunächst nur einen tendenziell signifikanten Vorteil des unterrichtsintervenierten Treatments ($F_{(2,334)}$ = 2,63; p<0.1). Dies ist, wie im Rahmen der Auswertung der offenen Aufgaben bereits erörtert, auf die nur geringe Leistungsdifferenz, wie sie sich unmittelbar im Anschluss an die Intervention ergibt, zurückzuführen. Dementsprechend ist bei einer direkten Gegenüberstellung der Leistungsentwicklung vom Pre-Test zum Follow-up-Test von einer Vergrößerung der Effekte auszugehen, da in diesem Fall die deutlich größere Behaltensleistung der in kooperativer Kleingruppenarbeit unterrichteten Lernenden zum Tragen kommt (Abb. 6.13). Die statistischen Kennwerte zeigen mit $F_{(1,167)}$= 5,1 und daraus folgend p= .02 die geringste Irrtumswahrscheinlichkeit der gesamten Leistungstestauswertung an, wenngleich aus der partiellen Effektgröße mit η^2= 0,04 auch in diesem Fall nur eine Varianzaufklärung von 4 Prozent resultiert. (Im Vergleich dazu klärt der Haupteffekt >Zeit< 71 Prozent der Varianz der Innersubjekteffekte bei gleichen Freiheitsgraden und einem F-Wert von 408,1.) Nach Umrechnung von η^2 in die allgemeine Effektgröße ε erhält man mit 0,20 einen Wert, der nach Bortz (1999) auf einen schwachen bis mittleren Effekt hindeutet.

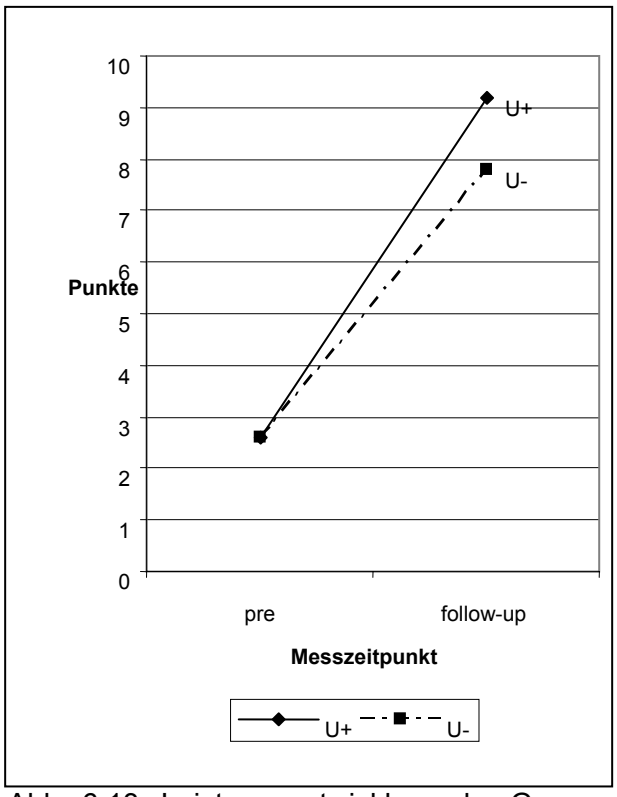


Abb. 6.13: Leistungsentwicklung der Gruppen U^+/U^-

Die entsprechenden Varianzanalysen der aggregierten Daten des Multiple-Choice-Tests weisen in die gleiche Richtung: Auch in diesem Fall zeigt sich die größte Signifikanz, wenn man die Ergebnisse des Pre-Tests mit denen des Follow-up-Tests vergleicht ($F_{(1,167)}$ = 4,47; p<.05; ε = 0,16), während die Varianzanalyse über alle drei Messzeitpunkte erneut nur einen tendenziell signifikanten Effekt aufweist.

Wenngleich die dargestellten Ergebnisse den statistischen Standards zur Überprüfung von Unterschiedshypothesen entsprechen, so sind die Größen der beschriebenen Effekte jedoch wenig anschaulich. Aus diesem Grund wurden zusätzlich die Leistungszuwächse der offenen Aufgaben und des Multiple-Choice-Tests (in Form der Differenzwerte zwischen dem ersten und dritten Messzeitpunkt) aufsteigend geordnet und zur Bildung von Extremgruppen in Terzile geteilt. Auf diese Art werden all jene Personen zusammengefasst, die zu dem Drittel gehören, die über die Zeit den geringsten bzw. den größten Lernzuwachs erfahren haben. Bei der Zusammensetzung der Terzile zeigt sich, dass in beiden Testformen diejenigen Schüler, die den

geringsten Lernzuwachs aufweisen, in etwa zu gleichen Teilen aus der U^+- wie aus der U^--Gruppe stammen. Dahingegen setzt sich die Gruppe der Lernenden mit hohem Lernzuwachs überwiegend aus Teilnehmern der Unterrichtsintervention zusammen. Im Falle des Multiple-Choice-Tests ist dieser Unterschied (Zweidrittel Anteil U^+) nach χ^2-Test hoch signifikant.

	Unteres Terzil			Oberes Terzil		
	Mittlerer Lernzuwachs	Zusammensetzung U^+	Zusammensetzung U^-	Hoher Lernzuwachs	Zusammensetzung U^+	Zusammensetzung U^-
Offene Fragen	1,8 Pkt.	46,4%	53,6%	10,3 Pkt.	60,0%	40,0%
Multiple-Choice	3,7 Pkt.	49,1%	50,9%	23,4 Pkt.	66,7%	33,3%

Tab. 6.6: Zusammensetzung der Extremgruppen in den Leistungstests

Berücksichtigt werden muss hierbei, dass die Terzile nur auf Basis der Leistungs*zunahme* generiert wurden, weshalb keine Angaben über die allgemeine Fachleistung der Schüler eines Terzils gemacht werden können. Das oberste Terzil kann somit nicht als das leistungsstärkste Drittel aller Lernenden bezeichnet werden. Ein leistungsstarker Schüler, der mit guten Ergebnissen im Pre-Test nur noch verhältnismäßig wenige Punkte maximal hinzu erwerben kann, wird daher mit hoher Wahrscheinlichkeit nicht im oberen Terzil zu finden sein. Soll die Lernwirksamkeit der Intervention in Abhängigkeit von der allgemeinen individuellen Fachleistung der Schüler untersucht werden, so muss ein externes Außenkriterium zur Differenzierung von leistungsschwachen und leistungsstarken Schülern (z.B. letzte Zeugnisnote im Fach) herangezogen werden (s. Kp. 6.4.2).

Zum Abschluss der vergleichenden Analyse der beiden Unterrichtsformen sei noch ein kurzer Blick auf die differenzierte Darstellung der offenen Aufgaben geworfen. Abbildung 6.14 zeigt, basierend auf den z-standardisierten Differenzwerten zwischen dem ersten und zweiten bzw. ersten und dritten Messzeitpunkt, den Leistungsvorteil des U^+-Treatments nach Post- und Follow-up-Test für jede Aufgabe differenziert.

Zwar liefern die Ergebnisse keine weiteren globalen Erkenntnisse über die unterschiedliche Lernwirksamkeit der beiden Unterrichtsformen, ihnen ist jedoch eine Reihe von bislang noch nicht identifizierten Aspekten zu entnehmen, die eine differenziertere Interpretation der Auswirkung der Unterrichtsintervention erlauben:

- Sofern signifikante Ergebnisse auf der Ebene der einzelnen Aufgabenauswertung zu verzeichnen sind, so begünstigen sie allesamt das U^+-Treatment.
- In den Ergebnissen der Aufgabe eins wird der Leistungsvorteil des U^+-Treatments über die Zeit noch weiter ausgebaut und führt bis zum Follow-up-Test zu einem hoch signifikanten Unterschied der Mittelwerte. Dies deutet darauf hin, dass insbesondere das einfache, reproduktive Anwenden von Faktenwissen im Regelunterricht nur äußerst zeitnah an die Bearbeitung eines Inhaltsbereichs erbracht werden kann, wohingegen im Rahmen der Unterrichtsintervention, als Folge einer aktiveren Auseinandersetzung mit dem Inhaltsbereich, dieses Faktenwissen längerfristig verankert zu sein scheint.

- Die Leistungsvorteile sind besonders ausgeprägt für die erste und fünfte Aufgabe, also für Aufgaben, deren Bearbeitung entweder basale Anforderungen - Reproduktion von Faktenwissen – an die Lernenden stellt (Aufgabe eins) oder besonders komplexe Kompetenzen (Aufgabe fünf) erfordert, wohingegen Aufgaben des mittleren Schwierigkeitsbereichs nahezu vergleichbar von beiden unterrichtlichen Treatments beantworten werden. Sofern in diesem mittleren Bereich Effekte auftreten, besitzen sie nur tendenziell signifikanten Charakter.

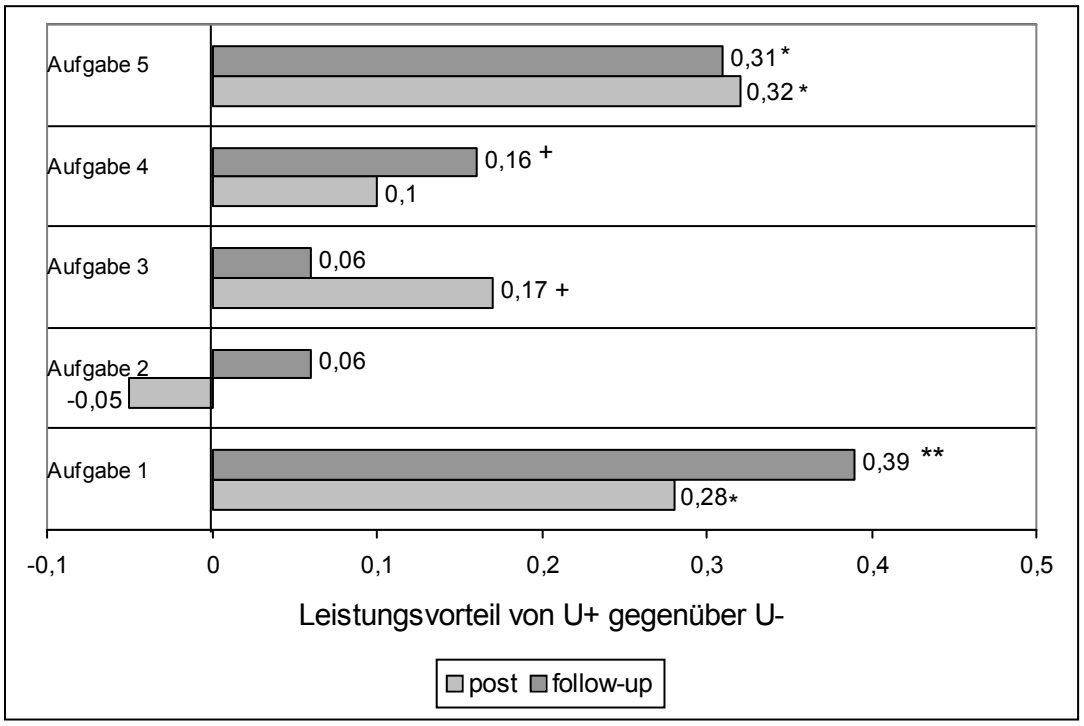


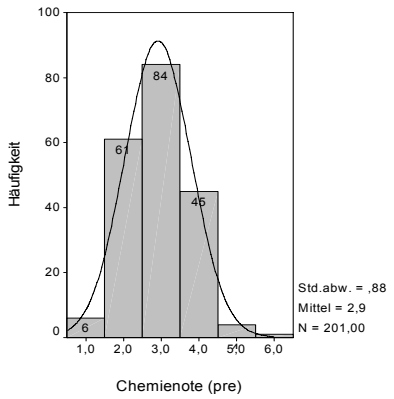
Abb. 6.14: Leistungsverhältnis von U^+ gegenüber U^- nach Post- / Follow-up-Test (offene Aufgaben, z-standardisiert)

Hieraus lässt sich die noch zu prüfende Vermutung ableiten, wonach von der Unterrichtsintervention in besonderem Maße Schülerinnen und Schüler profitieren sollten, die im Regelunterricht entweder zu den leistungsstärkeren oder zu den leistungsschwächeren Lernenden gehören.

6.4.2 Treatmenteffekte in Abhängigkeit von der Fachleistungsperformanz

Die im Rahmen des Hintergrundfragebogens erhobenen Schulleistungen in Form der letzten Zeugnisnote (schriftliche Hauptfächer und Naturwissenschaften) können als Variable zur Prüfung des individuellen Interventionserfolges in Abhängigkeit von der Fachleistung herangezogen werden. Da die Intervention sehr gerichtet auf die Chemiefachleistung abzielt und weniger korrelative Fragen zwischen der einzelnen Fachleistung und der allgemeinen bzw. naturwissenschaftlichen Schulleistung im Vordergrund stehen, wird als Kontrollvariable die letzte Zeugnisnote im Fach Chemie vor Beginn der Intervention herangezogen. Das entspricht in allen Klassen der Halbjahresnote der Jahrgangsstufe 7 (7/I). Vierzehn Schüler machten keine Angabe zu ihrer

Chemienote. Die Notenvergabe folgt klar einer Normalverteilung, wie dem folgenden Diagramm zu entnehmen ist.



Note	$N_{(t1)}$	$N_{(t3)*}$	
sehr gut - gut	67	62	U^+: 31
			U^-: 31
befriedigend	84	65	U^+: 31
			U^-: 34
ausreichend - ungenügend	50	31	U^+: 17
			U^-: 14

* Anzahl der Probanden vom jeweiligen Ausgangsterzil, die zum Messzeitpunkt 3 an der Intervention teilnahmen.

Abb. 6.15: Verteilung der Zeugnisnoten der Probanden im Fach >Chemie< vor der Intervention

Ausgehend von dieser Verteilung wurden drei Extremgruppen gebildet, wobei die Noten >sehr gut< und >gut< zu einem oberen Leistungsniveau, die Noten >ausreichend<, >mangelhaft< und >ungenügend< zu einem unteren Leistungsniveau zusammengefasst wurden. Lernende mit der Note >befriedigend< entsprechen hierbei einem mittleren Leistungsniveau.[9]

Die durch die Versetzungszeugnisse bedingte Reduzierung der Stichprobengröße zum Follow-up-Test (t_3) betrifft die drei Leistungsgruppen erwartungsgemäß in unterschiedlich starkem Ausmaß. Während aus der oberen Leistungsgruppe nur fünf Probanden die Versetzung in die achte Klasse nicht erlangen (respektive Schulwechsel), so sind dies aus der mittleren und unteren Leistungsgruppe jeweils neunzehn Personen. Die Verteilung zwischen unterrichtsintervenierten und nicht-intervenierten Probanden ist innerhalb eines Terzils jedoch annähernd gleich bzw. sogar identisch (obere Leistungsgruppe), so dass ein Vergleich des Lernzuwachses der Treatments innerhalb einer Extremgruppe zulässig ist. Hierzu wurde für jede Leistungsgruppe eine univariate Varianzanalyse mit dem Zwischensubjektfaktor >Unterrichtsintervention< und der abhängigen Variable >Lernzuwachs< (zwischen Pre- und Follow-up-Test) durchgeführt. Die Ergebnisse der Analyse sind in Abbildung 6.16 grafisch dargestellt.

Sehr gute und gute Schüler profitieren demnach tendenziell von der Unterrichtsintervention ($F_{(1,62)} = 2{,}9$; $p = .09$), bei einer partiellen Effektgröße von $\eta^2 = .046$. Lernende des mittleren Leistungsspektrums zeigen keine statistisch bedeutsamen Unterschiede zwischen Interventions- und Kontrollgruppe ($F_{(1,65)} = 0{,}5$; $p = .47$). Dahingegen zeigt die Unterrichtsintervention bei Lernenden des unteren Leistungsspektrums eine

[9] Die kategoriale Einteilung führt zwar zu drei ungleich großen Gruppen, berücksichtigt aber die Tatsache, dass bei einer rangbasierten Terzilbildung das mittlere Terzil durch die Gruppe der Schüler mit einer befriedigenden Note gezogen würde. Siebzehn Schüler dieser Notengruppe würden somit dem unteren Terzil zugeschlagen. Auf Grund der klassenbasierten Generierung der Identifikationsnummer würden diese Schüler alle aus einer Klasse (FH74) und somit einem Unterrichtstreatment stammen. Damit würde der Treatmentvergleich innerhalb dieses Terzils verzerrt.

hoch signifikante Wirkung ($F_{(1,31)} = 13,8$; $p = .001$). Die partielle Effektgröße von $\eta^2 = .32$ (32% geklärte Varianz) zeugt von einem starken Effekt des Faktor >Unterrichtsin-

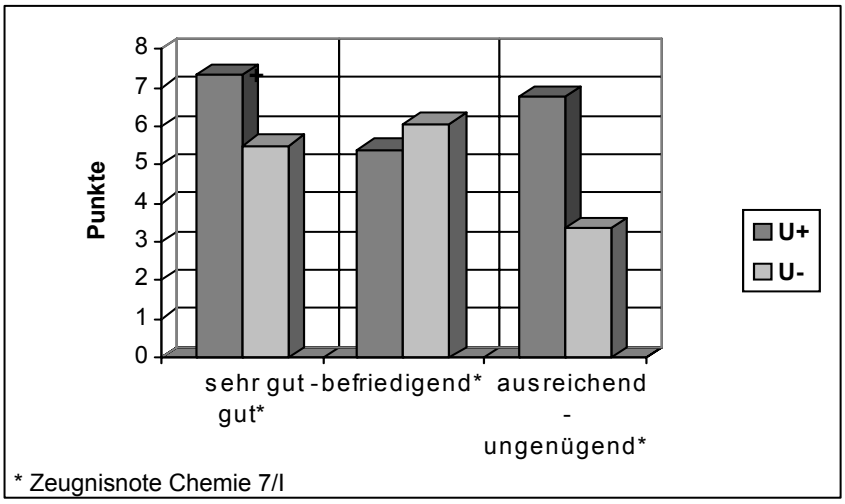


Abb. 6.16: Lernzuwachs (pre –follow-up) differenziert nach Fachleistung vor der Intervention

tervention< für diese Leistungsgruppe.

Eine Interpretation dieses zunächst überraschenden Verhältnisses von Fachleistung und Interventionserfolg könnte darin bestehen, dass sich in den Ergebnissen primär weniger die lernförderliche Stärke der Intervention als vielmehr die Schwäche des Regelunterrichts hinsichtlich der adäquaten Förderung von starken und schwachen Schülern widerspiegelt, wie er in der Kontrollgruppe realisiert wurde. Ein frontalunterrichtliches Lernarrangement ist hingegen so gut auf die zahlenmäßig große Gruppe der Schüler mittlerer Leistungsstärke adaptiert, dass der Erfolg dieser Methode nicht durch die konkrete Form der Unterrichtsintervention gesteigert werden kann. Im Vergleich dazu scheint die Förderung starker bzw. schwacher Schüler durch eine lehrerzentrierte Problemlösungsphase nur suboptimal zu gelingen, während genau diese Klientel durch die Unterrichtsintervention einen Vorteil in der Leistungsentwicklung erhält.

Versucht man ferner die lernwirksamen Faktoren, die dieser spezifischen Form der Kleingruppenarbeit zu Grunde liegen könnten, zu identifizieren, so ist es bedingt durch die großen Kompetenzdifferenzen zwischen leistungsstarken und leistungsschwachen Schülern evident, dass hierfür kein eindimensionaler Wirkmechanismus verantwortlich gemacht werden kann. Betrachtet man zunächst die Situation der leistungsschwachen Schüler, so lässt sich vermuten, dass durch die Unterrichtsintervention eine im Vergleich zum Frontalunterricht breitere kognitive Aktivierung initiiert wird. Die Form der Arbeitsaufträge, die eine theoretische und praktische Auseinandersetzung der Kleingruppe mit dem Sachthema erfordert, bietet dem einzelnen Schüler weniger Rückzugsmöglichkeiten als dies im Vergleich zum lehrerzentrierten Unterrichtsgespräch oder auch in der Kleingruppenarbeit mit dezidierter Versuchsvorschrift - demzufolge das „Experiment“ dann meist von einem einzelnen Schüler durchgeführt wird - möglich ist. Zurückhaltende Schüler, denen zuweilen auch der Mut fehlen mag, sich im Plenumsgespräch zu äußern, bekommen in dieser Lernform

eher die Gelegenheit einer aktiven Auseinandersetzung mit der gegebenen Aufgabenstellung.

Aus der Perspektive der leistungsstarken Schüler könnten auf der anderen Seite das mit der Aufgabenstellung verbundene hohe Anforderungsniveau sowie die Abwesenheit eines den Lernprozess strukturierenden Lehrers gerade diese Schülergruppe in einem über die Möglichkeiten des Frontalunterrichts hinausgehende Maße fördern.

Die lernförderliche Wirkung der Gruppenarbeitsphase würde dieser Interpretation zur Folge zum einen auf der durch die Sozialform bedingten breiten kognitiven Aktivierung und Involvierung einer Vielzahl der Schüler beruhen. Darüber hinaus bewirkt die gleichermaßen offene wie auch komplexe Struktur der Aufgabenstellungen, die aber durch die instruktionalen Hilfen dennoch eine optionale Strukturierung erhalten, eine auch den höheren Fertigkeiten leistungsstärkerer Schülern angemessene kognitive Anforderung.

6.4.3 Einfluss von Kontrollvariablen

Die bislang dargestellten Auswertungen der Leistungstests fokussieren gemäß den zentralen Forschungsfragen auf die Fachleistung in Abhängigkeit vom Unterrichtstreatment und der Lernvoraussetzung. Die vorhandenen Kontrollvariablen erlauben jedoch die Berücksichtigung weiterer, den Lernprozess beeinflussender, Faktoren. Als besonders prominenter Einflussfaktor wird in der Unterrichtsforschung neben dem Geschlecht der Probanden (vgl. Kessels & Hannover 2000) die Bedeutung des Lehrers (s. Kp.2) diskutiert, weshalb im Folgenden der Einfluss dieser beiden Kovariablen auf den Lernerfolg dargestellt wird.

Entwicklung der Fachleistung in Abhängigkeit vom Geschlecht

Zur Identifizierung eines Interaktionseffektes zwischen den Faktoren >Unterrichtsintervention< und >Geschlecht< wurden für beide Testformen eine zweifaktorielle Varianzanalyse mit Messwiederholung durchgeführt. Hierbei zeigt das Ergebnis der offenen Fragen auf der Ebene der Innersubjekteffekte eine signifikante Interaktion zwischen >Treatment> und >Geschlecht< über alle drei Messzeitpunkte ($F_{(2,330)} = 3{,}1$; $p = .046$). Ein Test der Zwischensubjekteffekte ergibt ferner eine hochsignifikante Interaktion zwischen den Gruppen >Geschlecht> und >Treatment< ($p = .004$), wodurch belegt wird, dass Jungen und Mädchen auf die Intervention unterschiedlich reagieren. Eine entsprechende Berechnung auf der Basis des Multiple-Choice-Tests liefert hingegen keine signifikanten Ergebnisse.

Auf Grund der identifizierten Interaktionseffekte wurde der Datensatz in männliche und weibliche Probanden gesplittet, um so die Auswirkung der Unterrichtsintervention auf männliche und weibliche Schüler differenziert betrachten zu können. Im Falle der offenen Fragen zeigt die Auswertung einen hochsignifikanten Leistungsvorteil des unterrichtsintervenierten Treatments bei den Jungen ($F_{(2,162)} = 7{,}2$; $p = .001$; $\eta^2 = .08$). Ausgehend von nahezu identischen Mittelwerten im Eingangstest ($M_{U+} = 2{,}9$;

M_{U-} = 2,6) entwickeln sich diese Werte zum Follow-up-Test deutlich auseinander (M_{U+} = 9,8; M_{U-} = 6,8).

Die analoge Berechnung für die Testergebnisse der weiblichen Probanden weicht deutlich von diesem Befund ab. Die Mittelwerte von Post- und Follow-up-Test zeigen hier sogar geringfügige Vorteile (< 1 Punkt) für die nicht intervenierte Gruppe, wenngleich diese Differenzen nicht signifikant sind ($F_{(2,168)}$ = .37; p = .692). Von einer Interaktion zwischen den Faktoren >Treatment< und >Zeit< kann somit auf der Basis dieser Testform nicht ausgegangen werden.

Die Ergebnisse der Varianzanalysen des Multiple-Choice-Tests zeigen in eine ähnliche Richtung, wenngleich die Differenzen bei dieser Testform nicht so ausgeprägt sind: Für die Jungen findet sich auch hier eine signifikante Interaktion zwischen den Faktoren >Zeit< und >Treatment<, allerdings mit einem kleineren Effekt ($F_{(2,162)}$ = 3,16; p = .045; η^2 = .038). Bei den Mädchen zeigen die Mittelwerte des Multiple-Choice-Tests zum Post- und Follow-up-Test nun höhere Werte zugunsten des unterrichtsintervenierten Treatments (Follow-up-Test: +2,3 Punkte), jedoch ist, wie bei den offenen Fragen, auch dieser Unterschied statistisch nicht bedeutsam (p = .355).

Zusammenfassend belegen die Ergebnisse einen deutlichen Lernvorteil der männlichen Probanden durch die Unterrichtsintervention. Ein derartiger Vorteil ist für die Mädchen in keiner Testform und für keine Unterrichtsform zu finden. Die Ergebnisse bedeuten *nicht*, dass durch die Unterrichtsintervention die Leistung der Mädchen gemindert wird. Innerhalb des intervenierten Treatments unterscheiden sich die beiden Gruppen nicht signifikant. Vielmehr kompensieren die Jungen, die an der Intervention teilnehmen, ein hoch signifikantes Leistungsdefizit, wie es in einem Geschlechtervergleich der nicht intervenierten Gruppe zu verzeichnen ist, wohingegen die Mädchen relativ resistent auf einen Wechsel der Unterrichtsform reagieren. Eine Interpretation dieses Befundes obliegt einem eigenständigen Forschungsfeld und übersteigt die Zielsetzung dieser Arbeit. Es soll daher an dieser Stelle darauf verzichtet werden.

Entwicklung der Fachleistung in Abhängigkeit vom Lehrer

Die an der Intervention beteiligte Anzahl der Lehrer wurde mit n=3 auf ein technisch mögliches Minimum reduziert. Zusätzlich unterrichteten alle Lehrer in unterrichtsintervenierten wie auch in nicht intervenierten Klassen, so dass für die oben dargestellten Treatmentvergleichen interindividuelle Lehrereffekte berechnet werden können. Daraus leiten sich angesichts des für die Unterrichtsqualität bedeutsamen Faktors der Lehrperson folgende Fragen ab:

1. Inwieweit unterscheiden sich die an der Studie beteiligten Lehrer in der erzielten Lernleistung der von ihnen unterrichteten Klassen, unabhängig vom Treatment?
2. Sind in der Umsetzung der beiden Unterrichtsformen intra- und interindividuelle Unterschiede in der Lernleistung zu verzeichnen?

Zur Berechnung der absoluten Lehrereffekte (unabhängig vom Treatment) wurde ebenfalls eine Varianzanalyse mit Messwiederholung und dem festen Faktor >Lehrer< durchgeführt. Für die Testform der offenen Fragen ergibt sich hierbei ein hoch signifikanter Interaktionseffekt (Zeit*Lehrer, $F_{(4,332)} = 3{,}8$; $p = .005$). Das hohe Signifikanzniveau darf in seiner Interpretation nicht dazu verleiten, die Lehrereffekte als bedeutsamer im Vergleich zu den Treatmenteffekten zu betrachten, da beiden Berechnungen unterschiedliche Freiheitsgrade zu Grunde liegen. Misst man den beschriebenen Interaktionseffekt an seiner praktischen Signifikanz, so liegen die Lehrereffekte mit einer Effektgröße von $\eta^2 = .043$ in einer Größenordnung mit den in Kapitel 6.2.2 beschriebenen Treatmenteffekten der offenen Aufgaben.

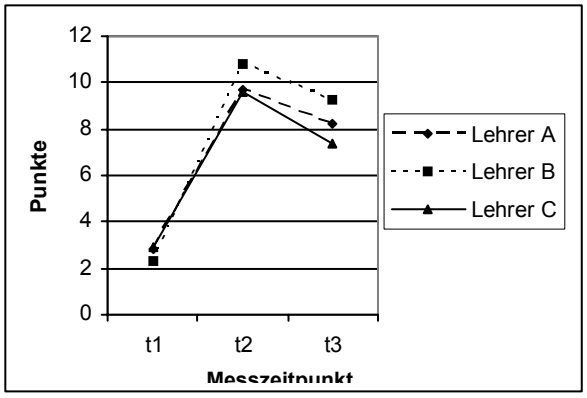


Abb. 6.17: Lehrereffekte „offene Aufgaben“

Bei der Betrachtung der Kurvenverläufe in Abbildung 6.17 fällt augenblicklich das gute Abschneiden der von Lehrer B unterrichteten Schüler sowohl zum Post- als auch zum Follow-up-Test auf. Dahingegen unterscheiden sich die Leistungsergebnisse der von Lehrer A und C unterrichteten Schüler augenscheinlich erst zum Follow-up-Test voneinander, so dass die oben geschilderte Interaktion durch die guten Ergebnisse von Lehrer B im Vergleich zu Lehrer A und C zustande gekommen scheint. Dies wird durch die Ergebnisse der Kontrasteberechnung einer One-Way-ANOVA bestätigt, wie sie in Tabelle 6.7 zusammengefasst sind. Hierzu wurden für alle Lehrer die Leistungsdifferenzen, wie sie sich aus der Kombination der drei Messzeitpunkte ergeben, berechnet. In der Tabelle sind die auf der Kontrasteberechnung basierenden T-Werte aufgeführt; signifikante Werte sind fett dargestellt. Man erkennt, dass sich Lehrer B zu beiden anderen Lehrern sowohl in einem Pre-Post-Vergleich ($T(t_1\text{-}t_2)$) als auch in einem Pre-Follow-up-Vergleich $T(t_1\text{-}t3)$) signifikant unterscheidet. Die Leistungsabnahme zwischen Post- und Follow-up-Test ($T(t_2\text{-}t_3)$) ist in keinem paarweisen Vergleich bedeutsam. Gleiches gilt auch für einen Vergleich zwischen den Lehrern A und C. Um die Effektgröße der hoch signifikanten Differenz des Pre-Follow-up-Vergleichs zwischen Lehrer B und Lehrer C bestimmen zu können, wurde für diesen Fall eine Varianzanalyse mit Messwiederholung über die betreffenden Messzeitpunkte durchgeführt. Die Kennwerte von $F_{(1,122)} = 12{,}1$; $p < .01$ führen zu einer Effektgröße von $\eta^2 = .090$.

Lehrer \ Zeitpunkt	$T(t_1\text{-}t_2)$	$T(t_2 - t_3)$	$T(t_1\text{-}t_3)$
A – B	**2,2***	-.1	**2,2***
B – C	**2,4***	1,0	**3,4****
A – C	.2	.9	1,1

Tab. 6.7: Paarweise Lehrervergleiche der Leistungszunahme (T-Werte)

Eine entsprechende Analyse des Multiple-Choice-Tests führt zu dem gleichen Bild: Auch hier gibt es zu keinem Zeitpunkt signifikante Unterschiede zwischen den Lehrern A und C, wohingegen der Pre-Post-Vergleich und ebenfalls der Pre-Follow-up-Vergleich zwischen den Lehrern A/B und B/C auf signifikante Unterschiede schließen

lässt. Die größte Effektstärke mit η^2 = .098 ist auch hier für den Pre-Follow-up-Vergleich zwischen den Lehrern B und C zu finden ($F_{(1,122)}$ = 13,3; p < .001).

Die Ergebnisse werfen die Frage auf, inwieweit diese Effekte auf eine unterschiedlich erfolgreiche Umsetzung der Unterrichtsintervention beruhen, oder ob sie auf weitergehende, mit der jeweiligen Lehrerpersönlichkeit konfundierte Terminanten der Unterrichtsqualität gründen. Zur Beantwortung dieser Frage wurde zunächst der Erfolg der Unterrichtsintervention auf intrapersonaler Ebene untersucht. In Tabelle 6.8 sind, differenziert für jeden Lehrer, die Mittelwerte für beide Test- und Unterrichtsformen aufgeführt. Die daraus resultierenden Interaktionseffekte (Intervention*Zeit) und Effektgrößen sind grau unterlegt.

	Offene Fragen								Multiple-Choice							
	U+			U-			Interaktion		U+			U-			Interaktion	
	M_{t1}	M_{t2}	M_{t3}	M_{t1}	M_{t2}	M_{t3}	$F_{(2,N-1)}$	η^2	M_{t1}	M_{t2}	M_{t3}	M_{t1}	M_{t2}	M_{t3}	$F_{(2,N-1)}$	η^2
Lehrer A	2,6	10,8	8,3	2,9	8,8	8,2	2,15	.048	27,1	42,3	42,2	27,9	41,5	38,0	$2{,}74^{+}$	.039
Lehrer B	2,4	11,1	10,6	2,2	10,4	7,6	**$7{,}83^{**}$**	**.089**	24,5	45,0	42,3	25,9	43,1	39,5	**$3{,}14^{*}$**	**.060**
Lehrer C	2,9	8,9	7,0	2,8	10,4	7,8	0,90	.022	27,8	42,2	36,2	25,9	40,5	37,7	1,05	.025

Anmerkungen: ** p < .01; * p < .05; p^{+} < 0.1

Tab. 6.8: Interaktionen (Intervention*Zeit) und Effekte auf intrapersonaler Ebene

Betrachtet man die Leistungsergebnisse der von Lehrer A unterrichteten Schüler, so ist für die Testform der offenen Fragen kein signifikanter Unterschied zwischen beiden Unterrichtsformen zu verzeichnen. Wie ein Vergleich der Mittelwerte zum Post-Test (U^+ 10,8 - U^- 8,8) bereits erahnen lässt, ist der Unterschied bei einem Pre-Post-Vergleich allerdings signifikant. Bedingt durch die rasche Leistungsabnahme des U^+-Treatments vom Zeitpunkt t_2 zum Zeitpunkt t_3, nivelliert sich jedoch der Interaktionseffekt über alle drei Messzeitpunkte. Im Gegensatz dazu ist der geringfügige Leistungsvorteil des unterrichtsintervenierten Treatments im Multiple-Choice-Test über die Zeit stabil, weshalb hier bei Lehrer A ein tendenziell signifikanter Interaktionseffekt vorliegt (p = 0.07).

Die Situation bei Lehrer B gestaltet sich sehr übersichtlich: Für beide Testformen zeigen die unterrichtsintervenierten Treatments zeitlich stabile Leistungsvorteile, was zu signifikanten Interaktionseffekten führt; im Falle der offenen Aufgaben sogar auf hoch signifikantem Niveau (p = 0.001), bei einer geklärten Varianz von fast neun Prozent.

Dahingegen deuten die Ergebnisse der von Lehrer C unterrichteten Schüler auf keinen leistungssteigernden Effekt der Unterrichtsintervention hin. In beiden Testformen ist kein signifikanter Unterschied zwischen den Unterrichtsformen über die Zeit feststellbar.

Damit ist geklärt, dass auf intrapersonaler Ebene die Unterrichtsintervention zu einer deutlichen Leistungszunahme der von Lehrer B und zu einer tendenziellen Leistungszunahme der von Lehrer A unterrichteten Schüler führt, wohingegen es bei

Lehrer C zu keinem lernförderlichen Effekt im Sinne einer Leistungssteigerung im Vergleich zur Kontrollgruppe kommt. Ungeklärt ist damit aber noch immer, inwieweit sich der Befund auf die erfolgreiche Umsetzung der Unterrichtsintervention beschränkt oder ob unabhängig davon eine grundsätzliche Divergenz in der Unterrichtseffektivität der Lehrer vorliegt. Diese Frage lässt sich prüfen, indem die Leistungsergebnisse beider Unterrichtsvariationen separiert voneinander auf einen interpersonellen Einfluss varianzanalytisch untersucht werden. Man erwartet hierbei auf Grund der bereits dargestellten Berechnungen für das unterrichtsintervenierte Treatment eine Interaktion zwischen den Faktoren >Lehrer< und >Zeit<. Unklar ist jedoch, ob sich dieser Effekt auch für das nicht intervenierte Treatment nachweisen lässt.

		Offene Fragen	Multiple-Choice
U+	**$F_{(4,164)}$**	**6,27*****	**6,39*****
	η^2	**.133**	**.135**
U-	$F_{(4,162)}$	1,81 n.s.	1,01 n.s.
	η^2	.043	.024

Tab. 6.9: F-Werte und Effektgrößen der Interaktion >Lehrer*Zeit<

Tabelle 6.9 gibt die F-Werte sowie die partiellen Effektgrößen der Interaktion >Lehrer*Zeit< differenziert für jede Unterrichtsform wieder (U+ grau unterlegt). Die zu Grunde liegenden Mittelwerte wurden nicht aufgeführt, da es sich um die gleichen wie in Tabelle 6.8 handelt.

Die Ergebnisse bestätigen die Befunde der intrapersonalen Varianzanalyse, wonach auf der Ebene des U^+-Treatments eine hoch signifikante Interaktion zwischen dem Lehrer und der Zeit zu verzeichnen ist. Berechnet man die Effekte zwischen den Lehrern B und C über die Messzeitpunkte Pre-Test und Follow-up-Test (hier ergaben sich auf der Ebene der Lehrervergleiche die größten Effekte, s.o.), so erhält man nun bei einem F-Wert von $F_{(1,63)} = 19{,}26$ ($p < .001$) eine partielle Effektgröße von $\eta^2 = .23{,}4$ was nach Bortz (1999) einem starken Effekt entspricht.

Dieser substantielle Einfluss der Lehrperson auf die Unterrichtsintervention lässt sich dahingegen für das nicht intervenierte Treatment nicht nachweisen. Weder liegt eine allgemeine Interaktion zwischen den Faktoren >Lehrer< und >Zeit< vor noch ergibt der direkte Pre-Follow-up-Vergleich zwischen den Lehrern B und C in diesem Fall einen bedeutsamen Unterschied ($F_{(1,57)} = 0{,}25$; n.s.). Ein Unterschied bezüglich der Effektivität der Lehrer ist daher für die nicht intervenierten Klassen in keinem Fall festzustellen.

6.5 Zusammenfassung der Ergebnisse der Schulleistungstests

Vor dem Hintergrund der in Kapitel 4.2.1 formulierten zentralen Forschungsfragen stellen sich die zentralen Ergebnisse der Schulleistungstests nach varianzanalytischer Auswertung zusammenfassend wie folgt dar:

- Die aus der Kombination der zwei unabhängigen Variablen entstehenden vier Treatments zeigen für beide Testformen einen signifikanten Lernvorteil für ein unterrichtsinterveniertes Treatment. Die Effekte sind umso deutlicher, wenn in einem direkten Pre-Follow-up-Vergleich auf die Behaltensleistungen sechs Mo-

nate nach Beendigung der Intervention fokussiert wird. Die stärksten Effekte ruft hierbei jedoch in fast allen Fällen das U^+/H^- -Treatment hervor, was generelle Fragen bezüglich der Lernwirksamkeit der Hausaufgabenintervention aufwirft.

- Die aus diesem Befund legitimierten Varianzanalysen zur Überprüfung des Einflusses der Hausaufgabenintervention zeigen keinen bedeutsamen Unterschied zwischen den Hausaufgabenvarianten. Ein durch die Intervention erworbener Lernzuwachs wird somit durch zusätzliche häusliche Intervention weder vergrößert noch geschmälert (s. 2. Forschungsfrage). Es kann im Rahmen dieser Arbeit nicht der Frage nachgegangen werden, inwieweit dieser Befund das Resultat einer suboptimalen Hausaufgaben*bearbeitung* – respektive der Aufgabenkonzeption sowie des Elterntrainings – ist und/oder ob sich die Lernwirksamkeit der Hausaufgabenintervention durch eine inhaltlich und zeitlich intensivere *Auswertung* in der Folgestunde steigern ließe. Dies weist auf eine Fehlstelle in der aktuellen Forschung hin, da in der fachbezogenen Unterrichtsforschung zu diesem Thema bislang keine weiteren Daten vorliegen, die als Referenz herangezogen werden könnten.

Aggregiert man beide unterrichtlichen Treatments und kontrastiert sie gegen die nicht intervenierten Treatments, so ergibt sich folgendes Bild (s. 1. Forschungsfrage):

- Die Wirkmächtigkeit der Unterrichtsintervention ist am größten, wenn man auf die langfristigen, interventionsüberdauernden Testergebnisse schaut. Die unmittelbaren Pre-Post-Vergleiche zeigen dagegen eine geringere Differenz zwischen den Treatments. Dies mag darin begründet liegen, dass das Bereitstellen von erworbenem Fachwissen unmittelbar im Anschluss an eine Thematik sowohl den Lerngewohnheiten der Schüler als auch den Lehrgewohnheiten der Lehrer entspricht. Eine frontalunterrichtliche Unterrichtsituation ist dieser Vermutung zur Folge an eine Leistungsüberprüfung zum Zeitpunkt des post-Tests optimiert, da für reguläre schriftliche Übungen der Inhalt von maximal sechs Unterrichtsstunden repetiert werden muss. Dahingegen gelangen langfristige Leistungsüberprüfungen im Sinne des Follow-up-Tests in der Schulpraxis nicht zum Einsatz. Die hohe Behaltensleistung der unterrichtlich intervenierten Probanden deutet auf eine hohe Schüleraktivität und eine große Prägnanz der Methode hin, die die Schüler dazu befähigt, sich auch nach einem halben Jahr wesentlicher Lerninhalte zu erinnern.
- Die Unterrichtsintervention zeigt insbesondere bei leistungsschwachen und (etwas geringer) bei leistungsstarken Schülern einen größeren Lernerfolg als in der jeweiligen Vergleichsgruppe. Damit fördert die Intervention eine Klientel, welche im Regelunterricht eher über- bzw. unterfordert wird, ohne jedoch die Schüler mittlerer Leistungsstärke zu benachteiligen.

Mit Hinblick auf den Einfluss von Kovariablen wurde die Lernleistung in Abhängigkeit vom Treatment und der Lehrperson bzw. des Geschlechts der Probanden diskutiert. Für beide Variablen zeigen sich signifikante Interaktionseffekte, die auf folgende Ursachen zurückgeführt werden können:

- Innerhalb der Gruppe der männlichen Probanden zeigen die Teilnehmer der Unterrichtsintervention einen größeren Lernerfolg als die Teilnehmer der Kontrollgruppe. Dieser Unterschied ist bei den weiblichen Probanden nicht festzustellen. Vergleicht man jedoch die geschlechtsspezifische Leistung auf der Ebene der Unterrichtstreatments, so zeigt sich im Falle der Unterrichtsintervention kein Leistungsunterschied zwischen den Geschlechtern, wohingegen die Leistungen der Mädchen in der Kontrollgruppe besser sind als die der Jungen. Durch die Intervention wird also ein Leistungsdefizit seitens der Jungen gegenüber den Mädchen kompensiert.
- Ebenso konnten bezüglich der Effektivität der Unterrichtsintervention individuelle Lehrereffekte nachgewiesen werden, wobei in keinem Fall die Unterrichtsintervention zu einer Abnahme der Lernleistung führte. Da es sich in der Jahrgangsstufe sieben um den Chemieanfangsunterricht handelt kann davon ausgegangen werden, dass unabhängig vom Lehrer alle Klassen gleichermaßen wenig Erfahrung im Umgang mit offen konzipierten Schülerexperimenten haben, weshalb ein Trainingseffekt als Begründung für das guter Abschneiden der von Lehrer B unterrichteten Schüler ausgeschlossen werden kann.
- Überdies liegen keine Unterschiede zwischen den Lehrern in der Effektivität des lehrerzentrierten Unterrichts vor. Die Ergebnisse widerlegen die weit verbreitete Annahme, wonach es pauschal „gute“ und „schlechte“ Lehrer gäbe, deren Lehreffektivität sich durch methodische Unterrichtsvariationen per se verbessern ließe. Vielmehr legen die Ergebnisse nahe, methodisch orientierte Lehrerweiterbildungen künftig stärker zu individualisieren und auf eine angemessene Passung zwischen vermittelter Methode und Lehrperson zu achten.

Zum Abschluss der Interpretation der Leistungsergebnisse sei noch eine vergleichende Bewertung der ermittelten Effektgrößen vorgenommen, um näheren Aufschluss über die praktische Bedeutsamkeit der diskutierten Variablen zu erhalten: Die durch die Unterrichtsintervention induzierte Effektgröße liegt in Abhängigkeit von der Testform und der Aggregationsebene in einem Bereich von $0.03 < \eta^2 < 0.05$, wodurch nach Bortz (1999) schwache bis mittlerer Effekte (ε= 0,15 – 0,2) dokumentiert werden. Berücksichtig man allerdings die Vielzahl der Variablen, die auf den Lernerfolg des Unterrichts einwirken (vgl .Kp.2), so ist es nahe liegend, dass an eine derartige Intervention eine geringere Erwartung hinsichtlich der praktischen Signifikanz gestellt werden kann als dies beispielsweise bei Vergleichsstudien der Fall ist, deren statistische Effektgröße allein von Zeiteffekten bestimmt wird (bspw. die Erhebung der Effektivität von Kursen zur Reduzierung der Flugangst).

Eine anschaulich(er)e, wenngleich statistisch nicht abgesicherte Einschätzung von der praktischen Bedeutsamkeit der Unterrichtsintervention erhält man, wenn man Post- und Follow-up-Testergebnisse an den maximal zu erlangenden Punktzahlen (= 100%) misst und auf dieser Basis die von den Schülern erreichten Punktzahlen in Prozentwerte umrechnet. Wendet man dieses, in der Schulpraxis zur Klausurnotenfindung etablierte Verfahren auf die vorliegenden Leistungsdaten an, so ist festzustellen, dass die Leistungsergebnisse der unterrichtlich intervenierten Schüler

durchweg 5 bis 6 Prozentpunkte über denen der nicht intervenierten Schüler liegen. Diese Zunahme entspricht gut einer Notentendenz (3 → 3+), was als anschauliches, aber auch realistisches Maß für den zu erwartenden Erfolg einer Unterrichtsintervention dieses zeitlichen Ausmaßes dienen kann.

7. Auswertung der Kleingruppenvideos

Die im vorangegangenen Kapitel dargestellten und diskutierten Ergebnisse der Leistungstests basieren auf Daten, die durch eine metrische Skalierung sowie eine große Probandenzahl gekennzeichnet sind. Erst diese beiden Bedingungen ermöglichen die wissenschaftliche Prüfung einer (Unterschieds)Hypothese mit Hilfe der dargestellten inferenzstatistischen Methoden. Die zufallskritische Absicherung und die Berechnung der praktischen Bedeutung der dargestellten Mittelwertunterschiede geben jedoch keine Auskunft über den Unterrichts*prozess*, in dem die in den Leistungsergebnissen abgebildeten Kompetenzen erworben wurden (vgl. Wild 2003). Um dennoch Informationen über das sich in den Kleingruppen vollziehende Interaktionsgeschehen zu erhalten, wurde in den unterrichtsintervenierten Klassen jeweils eine Kleingruppe längsschnittlich videographiert. Da die Tonqualität während der Gruppenarbeitsphasen auf Grund der hohen Umgebungsgeräusche zuweilen sehr schlecht war, wurde von allen Unterrichtssituationen ein Volltranskript erstellt. Das Transkriptions-Manual (s. Anhang) wurde in Anlehnung an Müller und Seidel (2001) erstellt. Die anschließende Kopplung des Transkripts an die Videodatei (das so genannte „Kalibrieren“) gewährleistet, dass beim Abspielen des Videos immer der der jeweiligen Filmsequenz zugehörige Text zu lesen ist. Die Kalibrierung und Kodierung der Videos erfolgte mit der Computersoftware CatMovie 4.08 (Wild 1999). Bedingt durch die geringe Fallzahl der Kleingruppen (n=4) wird von einer inferenzstatistischen Auswertung der Videos abgesehen. Da zusätzlich bislang nur wenige Prozessdaten bezüglich der Effektivität von experimentorientierter Kleingruppenarbeit vorliegen, hat dieser Teil der Untersuchung einen ausgeprägt explorativen Charakter.

Mit Blick auf die spätere Kodierung erfolgte die Kalibrierung nicht wie häufig üblich zeitbasiert (time-sampling), sondern nach dem „turnweise Verfahren“ (event-sampling). Ein „turn“ ist im Rahmen dieser Studie definiert als ein auf einen inhaltlichen Aspekt begrenzter, zusammenhängender Sprechakt. In der Kleingruppenphase entspricht ein turn in aller Regel einer Schüleräußerung; ein Sprecherwechsel führt somit immer auch zu einem Turnwechsel. Längere Lehrermonologe, wie sie während der Plenumsphasen zu beobachten sind, mussten immer dann als zwei (selten drei) turns kodiert werden, wenn der Lehrer eine inhaltlich neue Aussage begann, ohne vorher von einer anderen Person unterbrochen zu werden. (Wenn beispielsweise der Lehrer eine Hausaufgabenbesprechung zusammenfasste und sich nahtlos daran die Einführung in das Stundenthema anschloss.)

Mit Hinblick auf den Kategorisierungsvorgang ermöglicht die turnweise Kalibrierung die separate Kodierung jeder einzelnen Schüleräußerung. Dieses, im Vergleich zur zeitbasierten Kodierung eher mikroanalytische und sehr zeitaufwendige Vorgehen, ist notwendig, da die Schüleräußerungen während der Kleingruppenphasen in aller Regel sehr kurz sind und überdies häufig Paralleldiskussionen zwischen zwei Schülerpaaren geführt werden. Ein selbst in sehr kurzen Zeitintervallen kalibriertes alternatives time-sampling (z. B. alle 10 Sekunden) würde daher zu uneindeutigen Kodierentscheidungen führen und damit zu inakzeptablen Interrater-Reliabilitäten.

Abbildung 7.1 zeigt einen screenshot aus dem Transkriptionsdurchgang. Da aus der relativen Kürze der Schüleräußerungen, verbunden mit dem Auftreten von Parallel-diskussionen und Sprecherunterbrechungen, eine extrem hohe Informationsdichte resultiert, wurde die Kodieroberfläche so angelegt, dass jeweils der dem aktuellen turn vorausgehende bzw. folgende turn ebenfalls im Bildschirm zu sehen ist. Auf diese Weise ist es besser möglich, den Kontext des zu kodierenden turns zu erfassen. Während der Transkription wurden zusätzlich der Sender und der Empfänger des jeweiligen turns kodiert.

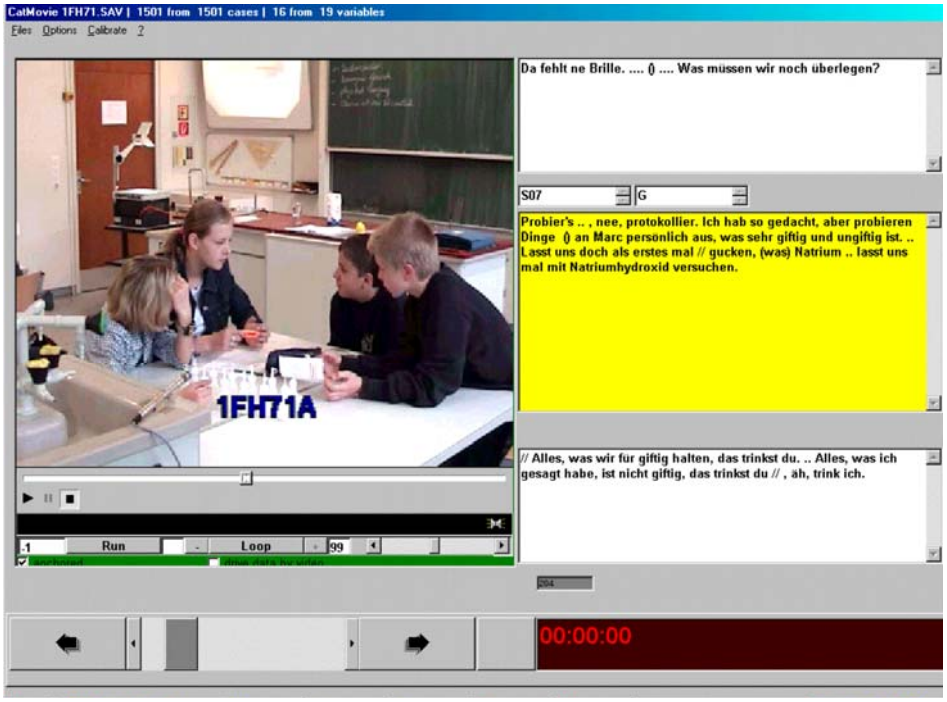


Abb. 7.1: Screenshot aus dem Transkriptionsmodus

An Hand der so aufbereiteten Videodaten wurden dann die Kleingruppensituationen näher analysiert. Zentrale Funktion des hierfür entwickelten Kategoriensystems ist es, den inhaltlichen Lernfortschritt einer Kleingruppensituation an Hand fachdidaktischer Konstrukte aufzuzeigen, weshalb das System den Titel „*Inhaltliche Progressionsanalyse (IPA)*“ trägt. Die Entwicklung der inhaltlichen Progressionsanalyse erfolgte mit Hilfe der Unterrichtsaufnahmen der Vorstunde (s. Kp. 5.2.1), aufbauend auf der Arbeit von Walpuski (2002).

7.1 Inhaltliche Progressionsanalyse (IPA) der Gruppenarbeitsphase

Ausgelöst durch die Ergebnisse der TIMSS-Videostudien (Stigler et al. 1999, Hiebert et al. 2003) wurden (und werden) in den letzten Jahren im deutschsprachigen Raum zahlreiche Folgestudien durchgeführt, die, in der Tradition der TIMSS-Videostudie, spezifische Ablaufmuster des Mathematik- und Physikunterrichts (so genannte „scripts“) identifizieren und vergleichen (Seidel 2002, Reusser & Pauli 2003, Reyer 2003). Vergleichbare Studien für den Chemieunterricht wurden bislang nicht durchgeführt; lediglich Schallies (1993) berichtet von einer Studie, bei der zur qualitativen

Rekonstruktion von 'Experimentierphasen mit freien Lösungsmöglichkeiten' Videomitschnitte angefertigt wurden, ohne diese jedoch kategoriengeleitet zu kodieren.

Im Zuge der oben genannten Videostudien wurden umfangreiche Kodiersysteme entwickelt, mit deren Hilfe die Organisationsform des videographierten Unterrichts theoriegeleitet analysiert werden kann (vgl. zur Übersicht Seidel 2003, Stigler & Fernandez 1995). Die Funktion derartiger Kodierschemata ist es, Informationen des beobachteten Unterrichts über einen zeitlichen Verlauf (meist 45 Minuten) zu erlangen. Da es sich bei der hier vorliegenden Studie allerdings um eine Interventionsstudie handelt, wird die Organisationsform des Unterrichts, Art und Zeitpunkt eines Aktions- und Sozialformwechsels sowie die gesamte Phasierung einer jeden Stunde dezidiert vorgegeben. Demzufolge ist die Kodierung der Stunde an Hand so genannter „Sichtstrukturen" (Seidel 2002), die Aufschluss über die Orchestrierung der unterrichtlichen Organisation geben, nicht zielführend.

Ein speziell auf die Choreographie eines experimentgestützten Physikunterrichts ausgerichtetes Kategoriensystem legen Tesch und Duit (2001) bzw. Tesch (2003) vor (Abb. 7.2).

Mit Hilfe des Systems lassen sich neben konkreten Angaben zur Experimentierphase, der Art der Datenerfassung, Organisationsform, Ablaufform und Funktion des Experiments auch weitere Aspekte, wie die Einbettung des Experiments in den Unterricht sowie die Art und Weise der Vor- und Nachbereitung des Experiments erfassen.

Da die Kleingruppenarbeitsphasen der Studie zentral von der Entwicklung und Durchführung eines Experiments geprägt sind, bietet dieses System eine erste Näherung für eine Kategorisierung. Dennoch wird auch in diesem Fall durch den präskriptiven Charakter der Studie die Kategorie nahezu aller Facetten festgelegt. In der Darstellung des Kategoriensystems von Tesch (2003) sind die in dieser Studie realisierten Kategorien grau unterlegt (Abb. 7.2).

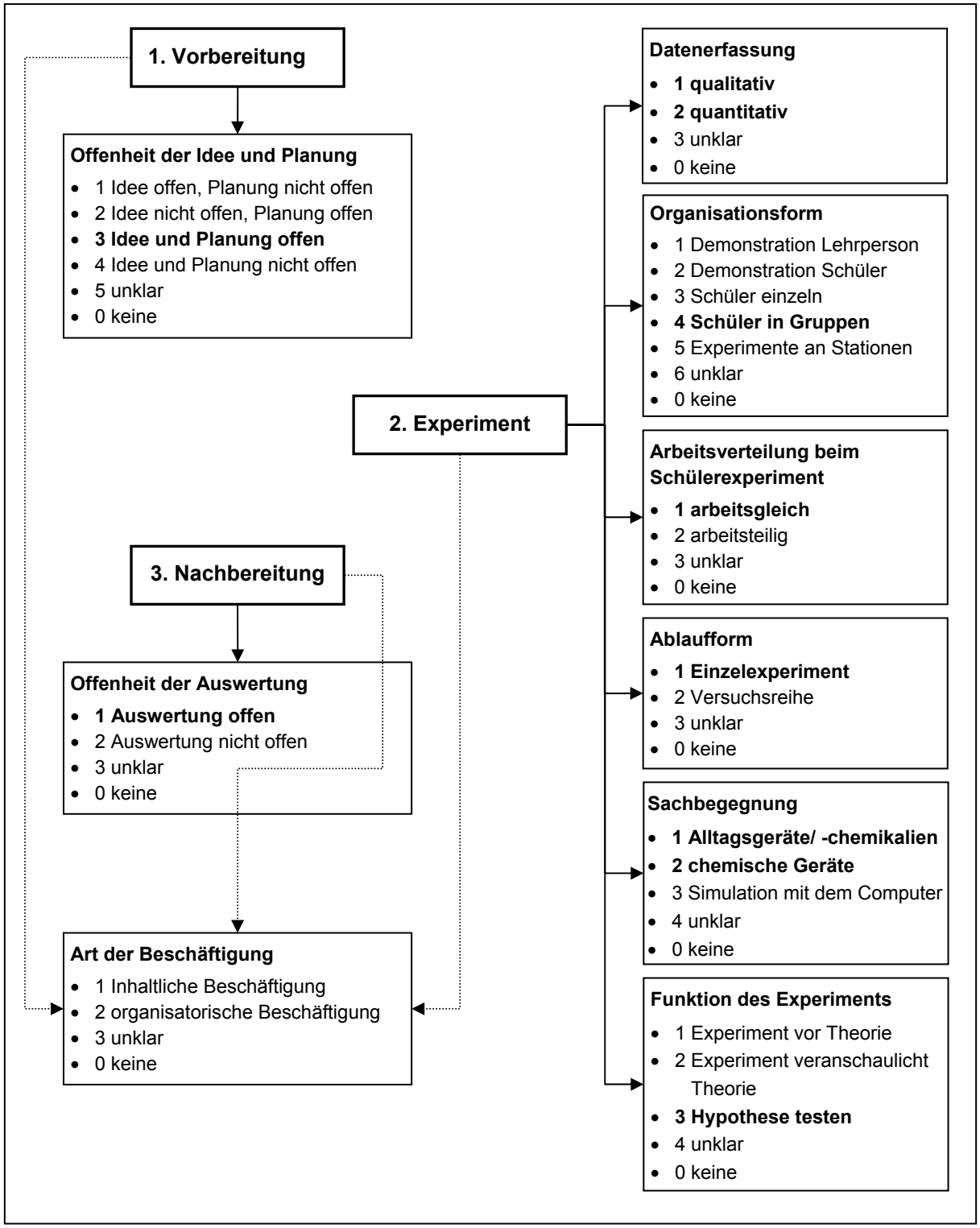


Abb. 7.2: Kategorien des Experimentierens. Verändert nach Tesch (2003) – In der Intervention realisierte Kategorien grau unterlegt

Das Konzept „Unterrichten mit Interaktionsboxen“ sieht innerhalb aller Kleingruppensituationen eine offene, d. h. nicht vom Lehrer vorgegebene Planung des Experiments vor. Die Datenerhebung verläuft überwiegend qualitativ (Messung des pH-Werts halb-quantitativ).

Ebenso sind durch die Methode die Kategorien „Organisationsform“ und „Ablaufform“ sowie die „Arbeitsverteilung“ festgelegt. Die Sachbegegnung beruht gleichermaßen auf chemischen Geräten wie auch auf Alltagschemikalien. Die Funktion des Experiments sollte generell im Testen von Hypothesen bestehen. Die Auswertung des Experiments verläuft innerhalb der Kleingruppen wiederum offen, wenngleich nach Abschluss der Kleingruppenphase eine vergleichende Auswertung im Plenum erfolgt, die dann vom Lehrer strukturiert und gelenkt wird.
Das Kategoriensystem von Tesch eignet sich von daher gut für eine umfassende, standardisierte Charakterisierung der in der Kleingruppenphase realisierten Experimente. Auf Grund des präskriptiven Charakters der Studie ist jedoch weder von einer inter- noch von einer intragruppalen Varianz bei der Anwendung des Systems auf die Kleingruppenphasen auszugehen. Dahingegen bietet die zusätzliche Facette „Art der Beschäftigung“ einen ersten Ansatz zur Kategorisierung der Gruppenarbeitsphasen. Im Gegensatz zum System von Tesch soll diese Facette neben der Phase der Vor- (1.) und Nachbereitung (3.) auch auf das Experiment (2.) selbst angewendet werden. Da in dieser Unterrichtsvariante der methodische Schwerpunkt der Kleingruppenarbeit auf dem gemeinsamen Kommunizieren der Entwicklung und Auswertung eines Experiments liegt, wird zur Betonung dieses Aspekts die Facette als >Fokus der Argumentation< (kurz >Fokus<) bezeichnet.

Im Rahmen der Kodierung dieser Facette wird jeder turn der Kleingruppenarbeitsphase dahingehend bewertet, welche Funktion, im Sinne von >inhaltlich<, >organisatorisch< oder >Sonstiges<, ihm im Kontext der Arbeitsphase zugeschrieben werden kann. Zur Gewährleistung der Reliabilität der Kodierung ist es von besonderer Bedeutung, trennscharfe Indikatoren für jede Kategorie zu formulieren. Unter „Indikatoren“ versteht man in diesem Zusammenhang „unmittelbar messbare Sachverhalte, welche das Vorliegen der gemeinten aber nicht direkt erfassbaren Phänomene […] anzeigen“ (Backhaus et al. 2003).

Mit der Kodierung des Fokus der Kommunikation werden gleichsam zwei Zielsetzungen verfolgt: Zum einen wird es so möglich, Erkenntnisse über die Zeitnutzung der Kleingruppen, das Verhältnis von organisatorischen zu inhaltlichen Äußerungen sowie den jeweiligen Anteil an Nebenkommunikation während der Gruppenarbeitsphasen zu erhalten. Zum anderen stellt diese Kodierung die Grundlage für den weiteren Kodiervorgang dar, da im Folgenden allein mit den inhaltlichen turns weitergearbeitet werden soll. Wie Abbildung 7.3 zu entnehmen ist, wird jeder inhaltliche turn in drei weitere Unterfacetten kodiert. Eine weitere Differenzierung der organisatorischen und sonstigen Äußerungen findet nicht statt. Die Kodierung erfolgt innerhalb aller Facetten disjunkt.

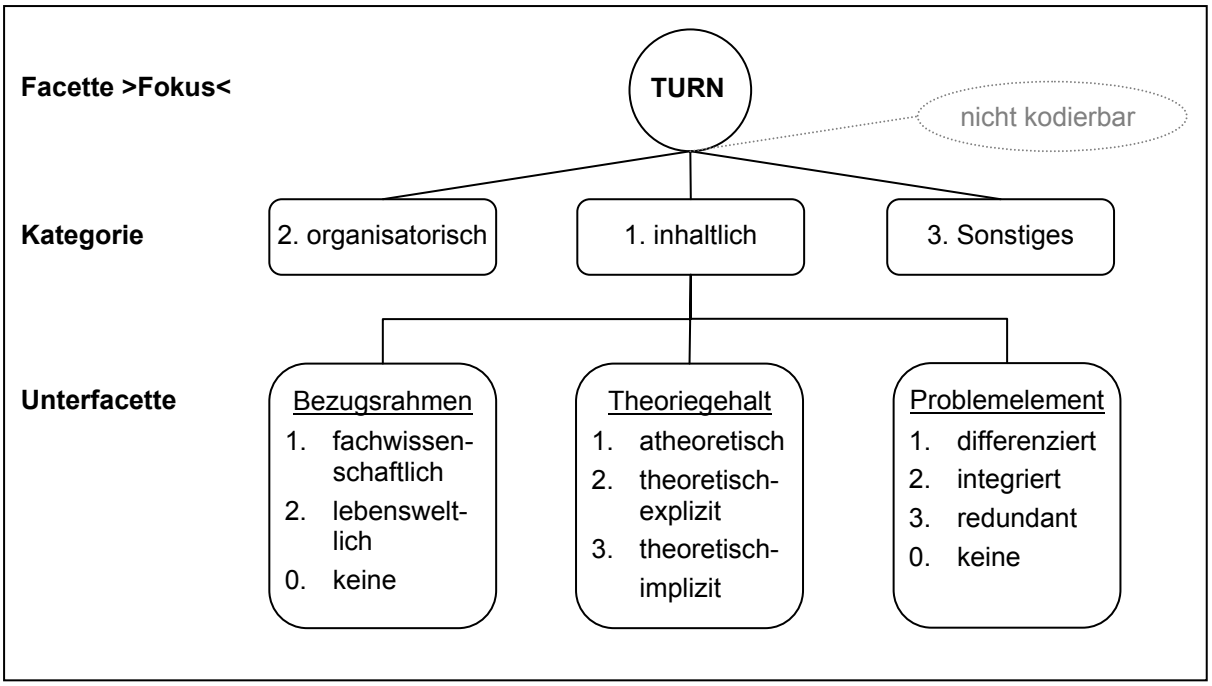


Abb. 7.3: Kategorienschema der „Inhaltlichen Progressionsanalyse“ (IPA)

Die inhaltliche Progressionsanalyse stellt ein aus fachdidaktischer Perspektive konzipiertes Instrumentarium zur qualitativen Bewertung komplexer, kooperativer Aufgabenstellungen dar. Als zentrale Beurteilungsaspekte werden der Grad der Theorieleitung der schülerinternen Argumentation, der zu Grunde liegende kognitive Bezugsrahmen sowie das Verhältnis von Integration und Diversifizierung beim kommunikativen Problemlösen erfasst. Die so erhaltenen Daten sollen schließlich auch als Hinweis für die Qualität schüleradäquater, kooperativer Aufgabenstellungen dienen. Im Detail erklären sich die Unterkategorien wie folgt:

7.1.1 Bezugsrahmen

Im Rahmen dieser Kodierung soll erfasst werden, inwieweit den inhaltlichen Äußerungen sprachlich ein Bezugsrahmen zugeordnet werden kann, der auf eine fachwissenschaftlich bzw. lebensweltlich geprägte Schülervorstellung hindeutet.

Auf der Beobachtungsebene sind die Indikatoren eng über den Inhaltsbereich „saure und basische Lösungen“ definiert: Die lebensweltlich geprägte Schülervorstellung attribuiert beispielsweise >Säuren< unabhängig von ihrem pH-Wert als prinzipiell gefährlich und ätzend, wohingegen >Basen< eher funktional als Reinigungsmittel definiert werden. Eine ebenfalls lebensweltliche Annahme ordnet allein den sauren Lösungen einen pH-Wert zu, dessen Zahlenwert mit der „Gefährlichkeit“ einhergeht. Säuren besitzen dieser Vorstellung zur Folge einen großen pH-Wert, wohingegen die Existenz des pH-Werts von Basen häufig zunächst über den Unterricht vermittelt werden muss und, so dies geschieht, dieser dann als tendenziell niedrig eingestuft wird.

Eine im Verlauf der Intervention abnehmende Artikulation von lebensweltlichen Schülervorstellungen bei gleichzeitiger Zunahme von fachwissenschaftlich angemessen

Äußerungen würde daher eine über die Ergebnisse der Schulleistungstests hinausgehende Qualität in der Lernprogression der Unterrichtsreihe aufzeigen.

7.1.2 Theoriegehalt

Mit dieser Unterfacette soll erhoben werden, in welchem Ausmaß die inhaltlichen Schüler-Schüler-Äußerungen durch eine theoriegeleitete Argumentation geprägt sind. Die bildungstheoretische Legitimation dieser Kategorie leitet sich unter anderem aus der im Kontext der Scientific Literacy-Debatte (Bybee 1997 & 2002) sowie der PISA-Studie aufgekommenen Diskussion um die Formulierung von naturwissenschaftlichen Kompetenzen ab. Prenzel et al. (2001) führen als einen von insgesamt vier Prozessen der naturwissenschaftlichen Grundbildung das Kommunizieren naturwissenschaftlicher Argumente an. Hierunter versteht man unter anderem die Fähigkeit, „bezogen auf eine bestimmte Situation und vorliegende Daten […] eine Argumentation zu entwickeln“ (ebenda, S. 199). Dieser Facette wird ein ausgesprochen hoher lernpsychologischer Schwierigkeitsgrad zugesprochen, da eine konkrete Ausformulierung nur für die beiden obersten Kompetenzstufen erfolgt (konzeptuelle und prozedurale Grundbildung).

Auf der Beobachtungsebene wird in dem hier vorliegenden System der Begriff >Theorie< sehr weit gefasst und beschreibt sprachstrukturell eher eine *begründete Idee*, welcher eine nicht explizit geäußerte subjektive Theorie zu Grunde liegt. Zielsetzung der Kodierung ist es, Problemlösevorschläge, die auf planlosem Ausprobieren beruhen und im Folgenden als >atheoretisch< bezeichnet werden, von solchen abzugrenzen, die sich durch eine systematische, theoriegeleitete Herangehensweise auszeichnen. Dieser Ansatz fußt auf der Typologisierung von Argumenten nach Toulmin (1996, Abb. 7.4), wonach eine Behauptung oder *Konklusion* zu trennen ist von den objektiven Daten, die zur Begründung der Tauglichkeit einer Behauptung herangezogen werden können. Um motivieren zu können, dass bestimmte Daten als Begründung für eine Konklusion dienen können, bedarf es einer weiteren Variablen, der Schlussregel. Die Schlussregel klärt gewissermaßen den Zusammenhang zwischen der Date und der Konklusion. Zur Veranschaulichung sei dieses Schema auf ein Beispiel der Unterrichtsreihe “saure und basische Lösungen“ übertragen: Die Schüler erhalten die Aufgabe, eine ihnen unbekannte Lösung (Säure) mit einem Angebot verschiedener Lösungen zu neutralisieren. Folgendes Schema gibt eine anzustrebende Argumentationskette wieder:

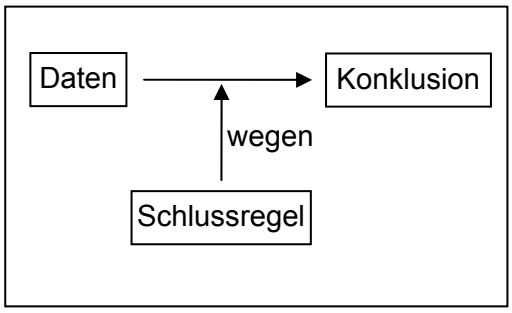


Abb. 7.4: Argumentationsschema nach Toulmin (1996)

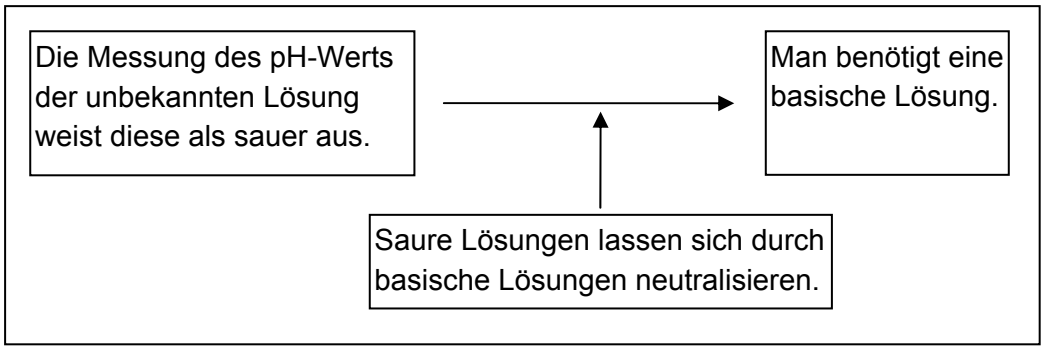


Abb. 7.5: Beispiel zu Toulmins Argumentationsschema

Ein atheoretischer Lösungsansatz zeichnet sich dahingegen dadurch aus, dass er weder auf Evidenzen (Daten) aufbaut noch eine Schlussregel enthält, sondern stattdessen eine (scheinbar) unmotivierte Behauptung aufweist. (Z. B.: „Nimm doch einfach mal eine Base.“)

Das bedeutet allerdings nicht, dass >atheoretisch< kodierte turns gleichsam den (spekulativen!) Rückschluss auf theoriefreie Kognitionen des sprechenden Schülers zulassen. Vielmehr soll durch diesen Kodierdurchgang das Ausmaß des offen kommunizierten, theoriegeleiteten Handelns der gesamten Gruppe erfasst werden.
Wie die Kodierung der Vorstunde ergeben hat, liegt zuweilen die Situation vor, dass die Schlussregel zu einer Konklusion erst einige turns später ausgesprochen wird. Daher wurde der Gesprächsverlauf nach jedem atheoretischen turn mit Hilfe das Transkripts auf diesen Aspekt untersucht und, sofern eine Begründung zu einem späteren Zeitpunkt nachfolgte, dieser turn in eine dritte Kategorie „theoretisch implizit“ umkodiert. Walpuski (2002) führt hierzu folgendes Beispiel an:

S_1: „Lass uns doch die Kreide als Docht nehmen.“
S_2: „Genau, die ist saugfähig.“

Die vom ersten Schüler getätigte, atheoretische Äußerung erhält erst durch den Zusatz von Schüler S_2 die Schlussregel, weshalb in diesem Fall beide turns inhaltlich zusammenzufassen und als theoretisch implizit zu kodieren sind.

7.1.3 Problemelemente

Der Terminus >Problemelement< wurde von Boos (1996) im Rahmen der Analyse komplexer Entscheidungsfindungen in Kleingruppen definiert als *die Gesamtheit aller von den Diskussionsteilnehmern explizit genannten Bestandteile des zu lösenden Problems.* Problemelemente stellen demnach die inhaltlich definierten Schlüsselbegriffe zu einem Diskussionsthema dar; die Anzahl der Problemelemente eines Inhaltsbereichs ist folglich begrenzt.

Bezogen auf die Unterrichtsreihe „saure und basische Lösungen“ sind als Problemelemente alle Gegenstände der Interaktionsboxen sowie hieraus abgeleitete Begriffe definiert. Weiterhin sind alle für den Inhaltsbereich relevanten abstrakten Begriffe, wie pH-Wert oder Neutralisation zu den Problemelementen zu zählen. Eine Übersicht aller im Rahmen der Unterrichtsreihe vorkommenden Problemelemente findet sich in Anhang G.

Alle Problemelemente sind kategorial hinsichtlich folgender Merkmalsausprägung zu unterscheiden (vgl. Boos 1990):

- Differenziertheit: In diese Kategorie werden alle erstmalig im Rahmen einer Kleingruppenphase genannten Problemelemente kodiert. Die Anzahl der differenzierten Problemelemente pro Zeiteinheit gibt demnach Aufschluss über die inhaltliche Vielfalt der Diskussion.
- Integration: Sofern ein Problemelement bereits in eine Kleingruppendiskussion eingebracht wurde, ist bei dessen wiederholter Nennung zu entscheiden, welche Funktion diese Wiederholung im Argumentationsprozess einnimmt. Die Kodierung *integrativ* erfolgt, wenn das Problemelement mit andern bereits genannten in Bezug gesetzt wird und somit eine Vernetzung der Argumente initiiert wird. Morguet (1992) spricht bei diesem Merkmal daher auch von Vernetztheit statt von Integration. Zu erwarten ist, dass zu Beginn einer Diskussion die differenzierten, zum Ende die integrativen Problemelemente überwiegen.
- Redundanz: Problemelemente, die ein zweites Mal im gleichen Kontext, mit der gleichen Eigenschaftszuschreibung versehen werden, sind als redundant zu kodieren. Inhaltlich wird damit die wiederholte Kennzeichnung derselben Information kodiert, ohne jedoch eine informationstheoretische Angabe darüber machen zu können, inwieweit es sich hierbei um eine förderliche oder um eine leere Redundanz handelt.

7.2 Ergebnisse der Inhaltlichen Progressionsanalyse

Nach Abschluss der Kalibrierungs- und Transkriptionsarbeiten wurden für alle 24 Kleingruppensituationen mit Hilfe des Transkripts der Beginn und das Ende der Gruppenarbeitsphase bestimmt. Der Beginn der Phase wurde hierbei definiert als erster turn der videographierten Kleingruppe nach dem letzten turn des Lehrers an das Plenum. Ebenso wurde das Ende der Gruppenarbeitsphase als letzter turn vor der ersten Äußerung des Lehrers zur Einleitung der Auswertung im Plenum festgelegt. Schülerinterne Kommunikation innerhalb der Kleingruppe, die nach diesem Zeitpunkt erfolgte, wurde nicht mehr kodiert.

Auf diese Weise ergibt sich folgende Gesamtsumme der zu kodierenden turns aller Kleingruppen:

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4	**Σ**
2127	2553	2464	1707	**8851**

Tab. 7.1: Gesamtturns der Gruppenarbeitsphasen

Die geringe Zahl der Gesamtturns von Gruppe vier erklärt sich aus einer tontechnischen Störung während der zweiten Unterrichtsstunde, in deren Folge nur 71 turns dieser Kleingruppenphase kodiert werden konnten. Im Vergleich dazu beläuft sich der entsprechende Mittelwert der übrigen drei Kleingruppen in der zweiten Unterrichtsstunde auf 487 turns. Addiert man zu Vergleichszwecken der vierten Gruppe

400 turns theoretisch hinzu, so liegt die Gesamtzahl mit über 2100 turns in einem mit den anderen Gruppen vergleichbaren Rahmen.

Die turns der Gruppenarbeitsphase wurden anschließend gemäß dem Kategorisierungsschema zunächst hinsichtlich ihrer funktionalen Orientierung kodiert. Die somit identifizierten inhaltlichen turns verteilen sich auf die Kleingruppen wie folgt:

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4	**Σ**
913	1323	1312	1016	**4564**

Tab. 7.2: Summen der inhaltlichen turns der Gruppenarbeitsphasen

Insgesamt gingen damit 8851 turns in den ersten Kodierdurchgang und 4564 turns in den zweiten Kodierdurchgang ein. Zur Beurteilung der Kodierungsreliabilität wurde ein Teil aller turns doppelt kodiert und die Beurteilerübeinstimmung nach Cohen's Kappa berechnet. Da im Gegensatz zu zeitbasierten Kodierungen, bei denen im Allgemeinen 10% aller Ereignisse doppelt kodiert werden, bei der ereignisbasierten Kodierung bislang nur wenig Erfahrung bezüglich der Kodierungsreliabilität vorliegt, wurden in beiden Kodierdurchgängen 20% aller turns von jeder Unterrichtssituation doppelt kodiert. Die doppelt zu kodierenden Sequenzen wurden zufällig aus den Transkripten gezogen.

Für die vier Kategorien der Inhaltlichen Progressionsanalyse ergeben sich demnach folgende Interraterübereinstimmungen:

Facette/ Unterfacette	turns	Wertebereich der Übereinstimmungskoeffizienten (Cohen's Kappa, ungewichtet)
Fokus	8851	$.78 < \kappa < .89$
Bezugsrahmen	4564	$.63 < \kappa < .95$
Theoriegehalt	4564	$.69 < \kappa < 1.0$
Problemelement	4564	$.62 < \kappa < .89$

Tab. 7.3: Beurteilerreliabilität der Inhaltlichen Progressionsanalyse

Betrachtet man die für die Güte der Kodierung relevanten unteren Grenzen der Wertebereiche, so ergibt sich ein insgesamt zufrieden stellendes Bild bezüglich der Reliabilität des Kategoriensystems. Wenngleich die Literaturlage keine einheitlichen Empfehlungen hinsichtlich der Gütekriterien von Übereinstimmungskoeffizienten hergibt, so sind die hier zu diskutierenden Werte als Hinweis auf eine statistisch zuverlässige Kodierung zu betrachten. Wirtz und Caspar (2002) betrachten einen Kappa-Wert von 0.6 bis 0.75 als Indikator für eine gute, einen Wert größer 0.75 als Indikator für eine sehr gute Übereinstimmung. Beide Autoren verweisen jedoch auch auf Frick und Semmel (1978), welche prinzipiell erst bei einem Kappa-Wert größer 0.75 von einer zufrieden stellenden Übereinstimmung sprechen. Landis und Koch (1977) betonen daher den willkürlichen Charakter derartiger Richtwerte und plädieren für „weichere" Intervalle zur Gütebeurteilung von Kodierübereinstimmungen. Die Autoren sprechen bei einem Wertebereich von 0.41 – 0.60 von „moderaten", bei einem Bereich von 0.61 – 0.80 von „substanziellen" und bei Koeffizienten größer 0.81 von „nahezu einwandfreien" („almost perfect") Übereinstimmungen. Unabhängig davon, wel-

cher Güteempfehlung man folgt ist jedoch festzustellen, dass die in dieser Studie erzielten Reliabilitätsmaße als hinreichend für eine weitere Interpretation der Daten anzusehen sind.

7.2.1 Ergebnisse der Facette >Fokus<

Nach Kodierung der Facette >Fokus< ergibt sich die in Abbildung 7.6 dargestellte Verteilung der funktionalen Orientierung der Kleingruppengespräche, aufsummiert über alle Kleingruppen und alle Unterrichtsstunden. Die häufigste Kodierung erfolgte in der Kategorie >inhaltlich<, was allerdings auf Grund des Bildungsauftrags von Unterricht auch zu erwarten ist. Überraschen mag mehr der mit 51,6% verhältnismäßig geringe prozentuale Anteil, der für eine inhaltliche Kommunikation aufgebracht wird, so man das Ergebnis dahingehend formuliert, dass nahezu die Hälfte der Gesamtkommunikation auf nicht-inhaltliche Aussagen entfällt.

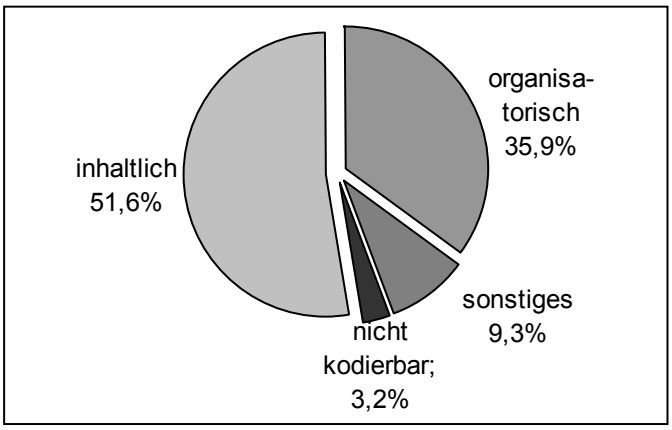


Abb. 7.6: Kategorienverteilung der Facette >Fokus<

Dieses scheinbar ungünstige Verhältnis relativiert sich, wenn man berücksichtigt, dass auch die organisatorischen Schüleräußerungen einer übergeordneten Zielerreichung dienen. Die Summe von organisatorischen und inhaltlichen Äußerungen (87,5%) spiegelt damit ein hohes Maß an Bereitschaft seitens der Schüler wieder, die Aufgaben innerhalb des vorgegebenen zeitlichen Rahmens lösen zu wollen. Der hohe Aufwand an organisatorischen Anteilen ist daher in erster Linie kritisch mit Hinblick auf die Ökonomie des Gruppenarbeitsprozesses zu sehen. Ferner ist in Anbetracht der offenen Lernsituation der mit 9,3% geringe Anteil an aufgabenirrelevanter Nebenkommunikation als positiv zu bewerten.

Der Anteil der nicht kodierbaren turns in Höhe von 3,2% ist auf die schlechte akustische Qualität zahlreicher turns zurückzuführen, die eine sichere Kodierung unmöglich machte.

Die bis dahin deskriptiven Daten gewinnen an Anschaulichkeit, wenn man folgende Vergleichsrechnung aufstellt: Im Rahmen der zweiten TIMSS-Videostudie konnten Reusser und Pauli (2003) an einer großen Stichprobe zeigen, dass in einer durchschnittlichen Unterrichtsstunde die Gesamtzahl der von den Schülern gesprochenen Worte zwischen 690 und 1018 liegt, gemittelt über alle Unterrichtsformen. (Wohingegen auf den Lehrer alleine die fünf- bis achtfache Redebeteiligung entfällt.) Orientiert man sich hierbei an dem gemessenen Höchstwert und legt mit n = 30 eine eher große Schüleranzahl pro Klasse fest, so resultiert hieraus eine durchschnittliche Redebeteiligung von 36 Wörtern pro Schüler und Stunde. Betrachtet man nun allein *die inhaltlichen turns einer Kleingruppenarbeitsphase*, so ergibt sich hierfür ein mittlerer Wert von 190, was bezogen auf eine Vierergruppe einem Wert von 48 turns pro

Schüler entspricht. Zieht man nun noch in Betracht, dass der durchschnittliche turn eine mittlere Länge von (abgerundet) sechs Wörtern aufweist, so kommt man auf eine durchschnittliche Redebeteiligung von 288 Wörtern pro Schüler und Kleingruppensituation, was trotz der konservativen Berechnungsgrundlage exakt dem achtfachen des von Reusser und Pauli ermittelten Wertes einer ganzen Unterrichtsstunde entspricht. Damit ist gewiss noch keine Angabe über die Qualität der Schüleräußerungen gemacht. Die Vergleichsrechnung stellt aber anschaulich das Potenzial der entwickelten und evaluierten Methode dar, den sachbezogenen Anteil an Schülerkommunikation zu erhöhen.

Betrachtet man abschließend die intergruppalen Differenzen der Facette >Fokus<, so ergibt sich folgender Befund: Der Anteil an inhaltlicher Kommunikation zeigt über die vier Gruppen mit einem Bereich von 42,9% (Gruppe 1) bis 59,4% (Gruppe vier) eine große Streuung. Dem Verlauf der Facette >Fokus< von Gruppe eins über alle sechs Stunden in Abbildung 7.7 ist zu entnehmen, dass diese Gruppe auch durch eine große intragruppale Schwankung bei der Besetzung der Kategorien gekennzeichnet ist.

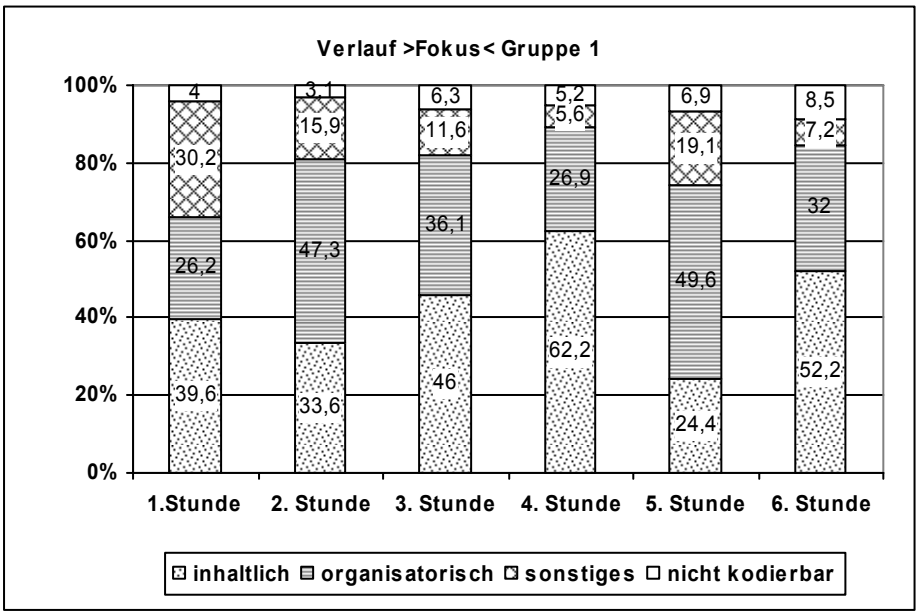


Abb. 7.7: Kategorienverteilung der Facette >Fokus< von Gruppe 1

Dies betrifft insbesondere den zuweilen hohen Anteil an sonstiger Nebenkommunikation. Vergleicht man dahingegen die Verteilung der inhaltlichen turns aller vier Gruppen untereinander (Abb. 7.8), so wird deutlich, dass im Vergleich zur ersten Gruppe alle anderen Kleingruppen eine geringere Schwankung (insbesondere die Gruppen drei und vier) als auch absolut höhere Mittelwerte aufweisen.

So beträgt der prozentuale Mittelwert der inhaltlichen turns bei Gruppe zwei 51,8%, bei Gruppe drei vergleichbar hoch 53,2% und bei Gruppe vier sogar 59,4%. Die Gründe für die große Varianz der ersten Gruppe können mit den im Rahmen dieser Studie eingesetzten Instrumenten nicht erforscht werden und bedürfen darüber hinaus vermutlich einer sozialpsychologischen Expertise.

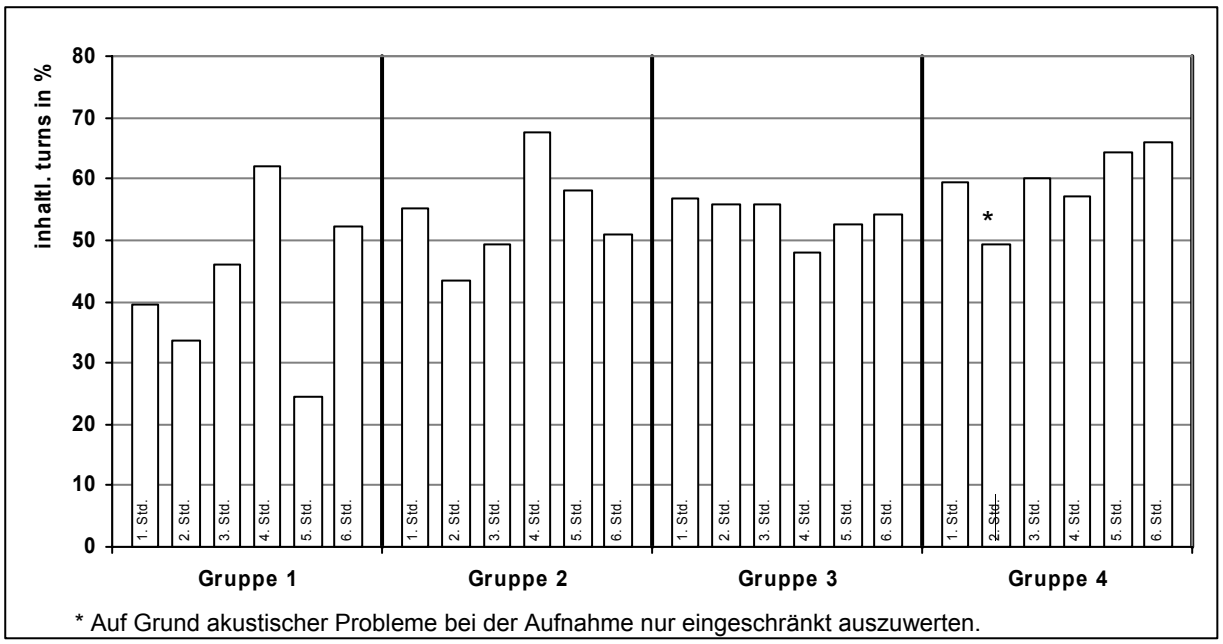


Abb. 7.8: Prozentualer Anteil inhaltlicher turns, differenziert nach Stunde und Gruppe

Die Ergebnisse veranschaulichen aber gut die große Spannbreite der qualitativen Befundlage. Eine der Abbildung 7.7 analoge Darstellung des Verlaufs der Facette >Fokus< von allen vier Kleingruppen ist Anhang F zu entnehmen.

7.2.2 Ergebnisse der Facette >Bezugsrahmen<

Die Ergebnisse der Facette >Bezugsrahmen< zeigen, dass im Mittel rund 90% aller inhaltlichen turns weder einen explizit fachwissenschaftlichen noch einen lebensweltlichen Aspekt aufweisen. Die unter diesem Gesichtspunkt kodierbaren turns verteilen sich auf die beiden Kategorien gemäß Tabelle 7.4.

Gruppe	1. Stunde		2. Stunde		3. Stunde		4. Stunde		5. Stunde		6. Stunde	
	fw	lw	fw	lw	fw	lw	fw	lw	fw	lw	fw	lw
1	0	27	3	2	17	10	16	12	4	20	7	3
2	5	45	0	2	1	10	8	10	5	12	16	11
3	7	36	4	10	1	4	3	21	7	11	9	5
4	1	2	--	--	8	1	7	4	23	6	16	2

Tab. 7.4: Anzahl fachwissenschaftlicher (fw) und lebensweltlicher (lw) turns

In Abhängigkeit von der Kleingruppe resultieren hieraus Spannweiten in den beobachteten Häufigkeiten von

- 2 – 5 Prozent für die Kategorie >fachwissenschaftlich< und
- 4 – 8 Prozent für die Kategorie >lebensweltlich.

Darüber hinaus ist in der Verteilung der Kategorien kein einheitlicher Trend über alle vier Kleingruppen festzustellen. Zwar gibt es partiell einige zeitliche Verläufe, die eine Entwicklung hin zu einer stärker an fachwissenschaftlichen Bezügen orientierten Argumentation nahe legen, wie Abbildung 7.9 verdeutlicht.

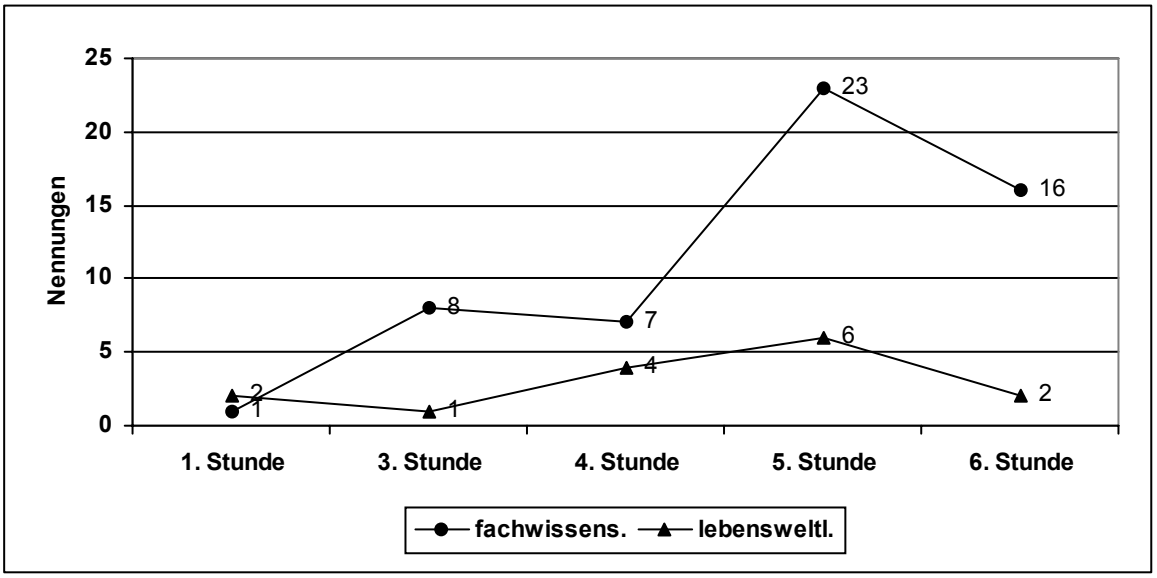


Abb. 7.9: Entwicklungsverlauf fachwissenschaftlicher und lebensweltlicher turns von Gruppe 4

Betrachtet man hingegen die Verteilung der Kategorien über die Zeit am Beispiel der Gruppe drei, so lässt sich eine derartige Entwicklung nicht feststellen. In diesem Fall kann lediglich von einem Rückgang der lebensweltlich geprägten turns ausgegangen werden.

Eine Gesamttendenz ergibt sich allerdings, wenn man die gruppenbezogenen Daten der ersten und sechsten Stunde über die beiden Kategorien aggregiert und gegenüberstellt, wie Abbildung 7.10 zu entnehmen ist. In diesem Fall ist der angestrebte Effekt zu verzeichnen, wonach über die Dauer der Intervention der Anteil lebensweltlich geprägter Konzepte zu Gunsten erklärungsmächtiger und fachwissenschaftlich angemessener Vorstellungen zurückgehen sollte.

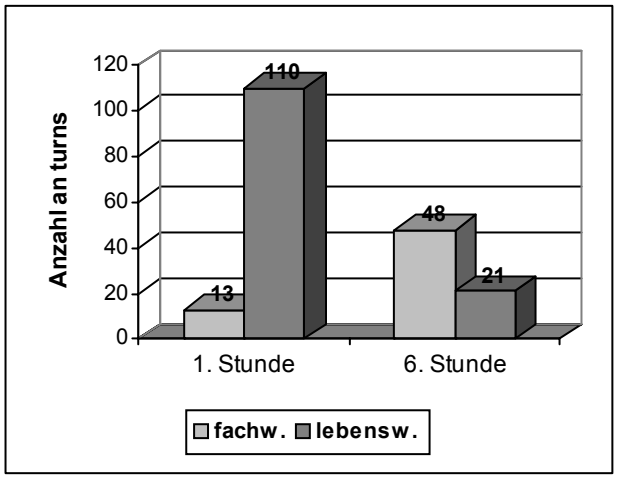


Abb. 7.10: Vergleich der Facette >Bezugsrahmen zwischen der ersten und der sechsten Unterrichtsstunde

Da dieser Befund jedoch erst mit der sechsten Unterrichtsstunde festzustellen ist und bis zu diesem Zeitpunkt die lebensweltlichen Konzepte im Stundenvergleich dominieren (bzw. etwa gleichhäufig mit den fachwissenschaftlichen auftreten; s. Tab. 7.4, 3. Stunde), ist diese Vermutung spekulativer Natur. Weil auf Grund der sechsstündigen Konzeption der Reihe keine weiteren Folgedaten vorliegen, kann somit nicht entschieden werden, inwieweit sich bei diesem Befund ein stabiler Konzeptwechsel abzuzeichnen scheint oder ob Alternativerklärungen in Erwägung gezogen werden müssen. So ist es beispielsweise auch denkbar, dass sich in dem Befund ein dirigierender Einfluss der Aufgabenstellung selbst widerspiegelt, welche im Falle der sechsten Stunde bedingt durch ihre Komplexität eine Vernetzung und Integ-

ration aller im Rahmen der Unterrichtsreihe erworbenen Fachbegriffe notwendig macht.

7.2.3 Ergebnisse der Facette >Theoriegehalt<

Die Auszählung der Kodierungen für die Facette >Theoriegehalt< zeigt bereits bei einer oberflächlichen Betrachtung eine nur äußerst geringe Besetzung in den relevanten Kategorien >theoretisch explizit< und >theoretisch implizit<. Über 97% aller inhaltlichen turns sind in keine der beiden Kategorien einzuordnen und gelten daher als >atheoretisch<. Damit liegt bei dieser Facette, vergleichbar mit der Facette >Bezugsrahmen<, ein ungünstiges Verhältnis zwischen den tatsächlich zu beobachtenden Ereignissen und der Gesamtzahl aller turns vor.

Gruppe	1. Stunde		2. Stunde		3. Stunde		4. Stunde		5. Stunde		6. Stunde	
	t_{ex}	t_{im}	t_{ex}	t_{im}	t_{ex}	t_{im}	t_{ex}	t_{im}	t_{ex}	t_{im}	t_{ex}	t_{im}
1	1	1	1	0	1	1	3	0	1	0	1	2
2	1	1	0	1	2	5	5	8	4	4	7	5
3	5	2	3	2	4	0	7	3	2	0	2	2
4	4	2	xx	xx	2	0	1	1	0	1	2	1

Tab. 7.5: Anzahl theoretisch expliziter (t_{ex}) und theoretisch impliziter (t_{im}) turns

Die übergroße Besetzung der Kategorie >atheoretisch< erklärt auch den extrem hohen Übereinstimmungskoeffizienten, der bei dieser Facette in einzelnen Stunden sogar den Wert 1,0 einnehmen kann. Damit verbunden stellt sich angesichts des nicht zu leugnenden Bodeneffekts die prinzipielle Frage nach der Aussagekraft dieser Unterfacette. Andererseits kann auf Grund fehlender Referenzstudien auf dem Gebiet der prozessbezogenen Unterrichtsforschung nicht gesagt werden, wie groß der Anteil an theoretisch fundierten, logisch begründeten Sachaussagen für eine erfolgreich arbeitende Kleingruppe sein muss. Sicherlich kann davon ausgegangen werden, dass für den einzelnen Gruppenprozess die Qualität einiger weniger turns bedeutsamer sein mag als die Quantität aller theoriebasierten Äußerungen. Insofern bieten die Werte einen ersten Anhaltspunkt für die Beschreibung des Ist-Zustandes des Ausmaßes an theoriebasierter schülerinterner Kommunikation.
Darüber hinaus soll nochmals darauf verwiesen werden, dass die kleine Fallzahl der videographierten Kleingruppen eine Verallgemeinerung der Ergebnisse in keinerlei Richtung zulässt. So zeigen beispielsweise die Kodierergebnisse der zweiten Gruppe für sich alleine betrachtet ein durchaus differenziertes und interpretierbares Bild (s. Abb. 7.11).

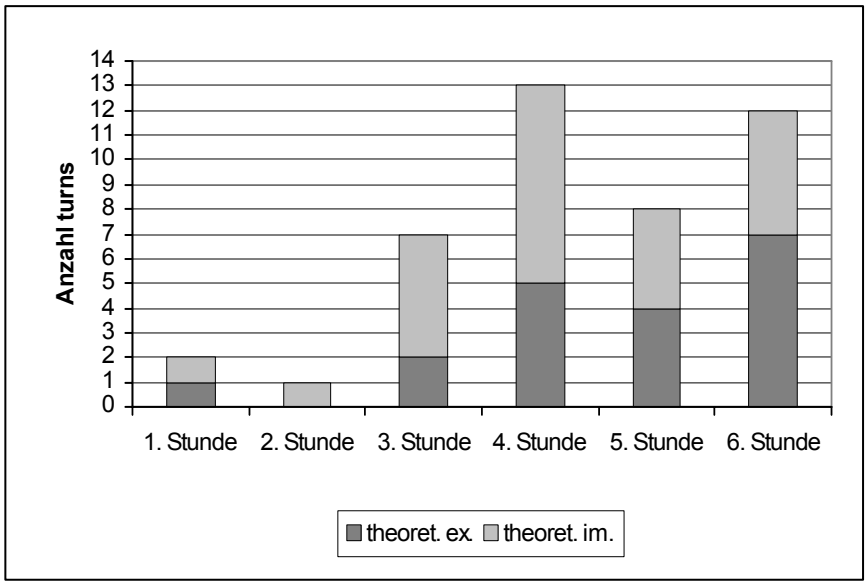


Abb. 7.11: Entwicklungsverlauf theoretischer turns von Gruppe 2

Der Entwicklungsverlauf dieser Kleingruppe zeigt eine deutliche Zunahme theoriebasierter Äußerungen über die Zeit. Bemerkenswert sind ferner die relativ hohen Werte der vierten und sechsten Stunde, was damit in Zusammenhang stehen könnte, dass in beiden Stunden die Aufgabenstellungen besonders komplex sind, so dass die Notwendigkeit einer begründeten Argumentation größer sein könnte, als dies bei Aufgabenstellungen der Fall ist, aus denen sich schnell und für alle Gruppenmitglieder einvernehmlich ein Handlungsplan ableiten lässt.

7.2.4 Ergebnisse der Facette >Problemelemente<

Die Kodierung der Unterfacette >Problemelemente< zeigt im Gegensatz zu den Unterfacetten >Bezugsrahmen< und >Theoriegehalt< eine ausgeprägte Besetzung der relevanten Kategorien (differenziert, integriert, redundant, s. Tab. 7.6). Dementsprechend hoch ist der Prozentsatz der turns, die ein Problemelement aufweisen; er nimmt in Abhängigkeit von der Gruppe und der Unterrichtsstunde einen Wert zwischen 52 und 64 Prozent an, wodurch auch die quantitative Bedeutsamkeit der Facette dokumentiert wird.

Gruppe	1. Stunde			2. Stunde			3. Stunde			4. Stunde			5. Stunde			6. Stunde		
	D	I	R	D	I	R	D	I	R	D	I	R	D	I	R	D	I	R
1	15	30	27	21	14	51	14	30	49	21	28	57	6	9	12	17	26	74
2	17	27	26	17	39	62	11	61	66	11	101	94	6	34	23	17	96	105
3	15	61	68	21	62	68	9	60	52	11	92	54	5	17	22	19	81	91
4	14	31	45	--	--	--	22	25	46	11	35	81	7	21	32	16	44	86

Tab. 7.6: Anzahl differenzierender (D), integrierender (I) und redundanter (R) turns

Auffällig sind die Kodierungsergebnisse der fünften Stunde, auf die – mit einer Ausnahme – die niedrigsten Summenwerte für alle Gruppen in allen drei Kategorien entfällt. Dieser Befund ist mit hoher Wahrscheinlichkeit auf die Aufgabenstellung der Stunde zurückzuführen, da es sich in diesem Fall um eine reine experimentfreie Textarbeit handelt. Aus diesem Grund ist die Zahl der potenziellen Problemelemente

in dieser Stunde geringer als das vergleichsweise in der vierten oder sechsten Stunde der Fall ist.

Betrachtet man den Verlauf der Kategorien innerhalb der Gruppen über die Zeit, so scheint sich keine nennenswerte Entwicklung innerhalb der sechs Unterrichtsstunden zu vollziehen. Lediglich die Kategorie >integriert< der Gruppe zwei zeigt hier eine deutliche Zunahme über die Zeit.

Betrachtet man die Gruppenmittelwerte der drei Kategorien über die sechs Stunden (Tab. 7.7), so liegen die Werte der Kategorie >differenziert< mit einem Bereich von 13,2 bis 15,6 turns pro Stunde innerhalb eines sehr engen Bereichs. Dies bedeutet, dass alle vier Gruppen in vergleichbarer Intensität die Problemstellungen differenzieren, woraus sich schließen lässt, dass der Differenzierungsgrad über die Aufgabenstellung vermittelt wird. Dahingegen scheinen sich in der Kategorie >integrierend< zwei Trendrichtungen abzuzeichnen. Die Gruppen zwei und drei weisen einen hohen Integrationsgrad auf, wohingegen von den Gruppen eins und vier hier nur halb so hohe Werte erreicht werden. Es scheint somit – unabhängig von der Aufgabenstellung – den vier Gruppen in unterschiedlichem Maße möglich, die differenziert eingebrachten Problemelemente auch inhaltlich zu integrieren und zu vernetzen. Damit bietet diese Facette für weitere Studien, die auf eine größere Fallzahl der Kleingruppen rekurrieren, die Möglichkeit, Unterschiede in den Diskussionsverläufen zwischen den Gruppen zu analysieren und mit der Lernleistung zu korrelieren.

Gruppe	M_D	M_I	M_R
1	15,6	22,8	45,0
2	13,2	59,6	62,6
3	13,3	62,2	59,2
4	14,0	31,2	58,0

Tab. 7.7: Mittelwerte der Facette >Problemelemente<

Ein weiterer Gesichtspunkt der Analyse dieser Unterfacette, der stärker den Prozesscharakter der Videodaten hervorhebt, fokussiert auf die Verteilung der Kategorien über den zeitlichen Verlauf der Kleingruppensituation. In Abbildung 7.12 ist dieser Verlauf exemplarisch für eine Gruppenarbeitsphase dargestellt.

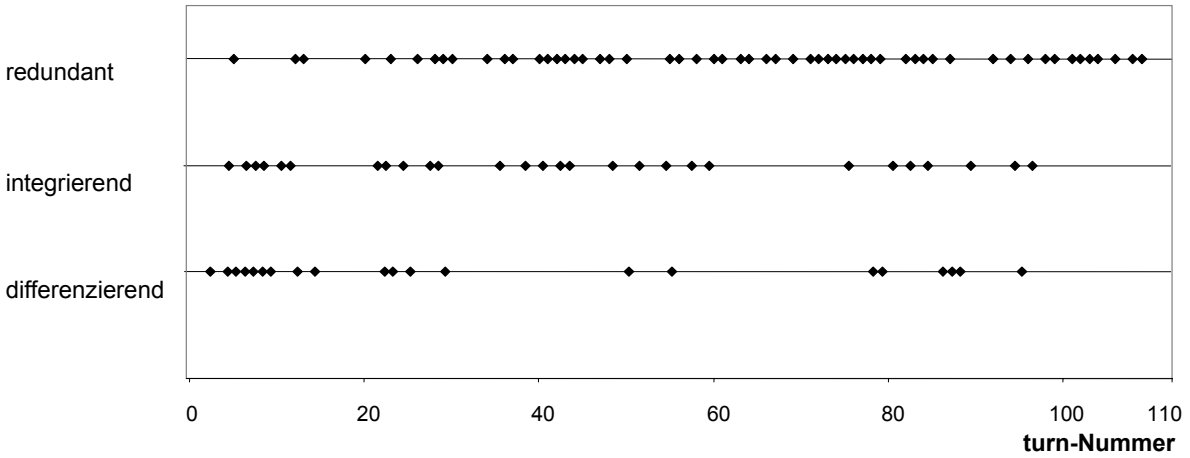


Abb. 7.12: Kategorienverlauf der Facette >Problemelemente< der Gruppe 1, vierte Stunde

In diesem Diagramm ist jeder turn der Unterfacette in Bezug auf die kodierte Kategorie (Y-Achse) in Abhängigkeit von der chronologischen Stellung in der Arbeitsphase (X-Achse) aufgetragen. Dadurch lässt sich ein halbquantitativer Verlauf der Kategorienzuschreibung über die Kleingruppensituation erstellen. Man sieht, dass zu Beginn der Arbeitsphase zunächst in sehr dichter Abfolge neun Problemelemente mit differenzierender Funktion in die Diskussion eingebracht und anschließend miteinander vernetzt werden. Im anschließenden Verlauf der Arbeitsphase werden weitere differenzierende Problemelemente nur noch punktuell eingebracht, wohingegen integrierende Ereignisse in insgesamt geringer Anzahl, aber mit kontinuierlicher Frequenz beobachtet werden. Mit einer gewissen zeitlichen Verzögerung ist etwa ab turn 40 eine deutliche Zunahme redundanter turns zu verzeichnen, deren Frequenz aber im letzten Drittel der Arbeitsphase wieder abnimmt.

Die beschriebene Abfolge der Kategorienhäufung über die Zeit erscheint einerseits nahe liegend für jede Form der diskursiven Problemlösung und von daher wenig spektakulär, andererseits gibt es bislang nur wenig Erkenntnisse darüber, ab welcher Alterstufe Schüler überhaupt in der Lage sind, komplexe Problemstellungen, die ein Kommunizieren naturwissenschaftlicher Argumente erfordern, in Schülergruppen zu bearbeiten. Trivial ist in diesem Zusammenhang allenfalls die differenzierende Nennung von Problemelementen zu Beginn der Gruppenarbeitsphase. Es kann aber gerade als Beleg *für* ein Kommunizieren naturwissenschaftlicher Argumente gesehen werden, dass die Schüler die Kompetenz besitzen, die zunächst isoliert (differenzierend) vorgebrachten Argumente im weiteren Verlauf der Gruppenarbeit aufzunehmen und zu vernetzen. Damit würde diese Unterfacette einen innovativen Beitrag zur prozessbezogenen Operationalisierung kommunikativer Kompetenzen liefern, der über ein sozialpsychologisches Verständnis von Kommunikationsfähigkeit hinausgeht.

Die Inhaltliche Progressionsanalyse dient, wie dargelegt, dem Zweck, den inhaltlichen Lernfortschritt einer Kleingruppe unter fachdidaktischen Aspekten zu explizieren. Diese Zielsetzung ist abschließend wie folgt zu bewerten: Die Kodierung des Argumentationsfokus erweist sich als eine nützlich und einfach zu realisierende, niederinferente Beschreibung der Gesprächsorientierung und der Gesprächsökonomie der Gruppenarbeitsphase. Die weitere Kodierung der inhaltlichen Aussagen mit dem Ziel der Identifizierung von lebensweltlichen und fachwissenschaftlichen Schüleraussagen erscheint auf Grund der geringen Anzahl an eindeutig zu kodierenden Aussagen als eine wenig geeignete Kategorie für eine auf größere Fallzahlen angelegte Untersuchung. Eventuell ist die Erklärungsmächtigkeit dieser Facette mehr im Kontext qualitativer Einzelfallstudien zu sehen. Dahingegen scheint die Facette >Problemelemente< einen hoffnungsvollen Ansatzpunkt zur Beschreibung einer bislang nur schwierig zur operationalisierenden Schlüsselkompetenz (Kommunikationsfähigkeit) darzustellen. Die Ergebnisse der Unterfacette >Theoriegehalt< zeigen über alle Gruppen betrachtet ein mit der Facette >Bezugsrahmen< vergleichbares und damit wenig aussagekräftiges Ergebnis. Es steht jedoch zu vermuten, dass es sich hierbei primär um ein Problem der Formulierung angemessener Indikatoren handelt. Möglicherweise sind die Begründungsformen, aus denen eine Kleingruppe ihren Hand-

lungsplan ableitet, sprachlich weiter zu fassen, als dies im Rahmen dieser Facette vorgenommen wurde. Andernfalls stellt sich die Frage, inwieweit tatsächlich davon ausgegangen werden muss, dass die Gruppen ihre Experimente durchführen ohne dies auf im Vorfeld geäußerte Überlegungen gründen zu können. Um diese Frage beantworten zu können wurde ein weiteres qualitatives Auswerteverfahren entwickelt, die so genannten Prozessgrafiken.

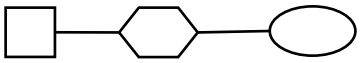
7.3 Prozessgrafiken zur qualitativen Analyse der Gruppenarbeitsphase

Prozessgrafiken dienen der Rekonstruktion des Gruppenarbeitsprozesses zum Nachvollzug der entwickelten und umgesetzten Lösungsansätze (Sumfleth, Rumann, Nicolai 2004). In Anlehnung an die im naturwissenschaftlichen Unterricht vorherrschende Form der Erkenntnisgewinnung erfolgt eine Phasierung der Lösungsansätze in die Schritte >Hypothesenbildung<, >Experimententwicklung und -durchführung< sowie >Auswertung des Experiments<. Als Datenquelle wird hierzu neben den Unterrichtsvideos vor allem das Transkript der Gruppenarbeitsphase herangezogen. Methodisch handelt es sich bei dieser Form der Datenanalyse um eine hermeneutische Inhaltsanalyse, deren Ergebnisse zur besseren Veranschaulichung grafisch dargestellt werden. Mayring (2000) verweist auf den *Verstehens*prozess bei der Anfertigung einer Inhaltsanalyse und betont, dass die Analyse nicht auf der Ebene der Oberflächeninhalte verhaftet bleiben darf (wie dies im Rahmen der Kodierung der Inhaltlichen Progressionsanalyse geschieht), sondern auf *latente Sinngehalte* abzielen muss.

Latente Sinngehalte erschließen sich in diesem Zusammenhang vor allem durch die Identifikation der von den Schülern vorgebrachten Ideen sowie durch die reflexive Interpretation, ob und wie die Schüler die experimentellen Ergebnisse mit Bezug auf ihre ursprüngliche Erwartungshaltung zu deuten wissen. Darüber hinaus werden die drei Elemente eines Lösungsansatzes hinsichtlich ihrer fachlichen Richtigkeit bewertet. Daraus resultiert für die Prozessgrafik folgende Symbolgebung:

- □/■ Quadrate symbolisieren eine Idee für eine Problemlösung, wobei die Idee fachwissenschaftlich richtig (weiß) oder falsch (grau) sein kann.
- ⬡/⬢ Sechsecke symbolisieren ein Experiment, das technisch richtig (weiß) oder falsch (grau) umgesetzt werden kann.
- ⬭/⬬ Ellipsen symbolisieren die Auswertung eines Experiments, wobei ein Rückbezug zur entsprechenden Idee des Lösungsansatzes herzustellen ist: Wird aus einem Experiment die fachlich richtige Schlussfolgerung gezogen, so wird dies mit einer weißen Ellipsen vermerkt, andernfalls wird die Farbe grau vergeben.

Im Idealfall resultieren hieraus folgende Lösungsabläufe:

A)  Eine fachwissenschaftlich richtige Idee wird durch ein richtig durchgeführtes Experiment verifiziert.

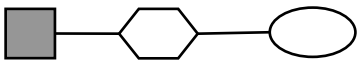
B)  Eine fachwissenschaftlich falsche Idee wird durch ein richtig durchgeführtes Experiment falsifiziert.

Abbildung 7.13 gibt ein typisches Beispiel für eine Prozessgrafik einer Kleingruppensituation. Auf der Y-Achse werden chronologisch die Lösungsansätze notiert, auf der Abszisse wird die Zeit aufgetragen.

Die Prozessgrafik ist wie folgt zu lesen: Schüler Nr. 16 formuliert eine Idee, die nicht beachtet wird. Anschließend formuliert derselbe Schüler eine weitere Idee, die dieses Mal in einer idealen Sequenz verifiziert wird.
Analysiert man auf diese Art die Prozessgrafik, so lassen sich folgende zentralen Fehlerquellen identifizieren (vgl. Sumfleth, Rumann, Nicolai 2004):

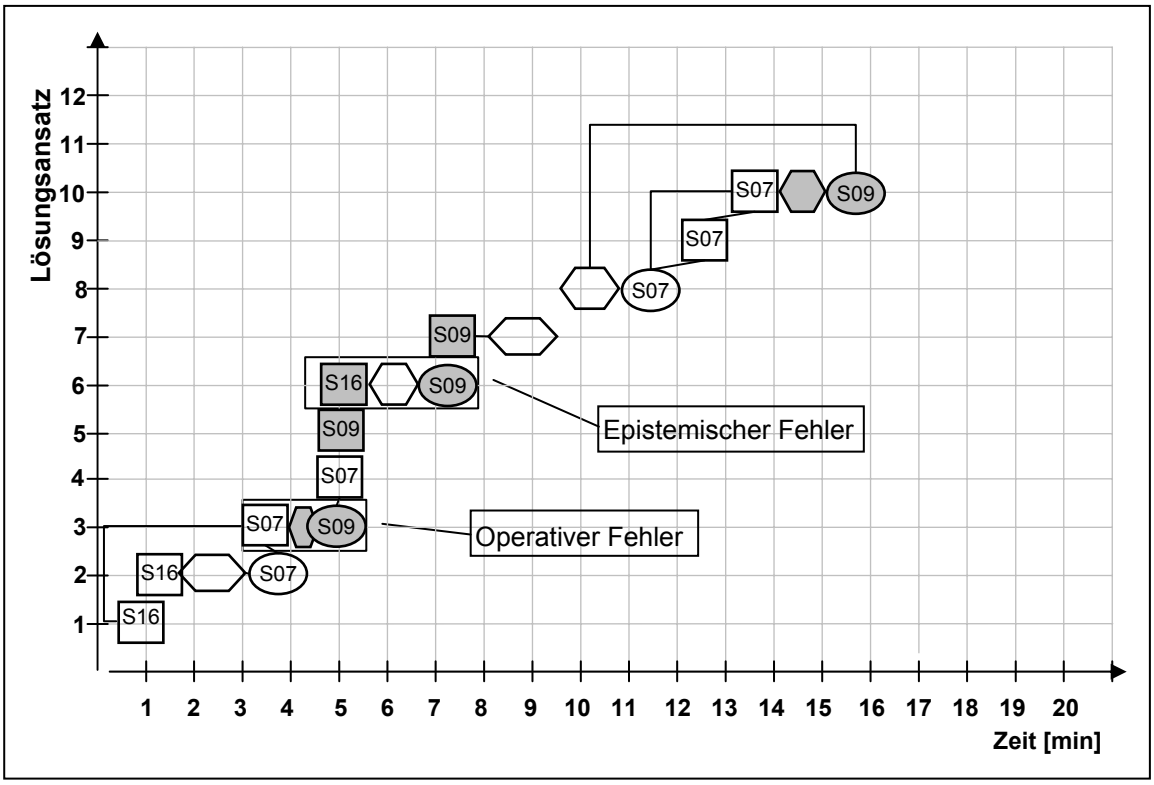


Abb. 7.13: Qualitative Prozessanalyse einer Kleingruppensituation (□ = Idee, ⬡ = Experiment, ⬭ = Auswertung, S16 = Schülercode, weiß = fachlich richtig, grau = fachlich falsch)

- In vielen Fällen bleiben die Lösungsansätze unvollständig; d.h. fachlich richtige, zielführende Ideen werden von der Gruppe nicht aufgegriffen (Ansatz 1, 4 und 9). Lösungsansatz Nummer sieben veranschaulicht eine Situation, in der eine fachlich falsche Idee experimentell umgesetzt und somit falsifiziert werden könnte. Allerdings bleibt die Schlussfolgerung bei diesem Lösungsansatz komplett aus.
- Die Lösungsansätze Nummer drei und zehn zeigen zwar einen dreischrittigen Verlauf, widersprechen aber den oben formulierten Idealfällen, da beide Ansätze eine fachlich richtige Idee durch einen Fehler bei der experimentellen Umsetzung fälschlicherweise falsifizieren. Dieser Fehler wird daher als operativer Fehler bezeichnet.
- Besonders folgenschwer erscheint ein weiterer Fehlertypus, der als epistemischer Fehler zu bezeichnen und in Lösungsansatz Nummer sechs zu erkennen ist. Ausgehend von einer fachlich falschen Idee wird ein zielführendes Experiment konzi-

piert und richtig durchgeführt. Aus dem Experiment wird aber die falsche Schlussfolgerung gezogen, so dass die eigentlich fachlich falsche Idee irrtümlicherweise „verifiziert“ wird. Offensichtlich scheint in diesem Fall die Schülervorstellung derart stabil zu sein, dass die Evidenz des Experiments widerspruchsfrei ignoriert werden kann und die Beobachtung der gewünschten Interpretation angepasst wird.

Vor dem Hintergrund dieser Fehlerquellen vergleicht Dettloff (2003) die Kleingruppensituation mit der entsprechenden Problemlösephase einer Frontalunterrichtssituation der Interventionskontrollgruppe und kommt dabei zu folgendem Ergebnis (s. Abb. 7.14).

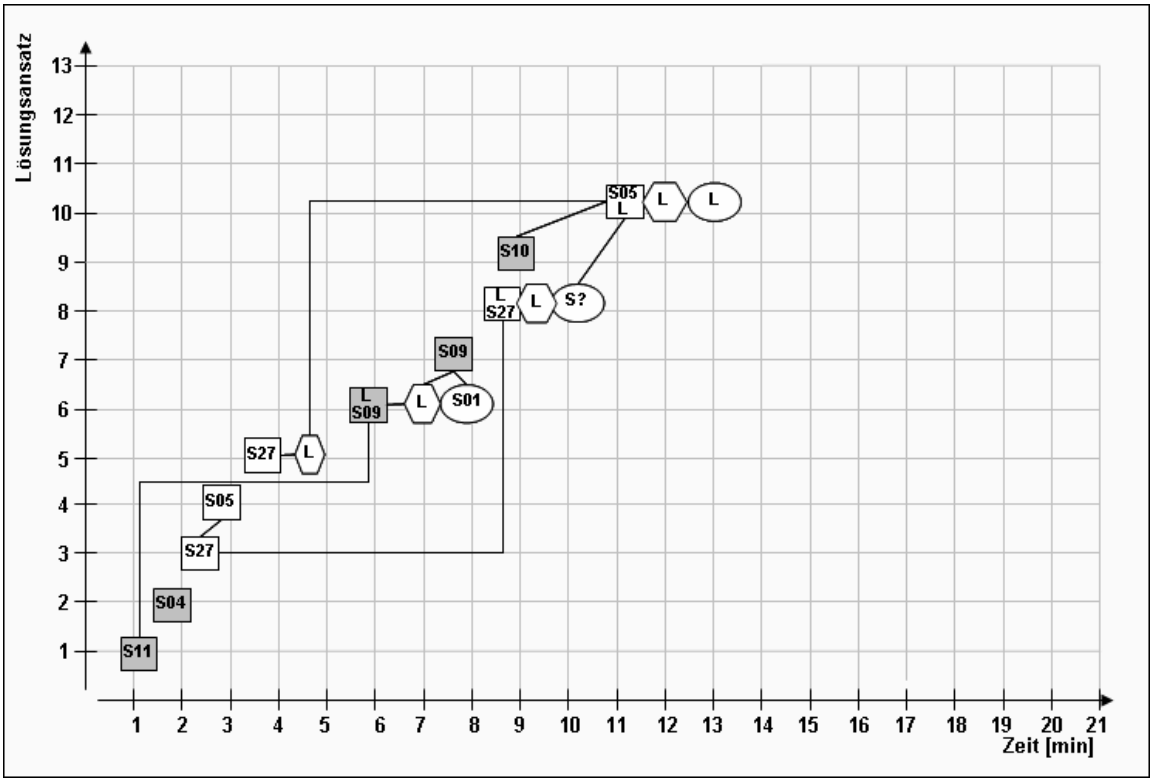


Abb. 7.14: Qualitative Prozessanalyse einer Frontalunterrichtssituation

Bereits bei einer oberflächlichen Betrachtung erscheint der Gesamteindruck der Grafik sehr viel übersichtlicher und deutlich strukturierter. Die eckigen schwarzen Linien deuten auf eine hohe Bezugsbildung zwischen den einzelnen Lösungsansätzen hin. Die ersten vier Lösungsansätze zeugen von einer Unterrichtsphase, in welcher der Lehrer die Ideen zunächst sammelt, ohne sie auf ihre fachliche Richtigkeit zu bewerten.

Im fünften Lösungsansatz greift der Lehrer eine zielführende Idee auf, deren Deutung aber unterbleibt, vermutlich weil sonst das Problem bereits gelöst wäre. Stattdessen thematisiert der Lehrer in der Folge eine fachlich falsche Idee, die er anschließend in einem Demonstrationsversuch falsifiziert. Im Folgenden greift der Lehrer nochmals die von Schüler siebenundzwanzig eingebrachte Idee auf, verifiziert diese und nimmt das Ergebnis als Grundlage für den abschließenden Lösungsansatz Nummer zehn.

Der Gesamtablauf der Problemphase erscheint kompakt und zielorientiert. Die Prozessgrafik der lehrerzentrierten Stunde zeigt aber auch, dass die gesamte Problemlösephase neben dem Lehrer von nur sieben weiteren Schülern bestritten wird.

Somit bleibt abschließend festzuhalten, dass die Schüler der Kleingruppenphase sehr wohl selbstständig in der Lage sind, ihre Experimente – wenn nicht aus Theorien – so doch zumindest aus Ideen planmäßig abzuleiten. Neben handwerklich-technischen Problemen bei der Umsetzung des Experiments zeigen die Schüler Schwierigkeiten bei der effektiven Regulation ihres Lernprozesses und insbesondere dem Abgleichen ihrer Vorannahmen mit den aus den Experimenten resultierenden Evidenzen.

8. Zusammenfassung und Ausblick

Das erwartungswidrig schlechte Abschneiden deutscher Schüler bei den großen internationalen Schulleistungsstudien der vergangenen Jahre lässt keinen Zweifel darüber, dass sich die Praxis des mathematisch-naturwissenschaftlichen Unterrichts grundlegend ändern muss. Auf der Ebene der Problemanalyse besteht hohes Einvernehmen darüber, dass der deutsche Unterricht im internationalen Vergleich durch eine Methodenarmut gekennzeichnet ist, die sich in einer Monokultur des fragend-entwickelnden Unterrichtsgesprächs ausdrückt. Einen hohen Redeanteil sowie die Formulierung von engen und konvergenten Fragestellungen beschreiben das Gesprächsverhalten des Lehrers in dieser direktiven Unterrichtsform, in der den Schülern ein überwiegend passiv-rezeptives Verhalten zugeschrieben werden kann. Es ist daher nicht überraschend, dass in Folge dieser unsymmetrischen Kommunikationsverteilung die Schüler gerade bei solchen Aufgaben große Schwierigkeiten zeigen, die ein flexibel anwendbares und transferfähiges Wissen erfordern. Weinert (1998) verdichtet die dominanten Defizite des mathematisch-naturwissenschaftlichen Unterrichts auf fünf Eigenschaften und stellt ihnen jeweils prospektiv einen stärker zu fördernden Aspekt gegenüber. Demnach ist der deutsche Unterricht:

(1) zu inhaltsbezogen ⟷ zu wenig verständnisintensiv

(2) zu leistungsbezogen ⟷ zu wenig lernorientiert

(3) zu themenbezogen ⟷ zu wenig begabungsdifferenzierend

(4) zu wissensbezogen ⟷ zu wenig nutzungsorientiert

(5) zu informierend ⟷ zu wenig qualifizierend

Andererseits besteht inzwischen Klarheit darüber, dass Lernprozesse suboptimal verlaufen, so sie ausschließlich auf selbstorganisierte Lernformen setzen. Im Rahmen dieser Studie wurde daher ein methodisches Konzept entwickelt, welches gleichermaßen konstruktivistisch-offene wie auch instruktional-gebundene Elemente aufweist.

Das Konzept sieht die Bearbeitung explorativer Aufgabenstellungen in kooperativer Kleingruppenarbeit vor. Im Gegensatz zur tradierten Form des Schülerexperiments, welche häufig durch eine reine operationale Zielsetzung gekennzeichnet ist, erhalten die Lernenden keine dezidierten Versuchsvorschriften, die ihnen die Versuchsdurchführung und die Beobachtungsaufgabe diktieren. Stattdessen werden die Schüler gehalten, ausgehend von einer gegebenen Problemstellung, selbstständig einen Versuchsplan zu erstellen, ein angemessenes Experiment zu entwickeln und durchzuführen sowie die entsprechenden Schlussfolgerungen in Abhängigkeit von ihrer Ausgangsvermutung zu ziehen.

Die Studie wurde in der siebten Jahrgangsstufe an zwei Gymnasien mit insgesamt acht Klassen im klassischen Interventions-Kontrollgruppen-Design durchgeführt. Während die Interventionsgruppe die Problemstellungen in der dargestellten Art und Weise in Kleingruppen bearbeiteten, wurden in der Kontrollgruppe die gleichen Inhaltsbereiche mit den gleichen Experimenten in einem lehrerzentrierten, fragend-

entwickelnden Unterrichtsgespräch realisiert. Dabei wurde großer Wert auf die Konstanthaltung oder, so dies der reguläre Unterrichtsablauf nicht zuließ, zumindest auf die Kontrolle möglichst vieler, die Lernleistung beeinflussender Störvariablen gelegt. Als besonders hervorzuhebende Gütekriterien können hierbei die Konstanthaltung der Lehrperson zwischen Interventions- und Kontrollgruppe einer Schule sowie die Konstanthaltung der effektiven Unterrichtszeit (time on task) beider Unterrichtstreatments erachtet werden.

Eine quantitative Leistungsevaluation erfolgte an Hand von Fachleistungstests, die vor, unmittelbar nach und sechs Monate nach Abschluss der Intervention durchgeführt wurden. Zusätzlich wurden längsschnittlich einzelne Schülergruppen während der Kleingruppenarbeit videographiert, um so verhaltensnahe Prozessdaten zu gewinnen.

Die aus den Varianzanalysen resultierenden Ergebnisse zeigen über alle Schüler gemittelt in beiden Testformen einen signifikanten Vorteil für die unterrichtlich intervenierten Schüler, der bei Berücksichtigung der Langzeitergebnisse des Follow-up-Tests am größten ausfällt. Die partielle Effektgröße belegt in allen Fällen einen schwachen bis mittleren Effekt, was umgerechnet auf die schulüblichen Prinzipien der Leistungsmessung einem Unterschied von einer Notentendenz entspricht. Bei einer differenzierten Betrachtung der Probanden stellt sich heraus, dass diese in unterschiedlichem Maße von der Intervention profitieren. Durch die Intervention werden insbesondere leistungsschwächere und – in etwas geringerer Signifikanzschranke – leistungsstärkere Schüler gefördert, während die Schüler des mittleren Leistungsspektrums keinen signifikanten Leistungsunterschied zur Kontrollgruppe zeigen. Die Unterschiede treten weiterhin besonders dann hervor, wenn man allein die männlichen Probanden betrachtet, wohingegen bei den weiblichen Probanden kein überzufälliger Leistungsunterschied zur Kontrollgruppe besteht.

Die Berücksichtigung der Person des Lehrers als Kontrollvariable erbrachte ferner einen bedeutsamen Einfluss seiner Person auf den Erfolg der Intervention, wohingegen bei Analyse aller Klassen der Kontrollgruppe untereinander kein Lehrereffekt zu verzeichnen war. Daraus lässt sich eine unterschiedliche Interventionssensitivität der Lehrperson ableiten, die unabhängig ist von dem Erfolg ihres traditionellen Lehrerverhaltens.

Die Ergebnisse der Videoanalysen belegen eine zufrieden stellende Aufgabenorientierung der videographierten Kleingruppen. Es konnte gezeigt werden, dass die Redebeteiligung der videographierten Schüler um das Achtfache über den Referenzwerten liegt, wie sie aus frontalunterrichtlichen Situationen bekannt ist. Die Kodierung der Schüleräußerungen zur Identifizierung von lebensweltlichen Vorstellungen erwies sich allerdings als nicht geeignet zur weiteren Differenzierung des Gruppenarbeitsprozesses. Dahingegen bietet die Entwicklung der Prozessgrafiken ein Instrumentarium, welches eine umfassende qualitative Beschreibung der Gruppenarbeitsprozesse erlaubt. Für die Zukunft wäre es wünschenswert, wenn es gelänge, einzelne Aspekte der Prozessgrafiken zu quantifizieren, um so die Grafiken untereinander besser vergleichen zu können.

Die Prozessgrafiken belegen ferner, dass die Schüler der Kleingruppenphase zu weilen große Schwierigkeiten mit dem technisch korrekten Durchführen der Experimente haben. Ebenso lassen sich Situationen nachweisen, in denen Schüler die Ergebnisse der Experimente ignorieren oder diese massiv uminterpretieren, sofern die Ergebnisse nicht ihren Erwartungen entsprechen. Dies kann als Beleg für die in dieser Alterststufe noch unzureichenden Selbstregulationskompetenzen gewertet werden. Für zukünftige Arbeiten sollte diese Fehlerquelle durch weitere instruktionale Unterstützungen gemindert werden. Schließlich bleibt zu wünschen, künftig auch Kleingruppenstudien mit größeren Fallzahlen an videographierten Gruppen durchführen zu können. Dies würde es ermöglichen, die beobachteten Prozessdaten mit den Testleistungsdaten zu triangulieren, um so einen Einblick in den Zusammenhang zwischen dem Lernprozess und der Lernleistung zu erhalten.

9. Literaturverzeichnis

Arnold, K., Franik, R. (2000): Chemie plus. Gymnasium Klasse 7 NRW. Berlin: Volk und Wissen.

Aronson, E., Blaney, N., Stephan, C., Sikes, J., Snapp, M. (1978): The Jigsaw Classroom. BeverlyHills, CA: Pergamon Press.

Asselborn, W., Jäckel, M., Risch, K. (2001): Chemie heute SI. Ausgabe NRW. Hannover: Schroedel Schulbuch Verlag.

Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2003): Multivariate Analysemethoden. Berlin u.a.O.: Springer.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., Weiß, M. (Hrsg.) (2002): PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. 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., Köller, O. (2000): Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos, R. Lehmann (Hrsg.): TIMSS/III – Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe, S. 271 – 315. Opladen: Leske & Budrich.

Baumert, J., Bos, W., Watermann, R. (1998): TIMSS/III: Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich. Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, J. (1998): Internationale Schulleistungsvergleiche. In J. Rost (Hrsg.): Handwörterbuch Pädagogische Psychologie, S. 219 – 225. Weinheim: Beltz.

Baumert, J., Lehmann, R., Lehrke, M. et al. (1997): TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: deskriptive Befunde. Opladen: Leske + Budrich.

Bierhoff, H. W., Rudinger, G. (1996): Quasi-experimentelle Untersuchungsmethoden. In E. Erdfelder et al. (Hrsg.): Handbuch quantitative Methoden, S. 47 – 58. Weinheim: Psychologie Verlags Union.

Bloom, B., Engelhardt, M., Furst, E., Hill, W., Krathwohl, D. (Eds.) (1956): Taxonomy of educationalobjectives: Handbook 1. Cognitive domain. New York: David McKay.

Bloom, B. S. (1971): Individual differences in school achievement: A vanishing point – Deutsche Übersetzung von Edelstein. In W. Edelstein, D. Hopf (Hrsg.): Bedingungen des Bildungsprozesses, S. 251 – 270. Stuttgart: Klett.

Bloom, B. S. (1976): Human characteristics and school learning. New York: Ballantine Books.

Blume, R., Kunze, W., Meloefski, R., Obst, H., Rossa, E., Schönemann, H. (1994): Chemie für Gymnasien, Sekundarstufe I - Lehrebuch zum Teilband 1 (Klasse 7). Berlin: Cornelsen Verlag.

Boos, M., Morguet, M., Meier, F., Fisch, R. (1990): Zeitreihenanalyse bei der Bearbeitung komplexer Probleme in Expertengruppen. Zeitschrift für Sozialpsychologie, S. 53-64.

Boos, M. (1996): Entscheidungsfindung in Gruppen. Bern, Göttingen, Toronto, Seattle: Verlag Hans Huber.

Borrows, P. (1989): Great chemical egg races! Education in chemistry, 26, S. 88 – 90.

Bortz, J. (1999): Statistik für Sozialwissenschaftler. Berlin u.a.O.: Springer.

Bortz, J., Döring, N. (1995): Forschungsmethoden und Evaluation. Berlin u.a.O.: Springer.

Bossert, S. T. (1988): Cooperative Activities in the Classroom. Review of research in education, 15, pp. 225 – 250.

Bouma, J., Maast, A. J. (1988): Miniprojekte – konstruktive Vorschläge zur Belebung des Chemie-unterrichts. Praxis der Naturwissenschaften Chemie, 37, Nr.8, S. 31 – 35.

Bowen, C. W. (2000): A Quantitative Literature Review of Cooperative Learning Effects on High School and College Chemistry Achievement. Journal of Chemical education, 77, No.1, pp. 116 – 119.

Bromme, R. (1997): Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert: Psychologie des Unterrichts und der Schule, Kp. 4. Göttingen, Bern u.a.O.: Hogrefe. Kp. 4 [Enzyklopädie der Psychologie: Themenbereich D: Serie I Pädagogische Psychologie, Band 3, Hrsg. N. Birbaumer et al.].

Bromme, R. (1995): Was ist 'pedagogical content knowledge'? Kritische Anmerkungen zu einem fruchtbaren Forschungsprogramm. In S. Hopmann, K. Riquarts, (zus. m. W. Klafki & A. Krapp) (Hrsg.): Didaktik und/oder Curriculum. Zeitschrift für Pädagogik, Beiheft 33, S. 105 – 115.

Brophy, J. E., Good, T. L. (1986): Teacher Behavior and Student Achievement. In M. C. Wittrock (Ed.): Handbook of Research on Teaching, pp. 328 – 375. New York: Macmillan. 3rd ed.

Bühler, A., Graf, E. (2003): PISA: Konsequenzen für den Chemieunterricht. PH FR, Zeitschrift der Pädagogischen Hochschule Freiburg, Nr. 1, S. 23 – 25.

Bulmahn, E. (2002): Bildung und Innovation. Rede im Deutschen Bundestag am 13. Juni 2002.

Bybee, R. W. (1997): Toward an understanding of science literacy. In W. Gräber, C. Bolte (Eds.): Scientific Literacy, S. 37 – 68. Kiel: Institut für Pädagogik der Naturwissenschaften (IPN).

Bybee, R. W. (2002): Scientific Literacy – Mythos oder Realität? In W. Gräber, P. Nentwig, T. Koballa, R. Evans (Hrsg.): Scientific Literacy – Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung, S. 21 – 43. Opladen: Leske + Budrich.

Caroll, J. B. (1973): Ein Modell schulischen Lernens. Deutsch von G. und H. Tütgen, In W. Edelstein, D. Hopf (Hrsg.): Bedingungen des Bildungsprozesses. Stuttgart: Ernst Klett Verlag.

Clarke, J. (1994): Pieces of the Puzzle: The Jigsaw Method. In S. Sharan (Ed.): Handbook of Cooperative Learning Methods. Westport and London: Greenwood Press.

Clausen, M. (2002): Unterrichtsqualität: Eine Frage der Perspektive? Münster, New York u.a.O.: Waxmann.

Cohen, E. G. (1994a): Restructuring the Classroom: Conditions for Productive Small Groups. Review of Educational Research, 64, No.1, pp. 1 – 35.

Cohen, E. G. (1994b): Designing groupwork: Strategies for the heterogeneous classroom. New York: Teachers College Press.

Cohen, E. G. (1993): Bedingungen für kooperative Kleingruppen. In G. Huber (Hrsg.): Neue Perspek-tiven der Kooperation. Grundlagen der Schulpädagogik, Band 6. Hohengeren: Schneider-Verlag.

Dämmgen, U., Frühauf, D. (1985a): „Saurer Regen“ im experimentellen Chemieunterricht. Wie sauer ist saurer Regen? Probleme der pH-Wert-Messung. Niu-P/Ch, Nr. 2, S. 58 – 61.

Dämmgen, U., Frühauf, D. (1985b): Bodenversauerung und Kalkdüngung. NiU-P/Ch, Nr. 8, S. 291 – 295.

Dansereau, D. F. (1988): Cooperative learning strategies. In C. E. Weinstein, E. T. Goetz, P. A. Alexander (Eds.): Learning and study strategies: Issues in assessment, instruction and evaluation, pp. 103 – 120. Orlando, FL: Academic Press.

Deci, E. L., Ryan, R. M. (1993): Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. Zeitschrift für Pädagogik, 39, Nr. 2, S. 223 – 238.

Dettloff, C. (2003): Videoanalyse von Frontal- und Kleingruppenunterricht am Beispiel einer Problemstellung zum Themenbereich Säure-Base. Universität Essen: 1. Examensarbeit.

Deutsch, M. (1949): A theory of cooperation and competition. Human Relations, Nr. 2, pp.129 – 152.

Didi, H.-J., Fay, E., Kloft, C., Vogt, H. (1993): Einschätzung von Schlüsselqualifikationen aus psychologischer Perspektive - Gutachten im Auftrag des Bundesinstituts für Berufsbildung. Bonn: Institut für Bildungsforschung.

Diehl, J. M., Staufenbiel ,T. (2002): Statistik mit SPSS. Eschborn: Verlag Dietmar Klotz.

Ditton, H. (2002): Unterrichtsqualität – Konzeptionen, methodische Überlegungen und Perspektiven. Unterrichtswissenschaft, 30, S. 197 – 212.

Doll, J., Prenzel, M. (2001): Das DFG-Schwerpunktprogramm „Bildungsqualität von Schule“ (BIQUA): Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen. In Bundesministerium für Bildung und Forschung (Hrsg.): TIMSS – Impulse für schule und Unterricht, S. 99 – 103. Bonn: BMBF.

Doll, J., Euler, M., Leutner, D., Prenzel, M., Reiss, C., Renkl, A., Sumfleth, E. (2002): Internes Arbeitspapier. Kiel: IPN.

Dörner, D. (1976): Problemlösen als Informationsverarbeitung. Stuttgart: Kohlhammer.

Eilks, I. (2003): Kooperatives Lernen im Chemieunterricht (Teil 1), MNU, 56, Nr. 1., S. 51 – 55.

Eilks, I. (2003): Kooperatives Lernen im Chemieunterricht (Teil 2), MNU, 56, Nr. 2., S. 111 – 116.

Einsiedler, W. (1997): Unterrichtsqualität und Leistungsentwicklung: Literaturüberblick. In F. E. Weinert, A. Helmke (Hrsg.): Entwicklung im Grundschulalter, S. 225 – 240. Weinheim: Beltz.

Einsiedler, W. (2000): Von Erziehungs- und Unterrichtsstilen zur Unterrichtsqualität. In M. K. W. Schweer (Hrsg.): Lehrer-Schüler-Interaktion. Pädagogisch-psychologische Aspekte des Lehrens und Lernens in der Schule, S. 109 – 128. Opladen: Leske + Budrich.

Eisner, W., Gietz, P., Justus, A., Schierle, W., Sternberg, M. (1994): elemente chemie I. NRW - Unterrichtswerk für Gymnasien. Stuttgart u.a.O.: Ernst Klett.

Fischer H.-E., Horstendahl, M. (1997): Research in Science Education. Special Issue about European Research in Science Education, 27, Nr.3, S. 411 – 424.

Fischer, H.-E., Sumfleth, E. (2002): Physik- und Chemieunterricht nach PISA, S. 64 – 74. Stuttgart: Raabe Verlag.

Flanders, N. A. (1965): Teacher influence, pupil attitudes, and achievement. Washington D.C.: U.S. Office of Education.

Fraser, B. J., Tobin, K. G. (Eds.) (1998): International Handbook of Science Education. Dordrecht, Boston, London: Kluwer.

Fraser, B. J., Walberg, H. J., Welch, W. W., Hattie, J. A. (1987): Syntheses of Educational Productivity Research. International Journal of Educational Research, Vol.11, pp. 145 – 252.

Gage, N. L., Berliner, D. C. (1996): Pädagogische Psychologie. Weinheim: Beltz.

Gärtner, H.-J., Scharf, V. (2001): Chemische „Egg-Races“ in Theorie und Praxis. Online-Version unter: http://ekaestrs.bildung-rp.de/staff/gae/eggrace/chemrace.pdf

Gärtner, H.-J., Scharf, V. (1997): Egg Race – Kreativität beim Experimentieren im Chemieunterricht. In Friedrich Jahresheft 1997. Stuttgart: Ernst Klett Verlag.

Gärtner, H.-J., von Borstel, G. (2003): Kohlenstoffdioxid und Wettbewerb. Naturwissenschaften im Unterricht Chemie, 14, Nr. 5, S. 19 – 21.

Geisler, A. (1999): Chemielernen als Veränderung von Schülervorstellungen am Beispiel Säure-Base. Essen: Univ., Diss.

Glass, G. V., McGaw, B., Smith, M. L. (1990): Meta-Analysis in social research. Newbury Park u.a.O.: Sage Publ.

Goss, L. M., Eddleton, J. E. (2003): A Demonstration of Acid Rain and Lake Acidification: Wet Deposition of Sulfur Dioxide. Journal of Chemical Education, 80, No.1, p. 39.

Gramm, A. (1991): Chemieunterricht - ein handlungstheoretisches Konzept. chimcia didactica, 17, Nr. 4, S. 43 – 57.

Gramm, A. (2000): Chemieunterricht – unde venis? Chemie in der Schule, 47, Nr. 2, S. 109 – 112.

Gräsel, C., Gruber, H. (2000): Kooperatives Lernen in der Schule. In N. Seibert (Hrsg.): Unterrichts- methoden - Kontrovers. Bad Heilbrunn: Klinkhardt.

Gruber, H., Mandl, H., Renkl, A. (2000): Was lernen wir in Schule und Hochschule: Träges wissen? In H. Mandl, J. Gerstenmaier, J. (Hrsg.): Die Kluft zwischen Wissen und Handeln. Göttingen u.a.O.: Hogrefe.

Gruehn, S. (1995): Vereinbarkeit kognitiver und nicht-kognitiver Ziele im Unterricht. Zeitschrift für Pädagogik, 41, S. 531 – 553.

Gudjons, H. (Hrsg.) (1993): Handbuch Gruppenunterricht. Weinheim, Basel: Beltz.

Häcker, H., Stapf, K. H. (Hrsg.) (1998): Dorsch – Psychologisches Wörterbuch, 13. Auflage. Bern u.a.O.: H. Huber.

Hage, K., Bischoff, H., Dichanz, H., Eubel, K.-D., Oelschläger, H.-J., Schwittmann, D. (1985): Das Methodenrepertoire von Lehrern. Opladen: Leske + Budrich.

Hagemeister, V. (1999): Was wurde bei TIMSS erhoben? Rückfragen an eine standardisierte Form der Leistungsmessung. Die Deutsche Schule, 91, Nr. 2, S. 160 – 177.

Hager, W., Hasselhorn, M. (2000): Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In W. Hager, J.-L. Patry, H. Brezing (Hrsg.): Evaluation psychologischer Interventionsmaßnahmen, S. 41 – 85. Bern u.a.O.: H. Huber.

Hand, B. (1989): Student Understanding Of Acids and Bases: A Two Year Study. Research in Science Education, 19, pp. 133 – 144.

Hand, B. M., Treagust, D. F. (1988): Application of a Conceptual Teaching Strategy to Enhance Student Learning of Acids and Bases. Research in Science Education, 18, pp. 53 – 63.

Häusler, K., Pavenzinger, W. (1992): Unterricht Chemie – Stundenbilder, Experimente, Medien. Band 1: Säuren und Basen. Köln: Aulis Verlag.

Häusler, K., Worowka, A. (1987): Rotkohl-Blaukraut - ein idealer Universalindikator. Niu P/Ch, 27, S. 291 – 294.

Hasselhorn, M., Mähler, C. (2000): Transfer: Theorien, Technologien und empirische Erfassung. In W. Hager, J.-L. Patry, H. Brezing (Hrsg.): Evaluation psychologischer Interventionsmaßnahmen, S. 86 – 101. Bern u.a.O.: H. Huber.

Helmke, A., Weinert, F. E. (1997): Bedingungsfaktoren schulischer Leistungen. In N. Birbaumer et al. (Hrsg.): Enzyklopädie der Psychologie. Themenbereich D, Serie I, Band 3, Psychologie des Unterrichts und der Schule. Göttingen u.a.O.: Hogrefe.

Helmke, A., Schrader, F.-W. (1998): Determinanten der Schulleistung. In D. H. Rost (Hrsg.): Handwörterbuch Pädagogische Psychologie, S. 60 – 67. Weinheim: Beltz.

Helmke, A. (2002): Unterrichtsqualität und Unterrichtsklima: Perspektiven und Sackgassen. Unterrichtswissenschaft, 30, Nr. 3, S. 261 – 277.

Helmke, A. (2003): Unterrichtsqualität – Erfassen, Bewerten, Verbessern. Seelze: Kallmeyer.

Hermanns, R., Wambach, H. (1984): Funktionen des Experimentes im Chemieunterricht. Der Chemieunterricht, 15, Nr. 2, S. 29 – 52.

Hiebert, J., Gallimore, R., Garnier, H. et mult. al. (2003): Teaching Mathematics in Seven Countries. Results from the TIMSS 1999 Video Study. Washington D.C.: NCES, U.S. Department of Education.

Hirsch, U., Horlacher, B. (1987): Saure und alkalische Haushaltsreiniger im Unterricht. NiU-Physik-Chemie, 35, Nr. 25, S. 25 – 28.

Huber, A. (1999): Bedingungen effektiven Lernens in Kleingruppen unter besonderer Berücksichtigung der Rolle von Lernskripten. Schwangau: Verlag Ingeborg Huber.

Huber, G. L. (1991): Methoden des kooperativen Lernens. In E. Meyer, R. Winkel (Hrsg.): Unser Konzept: Lernen in Gruppen. Grundlagen der Schulpädagogik, Band 2. Hohengehren: Schneider-Verlag.

Ingenkamp, K., Schreiber, K. (1989): Was wissen unsere Schüler? Überregionale Lernerfolgmessung aus internationaler Sicht. Weinheim: Deutscher Studienverlag.

Jencks, C. (1972): Inequality. A reassessment of the effect of family and schooling in America. New York: Basic Books.

Johnson, D. W., Johnson, R. T. (1975): Learning Together and Alone. Englewood Cliffs, NY: Prentice Hall.

Johnson, D. W., Maruyama, G., Johnson, R., Neslon, D. (1981): Effects of cooperative, Competitive, and Individualistic Goal Structures on Achievement: A Meta-Analysis. Psychological Bulletin, 89, No.1, pp. 47 – 62.

Johnson, D. W., Johnson, R. T. (1989): Cooperation and Competition: Theory and Research. Edina, Minnesota: Interaction Book Company.

Johnson, D. W., Johnson, R. T. (1994a): The new circles of learning. Cooperation in Classroom and School. Alexandria, Virginia: ASCD.

Johnson, D. W., Johnson, R. T. (1994b): Structuring academic controvesity. In S. Sharan (Ed.): Handbook of cooperative learning methods, pp. 66 – 81. Westport & London: Greenwood Press.

Johnson, D. W., Johnson, R. T. (1994c): Learning Together. In S. Sharan (Ed.): Handbook of cooperative learning methods, pp. 51 – 65. Westport & London: Greenwood Press.

Johnson, D. W., Johnson, R. T., Smith, K. A. (1995): Cooperative Learning and Individual Student Achievement in Secondary Schools. In J. E. Pedersen, A. D. Digby (Eds.): Secondary Schools and Cooperative Learning. New York & London: Garland Publishing.

Kahl, R. (2002): Der Fehler ist das Salz des Lernens. Das Magazin – Wissenschaftszentrum Nordrhein-Westfahlen, Nr. 1, S. 9 – 11.

Kerlinger, F. N. (1986): Foundations of Behavioral Research. New York: Holt, Rinehart, Winston.

Kerr, N. L. (1983): Motivation losses in small groups: A social dilemma analysis. Journal of Personalityand Social Psychology, 45, pp.819 – 828.

Kessels, U., Hannover, B. (2000): Situational aktivierte Identität in koedukativen und monoedukativen Lerngruppen. In R. Brechel (Hrsg.): Zur Didaktik der Physik und Chemie. Alsbach: Leuchtturmverlag.

Kirsch, I., de Jong, J., LaFontaine, D., McQuenn, J., Mendelovits, J., Monsieur, C. (2002): Reading for Change – Performance and Engagement across Countries. Results from PISA 2000. Paris: OECD.

Klauer, K. J.(1999): Situated Learning: Paradigmenwechsel oder alter Wein in neuen Schläuchen? Zeitschrift für Pädagogische Psychologie, 13, S. 117 – 121.

Klieme, E., Clausen, M. (1999): Identifying Facets of Problem Solving in Mathematics Instruction. Paper presented at the AERA Annual Meeting in Montreal.

Kogelnik, H.-J. (1982): Obst- und Gemüsesäfte als Farbindikatoren im Unterricht. PdN-Ch, 4, S. 123 – 125.

Kogelnik, H.-J.(1983): Frucht- und Gemüsesäfte als Säure-Base-Farbindikatoren. NiU-P/Ch, 5, S. 183 – 185.

Körner, H.-D., Breuer, D. (2004): Tandemgespräche – Konzeption einer neuen Unterrichtsmethode zum kooperativen Lernen in der Chemie. In A. Pitton (Hrsg.): Gesellschaft für Didaktik der chemie und Physik. Münster: Lit-Verlag.

Krauth, J. (1995): Testkonstruktion und Testtheorie. Weinheim: Psychologie Verlags Union.

Lange, H. (1999): Qualitätssicherung in Schulen. Die Deutsche Schule, 91, Heft, 2, S. 144 – 159.

Lange, H. (2001): Die Bildungspolitische Bedeutung von Schulleistungsvergleichen. In G. Kaiser, N. Knoche, D. Lind, W. Zillmer (Hrsg.): Leistungsvergleiche im Mathematikunterricht, S. 1 – 28. Hildesheim: Franzbecker.

Langensiepen, B. (1996): Zu »Thesen zur Weiterentwicklung der gymnasialen Oberstufe«. MNU, 49, Nr. 2, S. 118.

Lazarowitz, R., Hertz-Lazarowitz, R. (1998): Cooperative Learning in the Science Curriculum. In B. J. Fraser, K. G. Tobin (Eds.): International Handbook of Science Education, pp. 449 – 469. Dordrecht, Boston, London: Kluwer.

Landis, J. R., Koch, G. G. (1977): The Measurement of Observer Agreement for Categorial Data. Biometrics, 33, pp. 159 – 174.

Lehmann, L. (2002): Leitkartengesteuerter Chemieunterricht – Die Leika-Methode. Chemie und Schule, 17, Nr. 1, S. 2 – 5.

Leutner, D., Leopold, C. (2001): Selbstregulation von Lernstrategien: Förderung der Fähigkeit zur selbständigen Regulation des zielorientierten Einsatzes von Lernstrategien für den Erwerb von Wissen aus mathematisch-naturwissenschaftlichen Sachtexten. Zwischenbericht an die DFG.

Leutner, D., Fischer, H.-E. (2002): Diagnose und Förderung von Lern- und Problemlöseprozessen Im naturwissenschaftlichen Unterricht. Antrag auf Gewährung einer Sachbeihilfe, eingereicht bei der DFG.

Lienert, G., Raatz, U. (1998): Testaufbau und Testanalyse. Weinheim: Beltz.

Marcet, J. (1839): Unterhaltungen über Chemie, in welchen die Anfangsgründe dieser nützlichen Wissenschaft allgemein verständlich erklärt werden können. Weinheim: VCh.

Max-Planck-Gesellschaft (Hrsg.) (1986): Max-Planck-Gesellschaft - Jahrbuch 1986, S. 821 – 828. Göttingen: Vandenhoeck & Ruprecht.

Mayring, P. (2000): Qualitative Inhaltsanalyse. Weinheim: Deutscher Studienverlag.

Melenk, H. (2000): Einführende Anmerkungen zum nutzen empirischer fachdidaktischer Forschung. In R. Duit, C. v. Rhöneck (Hrsg.): Ergebnisse fachdidaktischer und psychologischer Lehr-Lern-Forschung, S. 11 – 16. Kiel: IPN.

Meyer, H. (1987): Unterrichtsmethoden I: Theorieband. Frankfurt a. M.: Cornelsen Scriptor.

Morguet, M. (1992): Umgang mit komplexen Problemen in Arbeitsgruppen. Konstanz: Hartung-Gorre.

Müller, A., Nieswandt, M. (1999): Arbeitsleitfaden zu Modul 3: „Aus Fehlern lernen“. Kiel: IPN.

Müller, C., Seidel, T. (2001): Transkriptions-Manual, basierend auf dem Manual des Pädagogischen Instituts PPII der Universität Zürich. In M. Prenzel, R. Duit, M. Euler, M. Lehrke, T. Seidel (Hrsg.): Erhebungs- und Auswertungsverfahren des DFG-Projekts „Lehr-Lern-Prozesse im Physikunterricht – eine Videostudie“, S. 27 – 40. Kiel: IPN.

Musahl, H.-P., Schwennen, C. (2001): Versuchsplanung. In G. Wenninger (Hrsg.): Lexikon der Psychologie. Heidelberg u. a.O.: Spektrum.

Nachtigall, C., Wirtz, M. (1998): Wahrscheinlichkeitsrechnung und Inferenzstatistik. Weinheim u. München: Juventa Verlag.

Nay, U., Glatzel, D. (1978): Schulkindliche Identifizierung von Säuren und Laugen. Sachunterricht und Mathematik in der Primarstufe, Nr. 6, S. 244 – 249.

Neber, H. (1998): Kooperatives Lernen. In J. Rost (Hrsg.): Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz.

Nicolai, N. (2004): Analyse der häuslichen Lernsituation im Chemieunterricht. Essen, unveröffentlichte Ergebnisse.

Niedderer, H. (1996): Überblick über Lernprozessstudien in Physik. In R. Duit, C. Reinders, C. von Rhöneck (Hrsg.): Lernen in den Naturwissenschaften, S. 119 – 144. Kiel: IPN.

O’Donell, A., M., O’Kelly, J. (1994): Learning from peers. Beyond the rhetoric of positive results. Educational Psychology Review, 6, pp. 323 – 349.

Palinscar, A. S., Brown, A. L. (1984): Reciprocal teaching of comprehension-fostering and com-prehension-monitoring activities. Cognition and Instruction, 1, pp.117 – 175.

Parchmann, I., Ralle, B., Demuth, R. (2000): Chemie im Kontext – Eine Konzeption zum Aufbau und zur Aktivierung fachsystematischer Strukturen in lebensweltlichen Kontexten. MNU, 53, Nr. 3., S. 132 – 137.

Pauli, C., Reusser, K., Waldis, M., Grob, U. (2003): “Erweiterte Lernformen” im Mathematikunterricht der Deutschschweiz. Unterrichtswissenschaft, 31, Nr. 4, S. 291 – 320.

Piaget, J. (1926): The language and thought of the child. New York: Harcourt Brace.

Pfeifer, P., Häusler, K., Lutz, B. (1997): Konkrete Fachdidaktik Chemie. München: Oldenbourg Verlag.

Postlethwaite, T. N., Weiler, H., Roeder, H. M. (Hrsg.) (1980): Schulen im Leistungsvergleich. Stuttgart: Klett-Cotta.

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. 192 – 248. Opladen: Leske + Budrich.

Qin, Z., Johnson, D. W., Johnson, R. T. (1995): Cooperative versus competitive Efforts and ProblemSolving. Review of Educational Research, 65, No. 2, pp. 129 – 143.

Ralle, B., Eilks, I. (Eds.) (2002): Research in Chemical Education – What does this mean? Proceedings of the 16th Symposium on chemical Education held at the University of Dortmund, 22-24 May 2002. Aachen: Shaker Verlag.

Rampf, H. (1987): Veranschaulichung des pH-Wertes in der Sekundarstufe I. NiU-P/Ch, 27, S. 304 – 305.

Renkl, A. (1995): Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. Learning and Instruction, Nr. 5, pp. 21 – 36.

Renkl, A. (1996): Träges Wissen: Wenn Erlerntes nicht genutzt wird. Psychologische Rundschau, 47, S. 78 – 92.

Renkl, A., Gruber, H., Mandl, H. (1996): Kooperatives problemorientiertes Lernen in der Hochschule. In J. Lompscher, H. Mandl (Hrsg.): Lehr- und Lernprobleme im Studium. Bern u.a.O.: H. Huber.

Renkl, A. (1997): Lernen durch Lehren. Wiesbaden: Deutscher Universitätsverlag.

Renkl, A. (1998): Träges Wissen. In D. H. Rost (Hrsg.): Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz.

Reusser, K., Pauli, C. (2003): Mathematikunterricht in der Schweiz und in weiteren sechs Ländern. Bericht mit Videobeispielen über die Ergebnisse einer internationalen und schweizerischen Video-Unterrichtsstudie. Universität Zürich: Pädagogisches Institut.

Reyer, T. (2003): Oberflächenmerkmale und Tiefenstrukturen im Unterricht. Universität Dortmund: Dissertation.

Rheinberg, F., Bromme, R., Minsel, B., Winteler, A., Weidenmann, B. (2001): Die Erziehenden und Lehrenden. In A. Krapp, B. Weidenmann (Hrsg.): Pädagogische Psychologie, S. 271 – 355. Weinheim: Beltz.

Rosenshine, B. V. (1979): Content, time, and direct instruction. In P. L. Peterson, H. J. Walberg (Eds.): Research on teaching, pp. 28 – 56. Berkley: McCutchan.

Rost, J. (1996): Testtheorie und Testkonstruktion. Bern u.a.O.: Hans Huber.

Rost, D. H. (Hrsg.) (1998): Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz.

Rost, J. (2000): Allgemeine Standards für die Evaluationsforschung. In W. Hager, J.-L. Patry, H. Brezing (Hrsg.): Evaluation psychologischer Interventionsmaßnahmen, S. 129 – 140. Bern u.a.O.: H. Huber.

Rotering-Steinberg, S., von Küngelen, T. (1986): Ergebnisse einer schriftlichen Befragung zum Gruppenunterricht. In Erziehungswissenschaft – Erziehungspraxis, 2, Nr. 1, S. 27 – 30.

Sarris, V. (1992): Methodologische Grundlagen der Experimentalpsychologie, Band 2: Versuchsplanung und Stadien. München, Basel: E. Reinhardt.

Schallies, M. (1993): Videoaufzeichnung als Hilfsmittel für die Gestaltung von freien Lösungsmöglichkeiten im experimentellen Chemieunterricht. In P. Buck, W. Ulrich (Hrsg.): Video in Forschung und Lehre, S. 274 – 285. Weinheim: Deutscher Studien Verlag.

Schallies, M. (1999): Chemieunterricht weiterentwickeln. In E. Sumfleth (1999): Chemiedidaktik im Wandel – Gedanken zu einem neuen Chemieunterricht, S. 28 – 48. Münster: Lit-Verlag.

Scharf, V., Gärtner, H. J. (1994): Konkurrenz und Kooperation – Gruppenarbeit als pädagogische Herausforderung im naturwissenschaftlichen Unterricht. In H. Behrendt (Hrsg.): Zur Didaktik der Physik und Chemie. Alsbach: Leuchtturm-Verlag.

Schecker, H., Klieme, E. (2001): Mehr Denken, weniger Rechnen. Physikalische Blätter, 57, Nr. 7/8, S. 113 – 117.

Schmidkunz, H., Lindemann, H. (1995): Das forschend-entwickelnde Unterrichtsverfahren. Magdeburg: Westarp Wissenschaften.

Schümer, G. (1998): Mathematikunterricht in Japan – Ein Überblick über den Unterricht in öffentlichen Grund- und Mittelschulen und privaten Ergänzungsschulen. Unterrichtswissenschaft, 26, S. 195 – 228.

Seidel, T. (2002): Lehr-Lernskripts im Unterricht. Kiel: Univ. Diss.

Seidel, T., Dalehefte, I. M., Meyer, L. (2001): Richtlinien für die Videoaufzeichnung. In M. Prenzel, R. Duit, M. Euler, M. Lehrke, T. Seidel (Hrsg.): Erhebungs- und Auswertungsverfahren des DFG-Projekts „Lehr-Lern-Prozesse im Physikunterricht – eine Videostudie“, S. 5 – 26. Kiel: IPN.

Seidel, T. (2003): Videobasierte Kodierverfahren in der IPN Videostudie Physik – ein Überblick. In T. Seidel, M. Prenzel, R. Duit, M. Lehrke (Hrsg.): Technischer Bericht zur Videostudie „Lehr-Lernprozesse im Physikunterricht, S. 99 – 11. Kiel: IPN.

Sharan, S., Hertz-Lazarowitz, R. (1980): A group investigation method of cooperative learning in the classroom. In S. Sharan, P. Hare, C. Webb, R. Hertz-Lazarowitz (Eds.): Cooperative in education, pp. 14 – 46. Provo, UT: Brigham Young University Press.

Sharan, S., Sharan, Y. (1992): Expanding cooperative Learning through Group Investigation. New York: Teachers College Press.

Sharan, Y., Sharan, S. (1994): Group Investigation in the cooperative Classroom. In S. Sharan (Ed.): Handbook of Cooperative Learning Methods. Westport CT and London: Greenwood Press.

Shuell, T. J. (1996): Teaching and Learning in the Classroom Context. In D. C. Berliner, R. C. Calfee (Eds.): Handbook of Educational Psychology, pp. 726 – 764. New York: Macmillan.

Sisovic, D., Bojovic, S. (2000): Approaching the Concepts of Acids and Bases by Cooperative Learning. Chemistry Education: Research and Practice in Europe, 1, No. 2, pp. 263 – 275.

Slavin, R. E. (1978): Student teams and achievement divisions. Journal of Research and Development in Education, 12, pp. 39 – 49.

Slavin, R. E. (1983): Cooperative Learning. New York: Longman.

Slavin, R. E. (1989): A theory of school and classroom organization. In R. E. Slavin (Ed.): School and classroom organization. Hillsdale, NJ: Erlbaum.

Slavin, R. E. (1993): Kooperatives Lernen und Leistung. In G. Huber (Hrsg.): Neue Perspektiven der Kooperation. Hohengehren: Schneider-Verlag.

Slavin, R. E. (1995): Cooperative Learning: theory, research and practise. Boston, London u.a.O.: Allyn and Bacon, 2nd ed.

Slavin, R. E. (1996): Research on cooperative learning and achievement: What we know, what we need to know. Contemporary Educational Psychology, 21, pp.43 – 69.

Slavin, R. E. (1997): Research on Cooperative Learning and Achievement: A Quarter Century of Re-search. Paper presented at the annual meeting of the German Psychologist association, Frankfurt.

Springer, L., Stanne, M. E., Donnovan, S. S. (1999): Effects of Small-Group Learning on Under-graduates in Science, Mathematics, Engineering, and Technology: A-Meta-analysis. Review of Educational Research, 69, No.1, pp. 21 – 51.

Stanat, P., Kunter, M. (2001): Kooperation und Kommunikation. In J. Baumert et al. (Hrsg.): PISA 2000, S. 299 – 322. Opladen: Leske + Budrich.

Stigler, J., Fernandez, C. (1995): Videotape Classroom Study. Washington D.C.: NCES IEA, U.S. Department of Education.

Stigler, J. W., Gonzales, P., Kawanka, T., Knoll, S., Serrano, A. (1999): The TIMSS Videotape Classroom Study. Washington D.C.: NCES, U.S. Department of Education.

Sumfleth, E. (1987): „Über den Zusammenhang zwischen Schulleistung und Gedächtnisstruktur - Eine Untersuchung zum Thema Säure-Base-Theorien“; NiU-P/Ch, 35, S. 29 – 35.

Sumfleth, E. (1988): Lehr-Lernprozesse im Chemieunterricht. Frankfurt a.M.: Peter Lang.

Sumfleth, E. (1996): Chemieverstehen in Abhängigkeit vom Vorwissen. In R. Duit, C. von Rhöneck (Hrsg.): Lernen in den Naturwissenschaften, S. 228 – 249. Kiel: IPN.

Sumfleth, E., Geisler, A. (1999): „Veränderungen von Schülervorstellungen im Bereich Säuren und Basen“. In R. Brechel (Hrsg.): Zur Didaktik der Chemie und Physik, S. 205 – 207. Alsbach: Leuchtturm Verlag.

Sumfleth, E., Ploschke, B., Geisler, A. (1999): Schülervorstellungen und Unterrichtsgespräche zum Thema Säure-Base. In E. Sumfleth (Hrsg.): Chemiedidaktik im Wandel – Gedanken zu einem neuen Chemieunterricht, S. 91 – 115. Münster: Lit-Verlag.

Sumfleth, E., Telgenbüscher, L. (2000): Zum Einfluss von Bildmerkmalen und Fragen zum Bild beim Chemielernen mit Hilfe von Bildern – Beispiel Massenspektrometrie. Zeitschrift für Didaktik der Naturwissenschaften (ZfDN), 6, S. 59 – 78.

Sumfleth, E., Wild, E. (2000): Schulische und familiale Bedingungen des Lernens und Lernmotivation im Fach Chemie: Evaluation eines integrierten Interventionskonzeptes zur Säure-Base-Thematik. Antrag auf Gewährung einer Sachbeihilfe eingereicht bei der DFG.

Sumfleth, E., Wild, E. (2001): Schulische und familiale Bedingungen des Lernens und Lernmotivation im Fach Chemie: Evaluation eines integrierten Interventionskonzeptes zur Säure-Base-Thematik. Zwischenbericht an die DFG.

Sumfleth, E., Wild, E., Rumann, S., Exeler, J. (2002): Wege zur Förderung der naturwissenschaftlichen Grundbildung im Chemieunterricht: Kooperatives Problemlösen im schulischen und familialen Kontext zum Themenbereich Säure-Base. Zeitschrift für Pädagogik, 45. Beiheft, S. 207 – 221.

Sumfleth, E., Rumann, S. und Nicolai, N. (2004): Kooperatives Arbeiten im Chemieunterricht im schulischen und häuslichen Kontext. Münster: Waxmann im Druck.

Tausch, M., von Wachtendonk, M. (1996): Stoff – Formel – Umwelt. Bamberg: C. C. Buchner.

Terhart, E. (2000): Lehr-Lern-Methoden. Weinheim u. München: Juventa.

Tesch, M., Duit, R. (2001): Videoanalysen – Kategorien zur Analyse des Experiments. In M. Prenzel, R. Duit, M. Euler, M. Lehrke, T. Seidel (Hrsg.): Erhebungs- und Auswertungsverfahren des DFG-Projekts „Lehr-Lern-Prozesse im Physikunterricht – eine Videostudie“, S. 95 – 110. Kiel: IPN.

Tesch, M. (2003): Experimente im Physikunterricht. In T. Seidel, M. Prenzel, R. Duit, M. Lehrke (Hrsg.): Technischer Bericht zur Videostudie „Lehr-Lernprozesse im Physikunterricht, S. 129 – 150. Kiel: IPN.

Toulmin, S. (1996): Der Gebrauch von Argumenten. Weinheim: Beltz Athenäum Verlag.

Walberg, H. J. (1986): Synthesis of research on teaching. In M. C. Wittrock (Ed.): Handbook of Research on Teaching, pp. 214 – 229. New York: Macmillan.

Walberg, H. J. (1981): a psychological theory of educational productivity. In F. H. Farley, N. J. Gordon (Eds.): Psychology and Education. The state of union, pp. 81 – 108. Berkeley, CA: McCutchan.

Walpuski, M. (2002): Entwicklung eines Kategoriensystems zur Interaktions- und Argumentationsanalyse von Kleingruppenarbeitsphasen im Chemieunterricht. Universität Essen:1. Examensarbeit.

Walpuski, M. (2003): Gestaltung der Kleingruppensituationen zum DFG-Antrag „Prozessorientiertes Experimentieren in kooperativer Kleingruppenarbeit“. Universität Essen, unveröffentlichtes Konzeptpapier.

Wang, M. C., Haertel, G. D., Walberg, H. J. (1990): What Influences Learning? A Content Analysis of Review Literature. Journal of Educational Research, 84, No.1, pp. 30 – 43.

Wang, M. C., Haertel, G. D., Walberg, H. J. (1993): Toward a Knowledge Base for School Learning. Review of Educational Research, 63, No.3, pp. 249 – 294.

Webb, N. M., Palinscar, A. S. (1996): Group Processes in the Classroom. In C. D. Berliner, R. C. Calfee (Eds.): Handbook of educational psychology, pp. 841 – 873, New York: Macmillian.

Weinert, F. E., Schrader, F. W., Helmke, A. (1989): Quality of Instruction and Achievement Outcomes. International Journal of Educational Research, 13, pp. 895 – 914.

Weinert, F. E., Helmke, A. (Hrsg.) (1997): Entwicklung im Grundschulalter. Weinheim, Basel: Beltz.

Weinert, F. E. (1998): Psychologische Theorienbildung auf dem Prüfstand. Zeitschrift für Pädagogische Psychologie, 12, S. 205 – 209.

Wiechmann, J. (Hrsg.) (2000): Zwölf Unterrichtsmethoden. Weinheim: Beltz.

Wild, K.-P. (1999): Manual zu CatMovie - Version 4.08. www.catmovie.de

Wild, K.-P. (2003): Videoanalyse als neuer Impulsgeber für eine praxisnahe prozessorientierte empirische Unterrichtsforschung. Unterrichtswissenschaft, 31, Nr. 2, S. 98 – 102.

Wild, E., Exeler, J., Gerber, J. (2003): Auswertungsergebnisse des Jugendlichenfragebogens. Unveröffentlichte Ergebnisse.

Wirtz, M., Caspar, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen u.a.O.: Hogrefe.

Wittrock, M. C. (1986): Students' Thought Processes. In M. Wittrock (Ed.): Handbook of Research on Teaching, pp. 297 – 314. New York: Macmillan.

Woest, V. (1995): Offener Chemieunterricht. Konstruktion – Erprobung – Bewertung. Alsbach: Leuchtturm-Verlag.

Wykotski, L. S. (1978): Mind in society: The development of higher psychological processes. Cam-bridge, MA: Havard University Press.

10. Abbildungsverzeichnis

		Seite
Abb. 2.1:	Modell des schulischen Lernens von Bloom	12
Abb. 2.2.	Lernerfolg in Abhängigkeit von der Diagnosefähigkeit des Lehrers und dem Strukturierungsgrad des Unterrichts	16
Abb. 2.3	Angebots-Nutzungs-Modell des Unterrichts nach Helmke 2002	19
Abb. 3.1	Dimensionen der Unterrichtsmethode	22
Abb. 3.2	Häufigkeitsverteilung von 30 Effektgrößen zum Einfluss kooperativer Lernformen auf die Leistung	29
Abb. 3.3	Häufigkeitsverteilung der relativen methodischen Qualität von Studien zum kooperativen Arbeiten	30
Abb. 3.4	Aufteilung der TGT-Turniertische	35
Abb. 4.1	Stellenwert von Variablen in einer Untersuchung	42
Abb. 4.2	One-group before-after design	42
Abb. 4.3	Übersicht experimenteller Versuchspläne	43
Abb. 4.4	Auswirkung der Versuchsanordnung auf die interne bzw. externe Validität	44
Abb. 4.5	Unabhängige und abhängige Variablen der Untersuchung	45
Abb. 4.6	2*2-Design der Unterrichts- und Hausaufgabenintervention	49
Abb. 4.7	Konsistenz der Proportion „Säure- pH-Wert“ in Abhängigkeit von einem dritten differierenden Begriff	52
Abb. 4.8	Kamerapositionen während der Unterrichtsintervention	54
Abb. 4.9	Arbeitsplan während der Intervention	55
Abb. 5.1	Verlaufsschema der Unterrichtsintervention	57
Abb. 5.2	Wege zum Säure-Base Begriff	60
Abb. 5.3	Inhaltlicher Verlauf der Unterrichtsintervention	61
Abb. 5.4	Stundenphasierung der Unterrichtsreihe	62
Abb. 5.5	Improvisierte Petroleumlampe	64
Abb. 5.6	Instruktionale Hilfe zur zweiten Stunde	69
Abb. 5.7	Instruktionale Hilfe zur dritten Stunde	71
Abb. 5.8	Zahlenstrahl der pH-Skala	73
Abb. 5.9	Informationstext zum sauren Regen	76
Abb. 5.10	Hinweise zu den Feststoffen der sechsten Stunde	78
Abb. 6.1	Vergleich der Leistungseingangsdaten	82
Abb. 6.2	Haupteffekt >Zeit< der vier Treatments	84
Abb. 6.3	Leistungsentwicklung „offene Fragen“	86
Abb. 6.4	Leistungsentwicklung „offene Fragen“ $t_0 – t_2$	87

Abb. 6.5	Interaktionsdiagramm des Faktors >Gruppe<	88
Abb. 6.6	Interaktionsdiagramm des Faktors >Zeit<	88
Abb. 6.7	Leistungsentwicklung Aufgabe 1	91
Abb. 6.8	Offene Fragen, Aufgabe 5	92
Abb. 6.9	Leistungsentwicklung Aufgabe 5	92
Abb. 6.10	Haupteffekt >Zeit< der vier Treatments im Multiple-Choice-Test	95
Abb. 6.11	Diagramm zur One-Way-ANOVA bezogen auf die Ergebnisse des Multiple-Choice-Test	96
Abb. 6.12	Interaktionsdiagramme der Haupteffekte >Hausaufgabenintervention< und >Unterrichtsintervention<	97
Abb. 6.13	Leistungsentwicklung der Gruppen U^+/U	98
Abb. 6.14	Leistungsverhältnis von U^+ gegenüber U^- nach Post- / Follow-up-Test (offene Aufgaben, z-standardisiert)	100
Abb. 6.15	Verteilung der Zeugnisnoten der Probanden im Fach >Chemie< vor der Intervention	101
Abb. 6.16	Lernzuwachs (pre – follow-up) differenziert nach Fachleistung vor der Intervention	102
Abb. 6.17	Lehrereffekte „offene Aufgaben“	105
Abb. 7.1	Screenshot aus dem Transkriptionsmodus	112
Abb. 7.2	Kategorien des Experimentierens. Verändert nach Tesch (2003) – In der Intervention realisierte Kategorien grau unterlegt	114
Abb. 7.3	Kategorieschema der „Inhaltlichen Progressionsanalyse“ (IPA)	116
Abb. 7.4	Argumentationsschema nach Toulmin (1974)	117
Abb. 7.5	Beispiel zu Toulmins Argumentationsschema	118
Abb. 7.6	Kategorienverteilung der Facette >Fokus<	121
Abb. 7.7	Kategorienverteilung der Facette >Fokus< von Gruppe 1	122
Abb. 7.8	Prozentualer Anteil inhaltlicher turns, differenziert nach Stunde und Gruppe	123
Abb. 7.9	Entwicklungsverlauf fachwissenschaftlicher und lebensweltlicher turns von Gruppe 3	124
Abb. 7.10	Vergleich der Facette >Bezugsrahmen zwischen der ersten und der sechsten Unterrichtsstunde	124
Abb. 7.11	Entwicklungsverlauf theoretischer turns von Gruppe 2	126
Abb. 7.12	Kategorienverlauf der Facette >Problemelemente< der Gruppe 1, vierte Stunde	127
Abb. 7.13	Qualitative Prozessanalyse einer Kleingruppensituation	130
Abb. 7.14	Qualitative Prozessanalyse einer Frontalunterrichtssituation	131

11. Tabellenverzeichnis

		Seite
Tab. 2.1	Schulleistungsrelevante Instruktionsdeterminanten	15
Tab. 3.1	Effekte kooperativer, kompetitiver und individualisierter Unterrichtsbedingungen auf die Lernleistung	31
Tab. 3.2	Typologie kooperativer Lernschemata	33
Tab. 4.1	Probandenverteilung auf die Untersuchungszellen	49
Tab. 4.2	Ausgewählte Skalen der eingesetzten Fragebögen	51
Tab. 4.3	Verteilung der Klassen auf die Treatments und Erhebungszeiträume	55
Tab. 5.1	Lebensweltliche Kategorisierung der eingesetzten Lösungen	67
Tab. 5.2	Farbreaktion der eingesetzten Lösungen mit Rotkohlsaft	68
Tab. 5.3	pH-Werte der eingesetzten Lösungen	72
Tab. 6.1	Deskriptive Statistiken der Daten des Eingangstests	81
Tab. 6.2	Offene Fragen - Deskriptive Statistiken der Varianzanalyse mit Messwiederholung	85
Tab. 6. 3	Paarweiser Vergleich der Untersuchungsbedingungen U^+/H^- gegen U^-/H^+	89
Tab. 6.4	Mittlere Leistungsergebnisse der >offenen Aufgaben< differenziert nach Aufgabe und Messzeitpunkt	90
Tab. 6.5	Multiple-Choice - Deskriptive Statistiken der Varianzanalyse mit Messwiederholung	94
Tab. 6.6	Zusammensetzung der Extremgruppen in den Leistungstests	99
Tab. 6.7	Paarweise Lehrervergleiche der Leistungszunahme (T-Werte)	105
Tab. 6.8	Interaktionen (Intervention*Zeit) und Effekte auf intrapersonaler Ebene	106
Tab. 6.9	F-Werte und Effektgrößen der Interaktion >Lehrer*Zeit<	107
Tab. 7.1	Gesamtturns der Gruppenarbeitsphasen	119
Tab. 7.2	Summen der inhaltlichen turns der Gruppenarbeitsphasen	120
Tab. 7.3	Beurteilerreliabilität der Inhaltlichen Progressionsanalyse	120
Tab. 7.4	Anzahl fachwissenschaftlicher (fw) und lebensweltlicher (lw) turns	123
Tab. 7.5	Anzahl theoretisch expliziter (t_{ex}) und theoretisch impliziter (t_{im}) turns	125
Tab. 7.6	Anzahl differenzierender (D), integrierender (I) und redundanter (R) turns	126
Tab. 7.7	Mittelwerte der Facette >Problemelemente<	127

12. Anhang

A. Allgemeine Lehrerinformation zur Unterrichtsintervention

Allgemeine Anmerkungen zur Intervention:

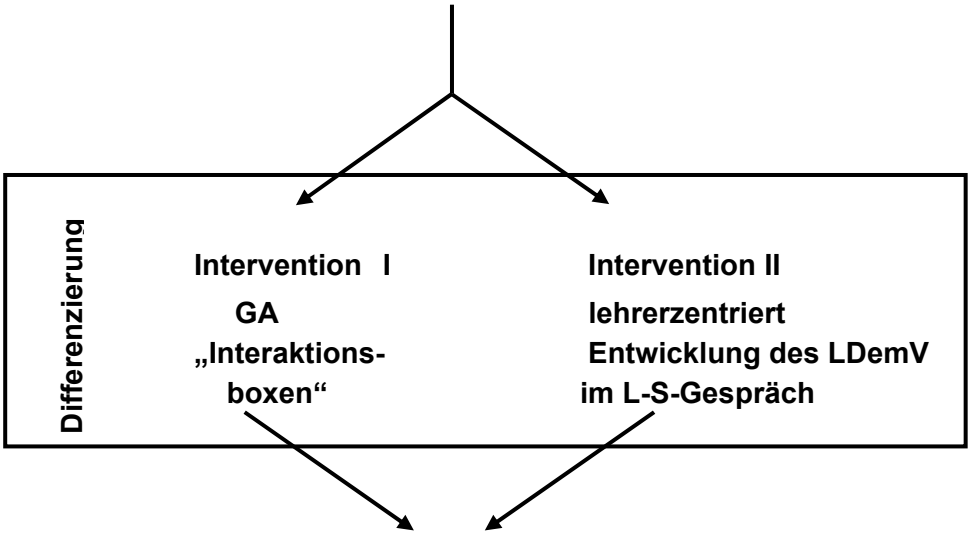
- Benötigtes Vorwissen:
  Element (Metall/ Nichtmetall) – Reinstoff – Stoffgemisch – Lösung; Oxide als Verbrennungsprodukte

Zu den Unterrichtsstunden der Intervention

- Da am Ende einer jeden Stunde ein kurzer Fragebogen (eine Seite) von den SchülerInnen ausgefüllt wird, muss der Unterricht ca. 3 Minuten vor dem Klingeln beendet sein.
- Jeder Lehrer unterrichtet in einer Welle eine Klasse interveniert und eine Klasse traditionell.
- Wichtig: die beiden Klassen einer Welle müssen inhaltlich synchron laufen, d. h. die effektive Lernzeit muss in beiden Klassen gleich sein!
- Charakteristikum der zweiten Unterrichtsintervention ist eine lehrerzentrierte Gesprächsführung. Experimente werden im Unterrichtsgespräch entwickelt und als LDemV oder SDemV vorgeführt.
- In der Interventionsgruppe I soll neben der Phase der Gruppenarbeit auch in der Auswertungsphase Wert auf einen möglichst hohen Redeanteil durch die SchülerInnen gelegt werden.
- Fachsprachlich sind für die Unterrichtsreihe zwei Formulierungen bedeutsam:
  - prinzipiell ist von sauren und basischen Lösungen zu sprechen, nicht von Säuren und Basen. Der Begriff „Lauge“ stellt eine umgangssprachliche Bezeichnung für eine basische Lösung dar, der Begriff „alkalische Lösung“ sollte wenn möglich vermieden werden oder ggfs. als Synonym für eine basische Lösung dargestellt werden.
- Der Zusammenhang von „pH-Wert“ und „Stärke“ einer sauren/basischen Lösung ist problematisch (Bsp.: Der pH-Wert gibt die Stärke einer sauren Lösung (noch schlimmer: Säure) an.). In der Oberstufe ist dann kaum noch ersichtlich, dass der pH-Wert einer starken Säure, niedriger Konzentration, geringer sein kann als der pH-Wert einer schwachen Säure mit hoher Lösungskonzentration. Die Begriffe Säurestärke und pH-Wert müssen ja nicht korrelieren, sind aber auf keinen Fall synonym.
- Besser geeignet ist daher:
  - Der pH-Wert ist ein Maß für die Konzentration einer sauren oder basischen Lösung.
  - Der pH-Wert gibt das Ausmaß des sauren/ basischen Charakters einer Lösung an.

Ein allgemeines Ablaufschema lautet:

- Einleitung /Besprechung der Hausaufgaben (~ 10min)
- Problemstellung durch den Lehrer (~ 2 min)
- Erarbeitung: Planung und ggf. Durchführung eines Experiments (~15 –20 min)



- Auswertung / Sicherung (~10 – 15 min)
  (Σ = 42min)
- Stundeninteressentest (3 min)

Zur Gliederung der Unterrichtsreihe:

1.Std. Charakterisierung von Haushaltslösungen:
Saure und basische Lösungen als Schema der Klassifikation.

2. Std. Zuordnung der Haushaltlösungen mit Indikatoren:
Farbumschlagsreaktionen mit Rotkohlsaft – hier oder unter 1.: neutrale Lösungen als dritte Gruppe.

3. Std. Der pH-Wert zur Charakterisierung der Konzentration einer Lösung:
Abstufung der unter 1. und 2. eingesetzten Lösungen hinsichtlich ihres pH-Werts.

4. Std. Die Neutralisationsreaktion:
Aufhebung der sauren/ basischen Eigenschaft einer Lösung.

5. Std. Reaktion von Nichtmetalloxiden mit Wasser:
Bildung einer sauren Lösung.

6. Std. Reaktion von Metalloxiden mit Wasser:
Bildung einer basischen Lösung (zur Neutralisation einer sauren Lösung).

B. Unterrichtsskripte Stunden 1- 6

Erste Unterrichtsstunde
Materialien:
großes Becherglas (500mL), Zucker, Glasstab, 7 Chemikalientabletts

Thema: Charakterisierung von Haushaltslösungen

Einstieg:
Lehrer informiert Schüler, dass das Thema der heutigen Stunde „Lösungen“ lautet, daher: Wiederholung des Lösungsbegriffes. LDemV: Zucker wird in einem Becherglas aufgelöst. Schüler beschreiben den Vorgang unter Verwendung der Fachsprache

Tafel

Flüssigkeiten, die Feststoffe, andere Flüssigkeiten oder Gase lösen, heißen Lösemittel. Wasser ist eines der wichtigsten Lösemittel. Die Löslichkeit eines Stoffes in Wasser ist eine messbare Stoffeigenschaft.

An dieser Stelle kann auch - nach Vorwissen der Schüler - über den Auflösevorgang an sich gesprochen und die gängigen Schülervorstellungen diskutiert werden. (Ist der Zucker verschwunden? Was passiert mit den Zuckerteilchen?)

Hinführung zur Erarbeitung:
Klassifizierung der zwölf Lösungen.
Zur Einleitung kann darauf verwiesen werden, dass folgende Aufgabe in einer abgespeckten Version bereits Bestandteil des Vortests war, den die Schüler bearbeitet haben. Bevor die einzelnen Gruppen die Aufgabenkarte und das Chemikalientablett erhalten, sollten noch ein paar Sätze zur Gruppenarbeit gesagt werden:

Lehrervortrag:
Organisatorisches zur Gruppenarbeit – Aufgaben sollten stündlich wechseln:

Organisator:	Holt Geräte
Sicherheitsexperte:	Holt Schutzbrillen
Entsorger:	Entsorgt Chemikalien am Stundenende
Protokollführer:	Schreibt Diskussions- Ergebnisprotokoll
bei Dreiergruppen:	Sicherheitsexperte übernimmt Entsorgung

Austeilen und Besprechen der Protokollbögen
Anschließend erhalten alle Gruppen ein Chemikalientablett, Schutzbrillen und eine Aktionskarte (laminiert, A5). Auf den Tabletts befindet sich:

Limonade	dest. Wasser	NaOH-Lösung
Zitronensaft	Feinwaschmittel	Fensterreiniger
WC-Reiniger	Zuckerwasser	Kalkwasser
Mineralwasser		Rohrreiniger
Speiseessig		

Gruppenarbeit (ca. 15‘):

Bearbeitung der Aufgabe in Gruppenrecherche. (Kalkwasser und NaOH-Lösung sollte den Schülern als Laborchemikalie vorgestellt werden, NaOH kann *umgangssprachlich* auch als Natronlauge angeführt werden.) In der lehrerzentrierten Klasse erfolgt diese Aufgabe in Stillarbeit. Alle 12 Lösungen sind hierzu auch in großen 500mL–Flaschen vorhanden.

Auswertung:
Die gruppenzentriert unterrichtete Klasse präsentiert ihre Vorschläge am OHP, durch das Folienpuzzle können rasch verschiedene Vorschläge realisiert werden.

Die lehrerzentrierte Klasse kann diese Aufgabe an der Tafel anfertigen. Wichtig ist, dass in beiden Klassen ein abschließender Konsens vorhanden ist, der in die Hefte übertragen werden kann.
Anm.: Beliebte Klassifizierungen sind zu diesem Zeitpunkt: Getränke – Reinigungsmittel – Säuren – [Wasser]

Hausaufgaben:
Suche in eurem Haushalt nach weiteren Lösungen und ordne sie in die Tabelle ein.

Wichtig:

- Am Ende der Stunde werden die Hausaufgabenhefter verteilt und in der hausaufgabenintervenierten Klasse die Anweisung zur Hausaufgabenintervention verlesen.
- Die Hausaufgabenstellung befindet sich bereits im Ordner. Die Schüler unbedingt daran erinnern, dass die HA in den Hefter geschrieben werden. (hierfür bitte mind. 5‘ einplanen)
- Zuletzt wird in beiden Klassen ein kurzer Interessentest durchgeführt (3‘)

Zur Erinnerung:
Dauerhaft ist darauf zuachten, dass immer von *sauren bzw. basischen* <u>*Lösungen*</u> gesprochen wird („alkalisch“, „Lauge“ bzw. „Säure“ und „Base“ sollen vermieden werden)

Verlaufsplan

Phase	Inhalt	Methode/ Medien	Zeit
Einstieg	Wiederholung des Lösevorgangs, Bsp. Zuckerlösung	UG LDemV	
Erarbeitung	Gruppenrecherche, Kategorisierung der 12 Lösungen	GA Lösungstablett	
Auswertung	Vergleich und Diskussion der Ergebnisse der Gruppenarbeit im Plenum, Abstimmung auf einen Klassenkonsens; Übernahme in die Hefte	UG Transparent	
Test	Austeilen der HA-Hefter, Besprechung der allg. HA-Instruktion, Hinweis auf die HA-Stellung, Durchführung des Interessentests, währenddessen werden die Gruppenprotokolle eingesammelt!	HA-Hefter Interessentest	

Aufgabenkarte zur ersten Stunde

Chemie - Box I

Auf dem Tablett findet ihr 12 Flaschen mit unterschiedlichen Lösungen. Eure Aufgabe besteht darin, diese Lösungen in Gruppen einzuteilen, so wie es euch sinnvoll erscheint. Gebt dann den Gruppen einen Namen, der die Gruppe kennzeichnet.

Wichtig: **Diskutiert** eure Ideen mit euren Mitschülern und **protokolliert** kurz eure Überlegungen!

Zweite Unterrichtsstunde

Materialien:

großes Becherglas (600mL) zur Entsorgung der Chemikalien

Thema: Unterscheidung saurer und basischer Lösungen mit Hilfe von Indikatoren

Einstieg:
Zum Einstieg sollen die Lebensmitteluntersuchungen der Hausaufgaben dargestellt und die tabellarische Sammlung der letzten Stunde um diese Stoffe ergänzt werden.

Erarbeitung:
Wichtig, vor der Bearbeitung der Aktionsboxen bitte zur Erinnerung eine Regel vorweg stellen: Als ersten Schritt bei der Gruppenarbeit lesen die Organisatoren die Aufgabenstellung laut vor! Für die Gruppen geht es darum, im vorgegeben Zeitrahmen die Aufgabe zu beantworten (sportiver Anreiz)! Austeilen der Boxen.
Dabei ist darauf zu achten, dass die Schüler möglichst schnell die Funktion der „Reagenzglas-Übersichtskarte“ erfassen. (Dazu dient auch der beigelegte Folienstift.)
Während der Gruppenarbeit müssen die Schüler ggfs. darauf hingewiesen werden, die Farbumschläge sofort zu notieren (Rohrreiniger schlägt sonst nach gelb um!)

Auswertung:

TAFEL: Wir untersuchen verschiedene Lösungen

Beobachtung: Versetzt man die zwölf Lösungen mit Rotkohlsaft, so entstehen unterschiedliche Färbungen.

rot *saure Lösungen**	violett-blau	grün-(gelb) *basische Lösungen**
Speiseessig Limonade Zitronensaft (Mineralwasser)	Dest. Wasser Zuckerwasser Feinwaschmittel (Mineralwasser)	Fensterreiniger Natriumhydroxidlsg. Kalkwasser Rohrreiniger

** = nachtragen*

Die Aktionskarte „Rheinischem Rotkohl wird Essig zugegeben“ kann die Schüler veranlassen, die rote Gruppe als saure Lösungen zu bezeichnen. Der Begriff „Base“ als Gegensatzpaar wird vermutlich vom Lehrer vorgegeben werden müssen. *Der Begriff „neutrale Lösungen“ muss an dieser Stelle noch nicht eingeführt werden, es sei denn, er wird durch die Schüler vorgegeben.*

Tafel:

Merksatz:	**Lösungen, die Rotkohlsaft rot färben, bezeichnet man als <u>saure Lösungen.</u>** **Lösungen, die Rotkohlsaft grün (oder gelb) färben, bezeichnet man als <u>basische Lösungen</u>.**

HA-Stellung:
Untersuche in deinem Haushalt Lösungen mit schwarzem Tee (AB) HA-Zettel austeilen und Protokolle einsammeln.

Verlaufsplan

Phase	Inhalt	Methode/ Medien	Zeit
Einstieg Besprechung der HA	Kurze Übernahme der untersuchten Haushaltslösungen in die Tabelle –	UG	
Erarbeitung Interaktionsboxen	Schüler untersuchen die 12 Lösungen mit Hilfe von Rotkohlsaft; gruppieren die Lösungen nach Färbung	Boxen + Becherglas (600mL)	
Auswertung Funktion von Rotkohlsaft	Auswertung der Ergebnisse: Rotkohlsaft färbt die Lösungen; Es lassen sich drei Gruppen identifizieren: rot, violett-blau und grün-gelb entsprechend der Zuordnung; sauer, (neutral), basisch. **Einsammeln der Protokolle und Austeilen der Hausaufgabenzettel**	Tafel	
Test	Schüler füllen Interessentest aus, währenddessen sammelt der Lehrer die Protokolle der Gruppenarbeit ein	Interessentest	

Aufgabenkarte und instruktionale Hilfe zur zweiten Stunde

Chemie - Box II

Eure Aufgabe für die heutige Stunde lautet: Überprüft mit Hilfe der zusätzlichen Lösung, der Box und der Hinweiskarten die Gruppeneinteilung der zwölf Lösungen aus der letzten Stunde. Welche weitere Gruppierungsmöglichkeit schlagt ihr vor?

Wichtig: Diskutiert eure Ideen mit euren Mitschülern und protokolliert kurz eure Überlegungen!

Zusatzinfo:

Das Gemüse „Rotkohl" trägt nur in Norddeutschland diesen Namen. In Süddeutschland kennt man es unter dem Namen „Blaukraut".

Während die süddeutsche Variante (*Blau*kraut) nur mit Wasser gekocht wird, gibt man bei der norddeutschen Art (*Rot*kohl) Essigsäure hinzu....

Dritte Unterrichtsstunde

Materialien:

Zahl der Lösungen auf sechs reduzieren (2*6)

KEINEN ROTKOHLSAFT reingegeben

großes Becherglas zur Entsorgung

große RG + Ständer

Schutzbrillen

Gr. A (4x)	Gr. B(3x)
WC-Reiniger	Speiseessig
Zitronensaft	Limonade
Zuckerwasser	destilliertes Wasser
Mineralwasser	Feinwaschmittel
Glasreiniger	Natriumhydroxid-Lösung
Kalkwasser	-----

Thema: Einführung des pH-Wertes zur Charakterisierung von sauren und basischen Lösungen

Einstieg:
Besprechung der Hausaufgaben. Einige Schüler haben vermutlich vergeblich auf eine ähnlich spektakuläre Farbreaktion wie die des Rotkohlsaftes gewartet. Gefordert wird bei dieser Hausaufgabe hingegen:

1. einen planvollen Einsatz der zu untersuchenden Lösungen,
2. ein sorgfältiges Beobachten der Reaktionen, sowie
3. eine Interpretation der Ergebnisse auf der Basis der Lernziele der zweiten Stunde.

Planvoll bedeutet, dass die Schüler klare Lösungen einsetzen sollten, vorzugsweise solche, bei denen eine begründete Vermutung hinsichtlich einer Gruppierung nach sauer/basisch besteht.
Sorgfältiges Beobachten bedeutet, dass ein genereller Trend zu erkennen ist, wenn man saure Lösungen zu schwarzem Tee gibt (Aufhellung) im Gegensatz zu basischen Lösungen (keine Aufhellung).
Kann kurz im LDemV veranschaulicht werden:
Etwas schwarzer Tee in vier Demo-Reagenzgläser vorlegen, zwei werden mit sauren Lösungen (WC-Reiniger u. Zitronensaft), zwei werden mit basischen Lösungen (Fenster- und Rohrreiniger) versetzt.→ Aufhellung bei sauren Lösungen, Verdunkelung bei basischen Lösungen. Daher kann die Aufhellung im Falle der sauren Lösungen nicht mit einem Verdünnungseffekt erklärt werden.
Abstraktion der Ergebnisse zum Rotkohlsaft bzw. schwarzen Tee („Welche Eigenschaften/Funktion haben diese beiden Lösungen?“) führt zu folgendem Merksatz:

Tafel:

Stoffe, wie schwarzer Tee oder Rotkohlsaft, mit denen man saure von basischen Lösungen unterscheiden kann, bezeichnet man als **Indikatoren.** (von lat. indicare = anzeigen)

Erarbeitung:
Austeilen der Aktionsboxen: Die Schüler sollen zum einen die pH-Skala zur Abstufung der sauren und basischen Lösungen entdecken und zum anderen als generelle Eigenschaften von sauren und basischen Lösungen angeben, dass der pH-Wert saurer Lösungen kleiner als sieben, der von basischen Lösungen größer als sieben ist. Aus Gründen der Zeitökonomie wird arbeitsteilig gearbeitet: vier Gruppen bekommen die ersten sechs Lösungen, die anderen drei Gruppen untersuchen die übrigen sechs Lösungen.

Auswertung:
Das Ergebnis ist im Tafelbild fest zu halten:
Der pH-Wert ausgesuchter Lösungen beträgt: (Folientransparent)

Lösung	**pH-Wert**	**Lösung**	**pH-Wert**
WC-Reiniger	1	dest.Wasser	7
Speiseessig	2	Feinwaschmittel	7-8
Zitronensaft	2	Fensterreiniger	9
Limonade	3	Kalkwasser	11
Mineralwaser	6	Natriumhydroxidlösung	12
Zuckerwasser	7	Rohrreiniger	13

Tafel:

Der pH-Wert gibt das Ausmaß des sauren bzw. basischen Charakters einer Lösung an. Je kleiner der pH-Wert desto saurer, je größer der pH-Wert desto basischer ist die Lösung. Lösungen mit einem pH-Wert von 7 bezeichnet man als neutrale Lösungen.

Warum ist Mineralwasser saurer als dest. Wasser/ oder Zuckerwasser? ➜ wg. der Kohlensäure (Mineralwasser kann (sollte) daher bei den sauren Lösungen eingruppiert werden.) Darüber hinaus ist im Unterrichtsgespräch festzuhalten, dass sowohl saure als auch basische Lösungen ätzende Eigenschaften haben – in Abhängigkeit vom pH-Wert kann eine basische Lösung daher stärker ätzend sein als eine saure (siehe Rohrreiniger– Information hierzu befindet sich auf Karte in den Boxen).

HA/ Test: HA-Stellung: praktische Untersuchung des pH-Werts von Handseife, Diskussion des Begriffs „pH-neutral“. Austeilen der HA-Zettel + Fragebogen, Einsammeln der Protokolle Bearbeitung des Stundentests.

Verlaufsplan

Phase	Inhalt	Methode/ Medien	Zeit
Einstieg	Besprechung der HA, ggf. LDemV Merksatz „Indikatoren“	LDemV Tafel	
Erarbeitung	Bearbeitung der Aktionsboxen	Boxen	
Auswertung	Definition pH-Wert, Merksatz, generelle Aussagen zu sauren / basische Lösungen.	Tafel Transparent	
HA /Test	Schüler füllen Interessentest aus, währenddessen sammelt der Lehrer die Protokolle der Gruppenarbeit ein.	HA-Blätter + Fragebogen Interessentest	

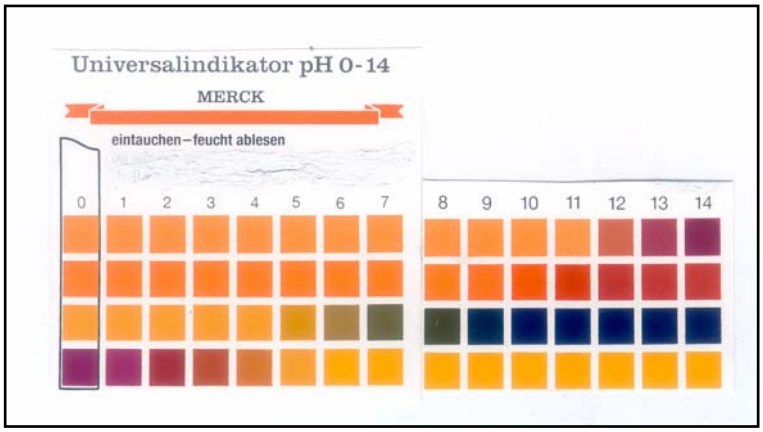
Aufgabenkarte und instruktionale Hilfen der dritten Stunde

Chemie - Box III

Für die heutige Stunde erhaltet ihr zwei Aufgaben: Zunächst sollt ihr mit Hilfe der Box herausfinden, mit welchem System sich die sauren und basischen Lösungen untereinander sinnvoll abstufen lassen.

Die zweite Aufgabe lautet: Mit welchen allgemeinen Eigenschaften lassen sich saure und basische Lösungen beschreiben?

Wichtig: **Diskutiert** eure Ideen mit euren Mitschülern und **protokolliert** kurz eure Überlegungen!



pH-Wert:

„Maß für den sauren oder basischen Charakter einer Lösung."

aus:
Meyers Taschenlexikon

Vierte Unterrichtsstunde

Aus den Chemikalientabletts entfernen: Essig, Zitronensaft, Mineralwasser, Feinwaschmittel Zusätzlich 8x Salzsäure und Rotkohlsaft; Becherglas zum Entsorgen der Lösungen

Thema: **Neutralisation einer sauren Lösung mit einer basischen Lösung**

Einstieg:
Besprechung der Hausaufgaben. Mündlicher Vergleich der pH-Werte. Die pH-Werte nicht hautneutraler Seifen dürften im pH-Bereich um sieben liegen. Für das Unterrichtsgespräch resultiert hieraus zweierlei (integrierte Wiederholung!):1. der normale pH-Wert der Haut ist offensichtlich leicht sauer (5,6) 2. hautneutral und ph-neutral ist nicht das gleiche. Die Verwendung „normaler“ Seifen kann daher zur Zerstörung des normalen pH-Werts der Haut führen. Der pH-Wert einer pH-hautneutralen Lotion kann ggfs. im LDemV getestet werden.

Folgeüberlegung: Warum bezeichnet man den sauren pH-Wert der Haut auch als “Säureschutzmantel“? (⇒ Schutz vor Infektionen / Krankheitskeime werden abgetötet) Daraus sollten im U’gespräch kurz folgende, allgemeine Eigenschaften von sauren /basischen Lösungen abgeleitet werden:

- Saure Lösungen können auch durchaus nützlich sein (Desinfektion, Konservierung von Lebensmitteln).
- Ätzend wirken sowohl saure Lösungen mit sehr geringem als auch basische Lösungen mit sehr hohem pH-Wert (Transparent „Rohrreiniger“ 3. Stunde).
- Warum lässt sich auch Mineralwasser als saure Lösung bezeichnen (auf Grund der Kohlensäure)?

Erarbeitung:
Mit dem Hinweis “In der heutigen Box findet ihr eine weitere saure Lösung, die den meisten von euch vom Namen bekannst sein dürfte?“ kann die Erarbeitungsphase eingeleitet werden. Sollten die Gruppen nicht zunächst auf die Idee kommen den pH-Wert von Salzsäure zu bestimmen, so muss diese Hilfe gegeben werden.

Auswertung:
Diskussion der Vorgehensweise im Unterrichtsgespräch (Transparent pH-Skala)
Tafelbild:

Die Neutralisationsreaktion **(Nachtragen)**

Die saure Eigenschaft der Salzsäure lässt sich durch Zugabe einer basischen Lösung beheben. Am besten eignet sich hierzu Natriumhydroxidlösung, weil sie ebenso basisch ist, wie Salzsäure sauer ist:

pH-Wert der Salzsäure: 2 (5 pH-Einheiten von neutral entfernt)
pH-Wert von Natriumhydroxidlösung: 12 (5 pH-Einheiten von neutral entfernt)

Das Aufheben der sauren Eigenschaften einer Lösung mit einer basischen Lösung (oder umgekehrt) bezeichnet man als Neutralisation.

Sicherungsfrage (sofern noch nicht im Unterrichtsgespräch gefallen): Wie groß ist der pH-Wert nach einer Neutralisationsreaktion?

HA / Test:

Austeilen der Hausaufgaben: Wirkungsweise eines Antacidums / Interessentest

Verlaufsplan

Phase	Inhalt	Methode/ Medien	Zeit
Einstieg	Besprechung der Hausaufgaben/ LDemV : pH-Wert von hautneutraler Lotion Ätzende Eigenschaft von sauren und basischen Lösungen ...	UG / LDemV	
Erarbeitung	Neutralisation von 0,01M HCl. Mit vorzugsweise Natronlauge.	GA Boxen	
Auswertung	Kurze Darstellung des Lösungswegs, Übernahme der Definition.	UG Tafel Transparent	
Test	Schüler füllen Interessentest aus, währenddessen sammelt der Lehrer die Protokolle der Gruppenarbeit ein	Interessentest	

Aufgabenkarte und instruktionale Hilfen der vierten Stunde

Chemie - Box IV

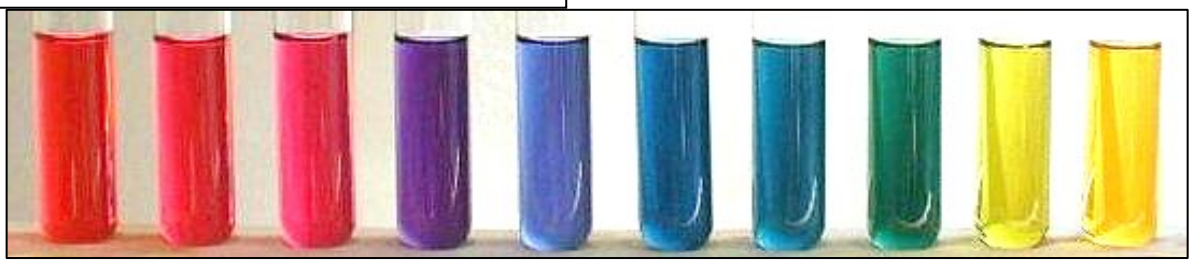
In der heutigen Box findet ihr eine neue Lösung mit der Aufschrift „Salzsäure". Eure Aufgabe besteht darin, mit Hilfe der Box die Salzsäure in ihrer sauren Wirkung unschädlich zu machen. Füllt hierzu ein Reagenzglas zu ca. einem Drittel mit Salzsäure. Beachtet jedoch, dass die entstehende Lösung ebenfalls unschädlich und nicht ätzend sein darf!

Wichtig: **Diskutiert** eure Ideen mit euren Mitschülern und **protokolliert** kurz eure Überlegungen!

pH-Wert-Übersicht:

Salzsäure	pH=___?	Glasreiniger	pH= 9
WC-Reiniger	pH= 1	Kalkwasser	pH=11
Limonade	pH= 3	Natriumhydroxid-Lösung	pH=12
dest. Wasser	pH= 7		
Zuckerwasser	pH= 7	Rohrreiniger	pH=13

Farbskala von Rotkohlsaft



pH 1 2 4 6 7 8 9 10 11 13

Die pH-Skala

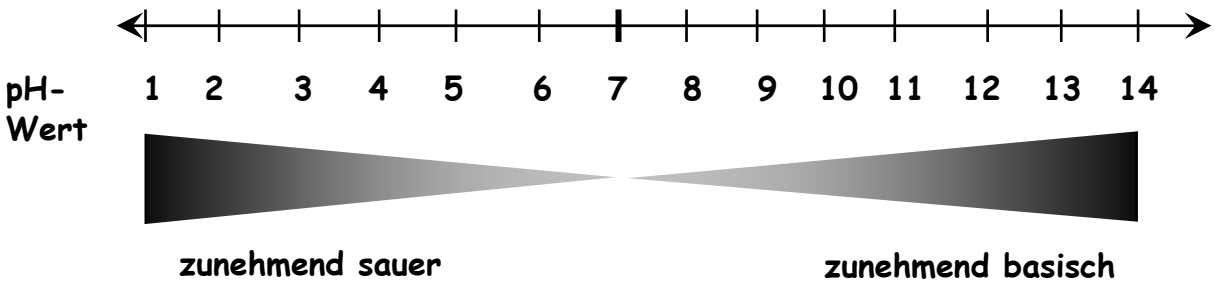


Foto der Interaktionsbox zur vierten Stunde

Inhalt der Interaktionsbox der vierten Stunde:

- 1 Reagenzglasständer
- 8-10 Reagenzgläser
- 5 Stopfen
- 3 kleine Bechergläser
- 1 Folienstift
- 1 Spritzflasche mit Indikatorlösung
- 1 Aufgabenkarte
- Informationskarten
- 7 pH-Teststäbchen
- 2 Pasteurpipetten.

Zusätzlich liegt den Materialien der vierten Stunde ein Spritzflaschentablett mit folgenden neun Lösungen bei:
Salzsäure, WC-Reiniger, Limonade, destilliertes Wasser, Zuckerwasser, Glasreiniger, Kalkwasser, Natriumhydroxidlösung und Rohrreinigerlösung

Fünfte Unterrichtsstunde

Thema: Integrierte Wiederholung – Reaktion von Nichtmetalloxiden mit Wasser

Einstieg:

Ergebnissicherung der dritten und vierte Stunde: Bearbeiten des Puzzles in den Kleingruppen: Lösungssätze:

Saure Lösungen haben eine pH-Wert, der kleiner ist als sieben. Der pH-Wert basischer Lösungen ist größer als sieben. Um eine saure Lösung zu neutralisieren benötigt man eine basische Lösung. Beträgt der pH-Wert einer basischen Lösung 12, so setzt man zur Neutralisation eine saure Lösung ein, deren pH-Wert 2 beträgt, weil der pH-Wert 12 genau so viele pH-Einheiten (nämlich 5) vom Neutralpunkt entfernt ist wie der pH-Wert 2 (ebenfalls 5).

Schüler lesen ihr Ergebnis im Plenum vor.
Besprechung der Hausaufgaben:
Die Funktion der Magensäure ist in der Abtötung von Keimen zu sehen (u. ggf. bei der Nahrungsverdauung)
Zur Wirkungsweise eines Antacidums sollte von den Schülern gesagt werden, dass es in jedem Fall basische Eigenschaften habe sollte, damit die Magensäure neutralisiert werden kann.
Hier kann weiter überlegt werden, ob es das Ziel eines Antacidums sein kann, denn gesamten Magenbrei auf pH 7 zu neutralisieren (natürlich nicht). Zweitens stellt sich die Frage, wie basisch eine Antacium sein darf. (Meist um pH 9, da höhere pH-Werte ebenfalls ätzend wirken würden.)

Erarbeitung:

Bearbeitung des Textes „Waldschäden durch sauren Regen“ in den Kleingruppen (inklusive Protokoll)

Auswertung:

1. Die Auswirkungen des sauren Regens sind dem Text zu entnehmen und werden im Plenumgespräch zusammengetragen.
2. Saurer Regen entsteht, wenn sich Schwefel-, Stickstoff- und Kohlenstoffoxide im Regenwasser lösen.
3. Alle erwähnten Gase sind Nichtmetalloxide.

Tafelbild

<u>Entstehung des sauren Regens</u>

Saurer Regen entsteht, wenn sich Abgase wie Schwefeloxide, Stickstoffoxide im Regenwasser lösen.
Allgemein gilt: Beim Lösen von Nichtmetalloxiden in Wasser entstehen immer saure Lösungen.

HA / Test:
Wdh. Entstehung des sauren Regens/ Gegenmaßnahmen auf der Verhaltensebene

Verlaufsplan

Phase	Inhalt	Methode/ Medien	Zeit
Einstieg	Sicherungspuzzle Besprechung der HA	GA UG	
Erarbeitung	Bearbeiten des Textes „Waldschäden durch sauren Regen“	GA	
Auswertung	Lösung von Nichtmetalloxiden in Wasser ergibt saure Lösung	UG Tafel	
Test	Schüler füllen Interessentest aus, währenddessen sammelt der Lehrer die Protokolle der Gruppenarbeit ein	Interessentest	

Aufgabenkarte und instruktionale Hilfen der fünften Stunde

Waldschäden durch sauren Regen

Unsere Wälder sind krank. Besonders betroffen sind die Nadelbäume; aber auch an den Laubbäumen werden zunehmend mehr Schäden beobachtet.

Wissenschaftler vermuten, dass eine der Hauptursachen für die Waldschäden die Luftverschmutzung ist. Kraftwerke, Industrieanlagen und Kraftfahrzeuge blasen jährlich Millionen Tonnen schädlicher Abgase in die Luft.

Der Ausstoß von Schwefeldioxid ist der Hauptverursacher des „sauren Regens". Hinzu kommen weitere Gase wie Stickstoffoxide und Kohlenstoffoxide. Beim Lösen dieser Gase im Regenwasser bilden sich saure Lösungen. Diese sind für die Ausbildung des sauren Regens verantwortlich.

Die Folge des „sauren Regens" ist eine Schädigung der Blätter und Wurzeln der Bäume und eine Herabsetzung ihrer natürlichen Widerstandskraft gegenüber tierischen und pflanzlichen Schädlingen. Außerdem sind auch Seen und Gewässer von den Auswirkungen des „sauren Regens" betroffen.

Aufgaben:

1. **Stellt die Auswirkungen des sauren Regens auf die Umwelt zusammen.**
2. **Erklärt, wie der saure Regen entsteht. Versucht eine Gemeinsamkeit aller am sauren Regen beteiligten Gase zu finden**

Saure Lösungen haben einen pH-Wert, ...

...der kleiner ist als sieben.

Der pH-Wert basischer Lösungen ...

...ist größer als sieben.

Um eine saure Lösung zu neutralisieren

..benötigt man eine basische Lösung.

Beträgt der pH-Wert einer basischen Lösung 12,

...so setzt man zur Neutralisation ...

... eine saure Lösung ein, deren pH-Wert ...

... 2 beträgt, weil der pH-Wert 12 ...

...genau so viele pH-Einheiten (nämlich 5)...

...vom Neutralpunkt entfernt ist,...

...wie der pH-Wert 2 (ebenfalls 5).

Sechste Unterrichtsstunde

Materialien:

Schwefel, Standzylinder mit Glasplatte Verbrennungslöffel, Spatel, Trichter, Indikatorpapier, 7 X 50 mL PE-Flaschen, 7 x BTB und 7 X 100 mL PE-Wasserflaschen 2 große Bechergläser zum Entsorgen

Thema: Reaktion von Metalloxiden in wässriger Lösung

Einstieg:
LDemV: Verbrennen von Schwefel im Standzylinder mit Wasser (150 – 200 mL). Zuvor ist der pH-Wert des Wassers zu messen. Die Verbrennung ist im abgedunkelten Abzug durchführen (ca. 3min). Währenddessen Wdh. „Lösung von Nichtmetalloxiden in Wasser“ im Schülervortrag. Zum Schluss den Standzylinder schütteln und pH-Wert messen. (Sollte um drei bis vier liegen, was einen realistischen Wert für sauren Regen darstellt.)
Durch den Schülervortrag zur Theorie des sauren Regens sollte Aufgabe eins der Hausaufgaben abgedeckt sein.
Aufgabe zwei der Hausaufgaben folgt im Unterrichtsgespräch. Dabei sollte deutlich werden, dass der Schadstoffausstoß nicht allein auf Industrieabgase zurückzuführen ist, sondern auch durch Privathaushalte verursacht wird: Kaminbrand, offenes Feuer (Gartenlaub), PKW-Verkehr, d. h. generell bei allen Verbrennungsreaktionen.

Erarbeitung:
Der hergestellte saure Regen wird auf die vorbereiteten 50mL PE-Fläschen verteilt (ca. 20mL pro Flasche, Trichter liegt bereit). Zusätzlich zu den Aktionsboxen wird nur Bromthymolblau und Wasser (PE 100mL, dient zum Lösen) ausgeteilt.
In GA soll der saure Regen neutralisiert werden.
Die Schüler sollten zum einen aus der vierten Stunde gelernt haben, dass der Indikator <u>frühzeitig</u> als pH-Kontrolle eingesetzt wird.
Problematisch könnten evtl. die Lösungsversuche sein: Während sich Zucker und Salz gut in Wasser lösen (pH 7) sind die Oxide nur schwerlöslich. Den Schülern muss ggf. der Hinweis gegeben werden, den pH-Wert von wässriger CaO/ MgO-Lösung zu testen (13 bzw. 10-11). Die schlechte Löslichkeit der Verbindungen spielt somit keine Rolle zum Lösen der Aufgabe.

Auswertung:
Die Schüler stellen im Plenum ihre Arbeitsergebnisse vor. Dabei sollte zweierlei deutlich werden:

1. Die exakte Neutralisation ist eine höchst kniffelige Angelegenheit, beträgt der pH-Wert des sauren Regens 3-4, so ist MgO-Lsg. das Mittel der Wahl (pH 10 –11, CaO geht aber natürlich auch).
2. Allgemein sollte herausgefunden werden, dass wässrige Lösungen von Metalloxiden basisch reagieren. (Die pauschale Antwort „alle Oxide können verwendet werden“ gilt nicht, da in der letzten Stunde auch saure Lösungen aus Oxiden hergestellt wurden.

Tafelbild

Wir neutralisieren sauren Regen

Die Lösungen von Magnesiumoxid und Calciumoxid reagieren basisch. Mit diesen Stoffen lässt sich eine Lösung, die den pH-Wert von saurem Regen besitzt, neutralisieren.

Allgemein gilt: Beim Lösen von Metalloxiden in Wasser entstehen immer basische Lösungen.

HA / Test:

Austeilen der HA-Blätter: Wdh: wässrige Lösungen von Nichtmetall- Metalloxiden. Einsatz von gebranntem Kalk bei der Bodenversauerung durch sauren Regen.
Durchführung des Interessentests

Verlaufsplan

Phase	Inhalt	Methode/ Medien	Zeit
Einstieg	LDemV: Verbrennung von Schwefel, Lösung von SO_2 in Wasser. Besprechung der HA	LDemV UG	
Erarbeitung	Aktionsboxen: Neutralisation von saurem Regen mit Feststoffen CaO, MgO, Salz Zucker	GA Aktionsboxen	
Auswertung	Abstraktion: Wässrige Lösungen von Metalloxiden reagieren basisch.	UG Tafel	
Test	Schüler füllen Interessentest aus, währenddessen sammelt der Lehrer die Protokolle der Gruppenarbeit ein	Interessentest	

Aufgabenkarte und instruktionale Hilfen der sechsten Stunde

Chemie - Box V

In der heutigen Stunde geht es wieder einmal um die Neutralisation. Heute sollt ihr versuchen den sauren Regen mit Hilfe der Box zu neutralisieren. Allerdings findet ihr dieses mal nur Feststoffe vor und die Anzahl der pH-Stäbchen ist begrenzt.

1. Versucht eure Probe des sauren Regens so gut wie möglich zu neutralisieren.
2. Welche Gemeinsamkeiten besitzen alle Stoffe, die sich zur Neutralisation eignen?

Wichtig: **Diskutiert** eure Ideen mit euren Mitschülern und **protokolliert** kurz eure Überlegungen!

Calciumoxid: **Hierbei handelt es sich um ein Metalloxid, welches das Metall Calcium mit Sauerstoff bildet.**

Zucker: **Zucker ist eine pflanzliche Verbindung, die aus den Elementen Kohlenstoff, Wasserstoff und Sauerstoff besteht.**

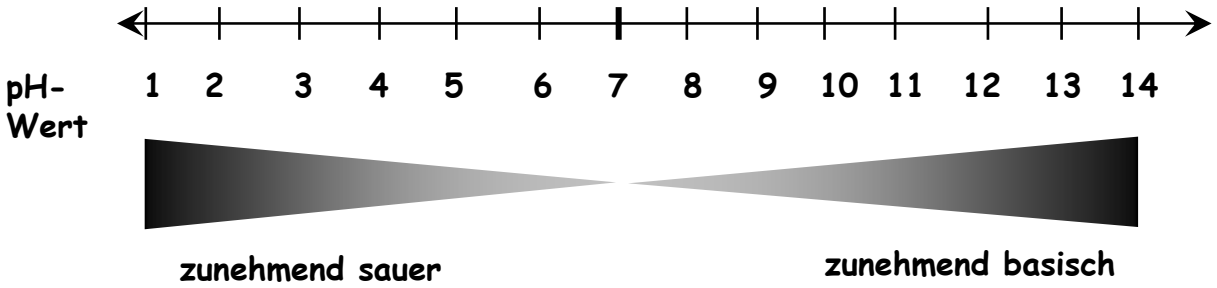
Kochsalz: **Der chemische Name dieser Verbindung lautet Natriumchlorid. Dieses Salz lässt sich aus den Elementen Natrium und Chlor herstellen.**

Magnesiumoxid: **Diese Verbindung gehört zu der Gruppe der Metalloxide. Magnesiumoxid entsteht bei der Reaktion des Metalls Magnesium mit Sauerstoff.**

Bromthymolblau ist ebenso wie Rotkohlsaft und schwarzer Tee ein **Indikator** mit typischen Farbveränderungen.

Saure Lösungen färbt er **gelb**, neutrale **grün** und basische **blau**.

Die pH-Skala



Kalkvorkommen in der Natur und Technik

Kalk (Calciumcarbonat) ist als Gestein in der Natur weit verbreitet. Es handelt sich meist um *Meeresablagerungen*. Bei ihrer Bildung waren viele Arten von Organismen beteiligt, wie Muscheln, Korallen, Schnecken, Seelilien u.a. Deren Strukturen (Versteinerungen) sind in vielen Kalkgesteinen noch erkennbar. Die Kristalle des *Marmors* haben sich nachträglich aus solchen Ablagerungen in der Tiefe der Erdrinde bei hohen Temperaturen und hohem Druck gebildet.
Kalkstein findet Verwendung als Baustein und zur Herstellung von Zement, Brannt- und Löschkalk.
In der Technik wird in Steinbrüchen abgebautes Calciumcarbonat auf ca. 1000°C erhitzt. Dabei zerfällt es in Kohlenstoffdioxid und Calciumoxid:

Calciumcarbonat reagiert zu Calciumoxid und Kohlenstoffdioxid

Dieser Vorgang heißt **Kalkbrennen.** Calciumoxid wird auch als **gebrannter Kalk** oder **Branntkalk** bezeichnet. Gebrannter Kalk reagiert mit Wasser unter Wärmeentwicklung zu **gelöschtem Kalk** oder **Löschkalk**.

Aus: Klett Chemiebuch „elemente"

C. Leistungstests

Offene Fragen

1. Teile folgende Flüssigkeiten in Gruppen ein, so wie sie dir sinnvoll erscheinen. Gib dann den einzelnen Gruppen einen Namen, der die Gruppe kennzeichnet.
   Mineralwasser, Limonade, Speiseessig, Kalkwasser, Fensterreiniger, Natriumhydroxidlösung, destilliertes Wasser, Feinwaschmittel, Zitronensaft, Rohrreinigerlösung, Zuckerwasser

2. In Autobatterien ist eine gefährliche Flüssigkeit, die so genannte „Batteriesäure“ enthalten. Welche der folgenden Flüssigkeiten würdest du benutzen, um solch eine Batteriesäure ungefährlich zu entsorgen? Wasser, Natriumhydroxidlösung (Natronlauge) oder Zuckerwasser. Begründe deine Antwort:

3. Der Magensaft des Menschen besitzt einen pH-Wert von 2 – 3. Der pH-Wert des Stuhls (Kot) beträgt jedoch 7 bis 8. Beschreibe so genau wie möglich die Lösungen in Magen und Darm und die Vorgänge, die im menschlichen Dünndarm ablaufen!

4. Welche der aufgeführten Stoffe sind deiner Meinung nach für den „sauren Regen“ verantwortlich? Calciumoxid (ein Metalloxid), Schwefeldioxid (Nichtmetalloxid), Stickstoffoxid (Nichtmetalloxid), Magnesiumoxid (Metalloxid). Begründe deine Antwort:

5. Du bekommst die Aufgabe, mit einem Experiment herauszufinden, ob es sich bei einem dir unbekannten Pulver um ein Metall oder um ein Nichtmetall handelt. Als Materialien erhältst du: Einen Bunsenbrenner, einen Verbrennungslöffel, ein Becherglas, einen Standzylinder mit Glasplatte, Wasser und ein pH-Teststäbchen.

   Anmerkung: Erhitzt man das Pulver an der Luft, so kommt es zu einer heftigen Reaktion.

   Beschreibe ein Experiment, mit dem sich diese Aufgabe lösen lässt.

Multiple-Choice

In diesem Fragebogen findest du einige Aussagen zur Chemie. Wir möchten dich bitten, jede Aussage anzukreuzen, die deiner Meinung nach richtig ist. Pro Aufgabe können eine, zwei oder auch mehrere Antworten richtig sein.

Folgende Stoffe sind Beispiele für saure Lösungen:

- ❒ Coca Cola
- ❒ Speiseessig
- ❒ Sprudelwasser (Mineralwasser)
- ❒ Nichts von alledem.

Kreuze an, welche Aussagen über Lösungen zutreffen.

- ❒ Basen sind ungefährlicher als Säuren.
- ❒ Säuren sind gefährlicher als Basen.
- ❒ Die Gefährlichkeit einer Lösung hängt vom pH-Wert ab.
- ❒ Da in Lösungen immer Stoffe enthalten sind, reagieren sie immer entweder sauer oder basisch.

Folgende Aussage(n) über saure Lösungen ist (sind) richtig.

- ❒ Sprudelwasser ist eine saure Lösung.
- ❒ Alle sauren Lösungen sind gefährlich.
- ❒ Salzsäure ätzt stärker als Essigsäure.
- ❒ Säuren können zum Konservieren und Zersetzen von Lebensmitteln benutzt werden.

Kreuze an, welche Aussagen über Säuren zutreffen.

- ❒ Der pH-Wert von Salzsäure beträgt 1 bis 3.
- ❒ Je größer der pH-Wert, desto saurer ist die Lösung.
- ❒ Je kleiner der pH-Wert, desto saurer ist die Lösung.
- ❒ Säuren enthalten Sauerstoff.

Der pH-Wert ...

- ❒ gibt an, ob eine Lösung sauer ist.
- ❒ ist in Seife enthalten.
- ❒ gibt an, ob eine Lösung basisch ist.
- ❒ entsteht im Mund beim Kauen.

Der pH-Wert einer basischen Lösung ...

- ❒ ist kleiner als der einer sauren Lösung.
- ❒ ist größer als der einer sauren Lösung.
- ❒ ist kleiner als der einer neutralen Lösung.
- ❒ ist größer als der einer neutralen Lösung.

Je saurer eine Lösung ist, desto ...

- ❒ kleiner ist der pH-Wert.
- ❒ größer ist der pH-Wert.
- ❒ weniger sind ihre basischen Eigenschaften.
- ❒ größer sind auch ihre basischen Eigenschaften.

Bei einer Neutralisation...

- ❒ werden nur die basischen Eigenschaften einer Lösung aufgehoben.
- ❒ werden nur die sauren Eigenschaften einer Lösung aufgehoben.
- ❒ werden weder die sauren noch die basischen Eigenschaften einer Lösung aufgehoben.
- ❒ werden die sauren und die basischen Eigenschaften einer Lösung aufgehoben.

Um eine saure Lösung zu neutralisieren benötigt man...

- ❒ Wasser.
- ❒ eine süße Lösung, z. B. Zuckerwasser.
- ❒ eine basische Lösung.
- ❒ eine neutrale Lösung.

Welche der folgenden Aussagen trifft zu?

- ❒ Indikatoren färben sich in Anwesenheit von Säuren rot.
- ❒ Ein pH-Papier ist ein Beispiel für einen Universalindikator.
- ❒ Säuren und Basen kann man mit Hilfe von Indikatoren unterscheiden.
- ❒ Indikatoren färben Säuren rot.

Löst man Metalloxide in Wasser, so ...

- ❒ entsteht eine saure Lösung.
- ❒ entsteht eine basische Lösung.
- ❒ passiert nichts, da man Metalloxide nicht in Wasser lösen kann.
- ❒ ist der pH-Wert der Lösung größer als 7.

Löst man Nichtmetalloxide in Wasser, so ...

- ❒ ist der pH-Wert der Lösung größer als 7.
- ❒ entsteht eine saure Lösung.
- ❒ entsteht eine Lösung, die Rotkohlsaft hellrot färbt.
- ❒ ist der pH-Wert der Lösung kleiner als 7.

Der pH-Wert einer sauren Lösung ...

- ❒ ist kleiner als der einer basischen Lösung.
- ❒ ist größer als der einer basischen Lösung.
- ❒ ist kleiner als der einer neutralen Lösung.
- ❒ ist größer als der einer neutralen Lösung.

Der pH-Wert ist definiert als ...

- ❒ eine Stoffeigenschaft.
- ❒ ein Maß für den sauren Charakter einer Lösung.
- ❒ die Stärke einer Säure.
- ❒ ein Maß für den basischen Charakter einer Lösung.

D. Videomanuale

1. Manual zum Kalibrieren der Videos

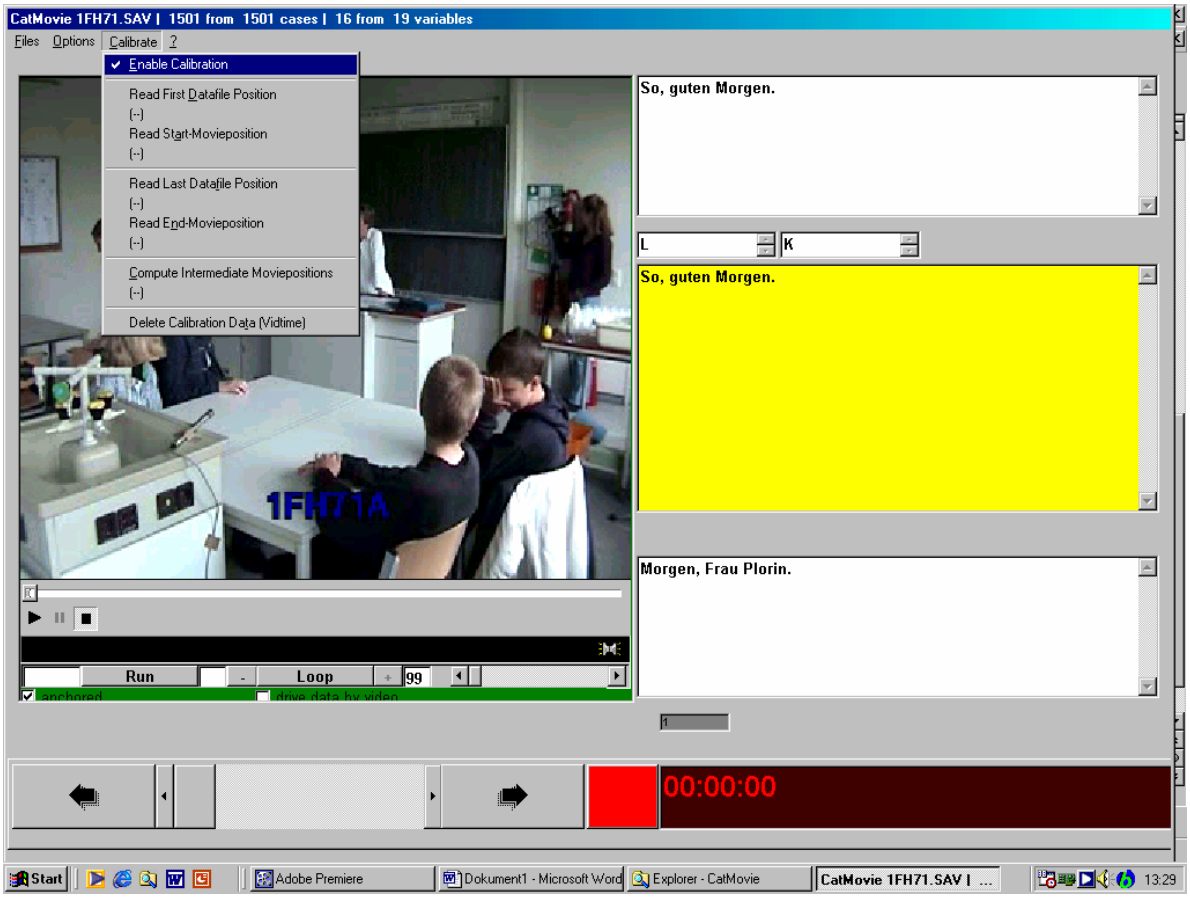
Die Vorbereitung

1. Neben dem Programm „CatMovie“ benötigt man die CDs der Unterrichts-/ Hausaufgabenvideos der A-Kamera, die CatMovie Job-Dateien, sowie die sav-Datei der entsprechenden Stunde/ Hausaufgabe.

2. Die CatMovie Job-Dateien und die sav-Dateien können unter dem Programm „Catmovie“ (meist c:\programme\catmovie) abgelegt werden.
   Es empfiehlt sich, auch die jeweils zu bearbeitenden Videos temporär zu speichern, ihr seid dann bei der Bearbeitung einfach schneller. (Beispielweise unter „c:\temp“ oder ihr legt einen Ordner an „c:\videos“) Praktisch ist in jedem Fall ein möglichst kurzer Pfadname, da ihr diesen in der CatMovie -Jobdatei von Hand eintragen müsst → hierzu öffnet ihr in CatMovie mit „files“ / „open“ die zu kalibrierende Datei (bspw. 1FH71A.CAT). Das im Vordergrund erscheinende schwarze Fenster wird direkt geschlossen.
   Dann öffnet ihr mit „files“ / „edit“ die „Job-Datei“. In der 10. oder 11. Zeile (das hängt von der Größe des Fensters ab) steht: **01,01, D:** „+Name der Datei“. Hier wird das „**D**“ gegen den Pfadnamen (z.B. c:\videos\ s.o.) ausgetauscht, unter dem ihr das Video abgelegt habt.
   Dann die Edit-Datei nach Datei/ Speichern verlassen. Ebenso CatMovie mit „files / exit Catmovie“ verlassen und neu starten.

3. Nach Neustart und Öffnen einer Datei muss man neben dem Videobild
   auch das entsprechende Texttranskript in den rechten Fenstern sehen. (Die Bedeutung der drei Fenster klärt sich später, maßgeblich zum Kalibrieren ist das mittlere, gelbe Fenster.

Das Kalibrieren



4. Die Optionen „anchored“ und „drive data by video” abwählen.
5. Aus dem Menü „Calibration“ den Punkt „Enable Calibration“ wählen.
6. Neben dem Datenschieber erscheint ein roter Punkt. Dieser Knopf wird immer dann gedrückt, sobald die im Transkriptionsfenster (gelb unterlegt) sichtbare Äußerung im Video zu hören ist.
7. Anschließend wird mit dem Datenschieber (großer Pfeil nach rechts) die nächste Äußerung ausgewählt und analog (ab Punkt 6) kalibriert.
8. Sollte man ein einer Stelle nicht mitgekommen sein, das Video stoppen, die Optionen „drive data by video“ und „anchored“ wählen und mit dem Video- oder Datenschieber zur letzten korrekt kalibrierten Äußerung zurückgehen, beide Optionen wieder abwählen und ab Punkte 6 weiterarbeiten.
9. Wenn das Video vollständig bearbeitet ist, wird „Enable Calibration“ wieder deaktiviert und die SPSS-Datei endgültig gespeichert. (Zwischenspeichern nicht vergessen!)
10. Die SPSS-Datei wird anschließend auf Lücken überprüft, diese kennzeichnen sich in der Spalte „vidtime“ durch den Parameter „-1“. Der Eintrag ist durch den tatsächlichen Zeitpunkt zu ersetzen und der Eintrag in der Spalte „realtime“ entsprechend zu korrigieren.

11. Die Spalten „vidtime“ und „realtime“ werden daraufhin überprüft, ob die turnweise nacheinander angeordneten Äußerungen auch zeitlich nacheinander kalibriert sind. (Beispiel: Wenn Zeile 15 bei „realtime“ den Eintrag 00:02:15 hat, muss die nächste Zeile den Eintrag 00:02:15 oder größer haben. Gleiches gilt für „vidtime“. Ggf. ist die SPSS-Datei manuell zu korrigieren.
12. Das kalibrierte Video noch einmal vollständig ansehen („anchored“ aus, „drive data by video“ an). Sollten sich an einzelnen Stellen deutliche zeitliche Unterschiede zwischen transkribiertem und gehörtem Text zeigen, die turn-Nummer notieren (graues Feld oberhalb der vidtime), damit die Stelle ggf. in der SPSS-Datei nachbearbeitet werden kann.

2. Videotranskriptions – Manual

Von Josef Exeler – Universität Bielefeld
Basierend auf den Erhebungs- und Auswertungsverfahren des DFG-Projekts "Lehr-Lern-Prozesse im Physikunterricht" - Teil II: Transkriptions-Manual Christoph Müller & Tina Seidel, IPN an der Universität Kiel

1. ALLGEMEINE HINWEISE ZUM TRANSKRIBIEREN

Die Transkription der Videos geschieht in zwei Durchgängen.

DURCHGANG 1

- Kennzeichnung des Senders und des Empfängers per Codenummer
- Erstellen des Transkripttextes

DURCHGANG 2

- Kontrolle des Transkripts (Text und Rechtschreibung)
- Einfügen von "handlungsleitenden" Zusatzinformationen

WÄHREND JEDES/NACH JEDEM DURCHGANG(S)

- Probleme und Vorgehen bei der Problemlösung bei allen Transkriptionsdurchgängen in der ***Transkriptions-Problemliste*** vermerken.
- Alle Arbeiten am Transkript im ***Transkriptions-Logbuch*** notieren. Hier bitte auch technische Störungen (z.B. Tonstörungen) mit Videozeitangabe vermerken.

2. TRANSKRIPTIONSREGELN

Allgemein:

- Es wird in Hochdeutsch transkribiert. Grammatische Fehler werden nicht korrigiert.
- Neben dem Sprecher/Sender (1. Spalte) wird auch der Empfänger (2. Spalte) angegeben. Wird ein Name im Text genannt, so wird auch dort der Sprecherkode eingefügt.
- Jeder Sprecher wird in eine neue Datenzeile geschrieben.
  Da in CatMovie keine Datenzeilen eingefügt werden können, bitte durch leere Klammern '( )' eine Datenzeile reservieren, falls die Möglichkeit besteht aus einer anderen Kameraperspektive noch weitere Informationen zu erhalten.
- Sind zwei Kameraperspektiven vorhanden, ist es zeitsparend nicht nach jedem Sprecherwechsel die Perspektive zu wechseln.
- Transkribiere **Lehrer-Schüler-Interaktionen** (LehrerIn spricht mit einzelnen/einer Gruppe von Schülerinnen und Schülern)
- Transkribiere die **Gespräche** von Schülerinnen und Schülern, **welche gerade im Film zu sehen sind**.

- Transkribiere allenfalls auch Gespräche von Schülerinnen und Schülern, die auf dem Film nicht zu sehen sind, wenn diese **Gespräche als Dialoge erkennbar und verständlich** sind.
  Also: Unzusammenhängende Äußerungen oder Wortfetzen werden nicht transkribiert.

2.1 SPRECHERCODES (gelten auch für den Empfänger)

L	**L**ehrer
S+Code (z.B. S12)	**S**chülerIn (z.B. der/die mit der Codenummer 12)
S+Code + S+Code (z.B. S12 + S27)	Mehrere **S**chülerInnen sprechen a) genau dieselben Wörter b) genau zur selben Zeit; aber nicht die ganze Klasse.
S?	Die Identität des Sprechers ist nicht klar.
Ss?	Die Identität mehrerer gleichzeitig das gleiche sprechender SchülerInnen ist nicht klar.
(S+Code) (S+Code/S+Code)	Man ist sich über die Identität eines Sprechers nicht ganz sicher, man hat aber eine Vermutung.
K	Die ganze **K**lasse spricht im Chor.
G	Die ganze Klein**g**ruppe spricht im Chor.
A	(**A**ndere) Es spricht eine Person, die als nicht dem Klassenzimmer zugehörig identifiziert wird, z.B. Unterbrechungen von außen (Person, die an die Tür klopft, SchülerIn aus anderer Klasse; LehrerkollegIn etc.)
V	**V**ideofilmerIn

2.1.A LEKTIONSBEGINN UND –SCHLUSS

- Das Transkript beginnt mit der ersten öffentlichen Äußerung des Lehrers/der Lehrerin an die ganze Klasse.
- Das Transkript endet mit der letzten Äußerung des Lehrers oder der Lehrerin an die ganze Klasse und die Antwort der SchülerInnen.

2.1.B ÜBERSCHNEIDUNGEN DER SPRECHSEQUENZ

//	Eine Person fällt der anderen ins Wort, z.B. die Lehrperson einem Schüler oder eine Schülerin einem Mitschüler. Der doppelte Schrägstrich kennzeichnet die Stelle, an der das gleichzeitige Sprechen beginnt. Die Sprechsequenz derjenigen Person, welche ins Wort fällt, wird am Anfang mit einem doppelten Schrägstrich gekennzeichnet.
	Beispiel: L Was glaubt ihr, wie man diese Lösung neutralisieren kann? S5 Durch Zugabe einer stark // basischen Lösung. S17 // Mit Wasser.
<	Zwei oder mehrere Sprecher beginnen gleichzeitig oder nacheinander zu sprechen und ihre Redebeiträge überschneiden sich. Die Wortbeiträge laufen parallel nebeneinander und scheinen sich nicht gegenseitig zu stören. Hier wird bei jedem beteiligten Sprecher **am Anfang** der Sprechsequenz ein 'kleiner als Zeichen' gesetzt. Dieses dient als Kennzeichen einer gleichzeitigen Rede.

	Beispiel: L Wie bezeichnet man Lösungen mit PH-Wert kleiner 7? S22 < Säuren. S13 < Laugen.
Wechsel der Arbeitsaktivität	Im Klassenzimmer passieren verschiedene Aktivitäten wie Lehrgespräch, Arbeit in Gruppen, Partnerarbeit, Experimente, SchülerInnenvorträge etc. Wenn ein solcher Wechsel während einer Gesprächssequenz erfolgt, muss ein neuer Sprechercode eingeführt werden.
	Beispiel: L Zum Schluss sollte die Lösung diese Farbe annehmen. L Holt Euch nun die Experimentierboxen ab und achtet auf die Aufgabenstellung in der Box.

2.3 SATZZEICHEN

Punkt, Fragezeichen, Komma, Semikolon, Ausrufezeichen etc. werden dort eingesetzt, wo es sinnvoll erscheint.

2.4 DIAKRITISCHE ZEICHEN UND ANDERE KONVENTIONEN

.. **...** **....**	**Zwei Punkte** mit führendem und nachfolgendem Leerzeichen: zeigen eine Sprechpause an, die kürzer als 3 Sekunden ist. **Drei Punkte** eine Sprechpause zwischen 3 und 5 Sekunden und **vier Punkte** eine Sprechpause größer als 5 Sekunden.
	Beispiel: L Ich gieße jetzt den Rotkohlsaft hinzu ... und ihr seht die Farbe schlägt um.
( )	Leere Klammern: Werden gesetzt, wenn ein Wort, ein Satz oder eine Sprechsequenz nicht verstanden werden kann.
	Beispiel: L Hat noch jemand die Aufgabe nicht verstanden? S2 Ja, können wir ( ).
(Wort)	Wort in Klammern: Beste Vermutung, was gesprochen wird; es kann aber keine Garantie gegeben werden, ob es dem tatsächlichen Wortlaut entspricht.
	Beispiel: L Können wir Rotkohlsaft als Indikator verwenden? S29 Ja, (das geht).
(Wort A/Wort B)	Wenn etwas nicht genau verstanden werden kann und zwei Alternativen möglich sind, sollten beide Möglichkeiten - in Klammern und getrennt durch einen Schrägstrich – aufgelistet werden.
	Beispiel: L Ist eine der Lösungen stärker eingefärbt, S23? S23 Ja, rechts (leicht/vielleicht).
Großbuchstaben für Symbole	Wenn Chemikalien, Lösungen, Reagenzgläser etc. mit Buchstaben bezeichnet werden, sollten diese in Großbuchstaben geschrieben

	werden.
	Beispiel: L Welche Lösung scheint eher eine saure zu sein? A oder B?
B-u-c-h-s-t-a-b-i-e-r-e-n	Wenn ein Sprecher ein Wort buchstabiert, dann wird mit Bindestrichen dazwischen geschrieben.
[Zusatz-informationen]	Zusatzinformationen in eckigen Klammern können angefügt werden, wenn der/die TranskribiererIn diese als wichtig für das Verstehen einer Sprechsequenz erachtet. Zur Anwendung kommt diese Regel beispielsweise bei ironischen oder sarkastischen Bemerkungen. Auch nonverbale Äußerungen des Lehrers/der Lehrerin können auf diese Weise erfasst werden, wenn dies für die Interaktion in der Klasse wichtig erscheint.
{}	Geschwungene Klammern dienen zur Beschreibung von nonverbalen Artikulationen der Klasse: {Gelächter}; {Stöhnen}; {Jauchzen} → bitte in Maßen einsetzen
	Beispiel: L Ich schreibe jetzt die Hausaufgaben an die Tafel. Ss {Stöhnen}

2.5 PARTIKEL (SPRECHGERÄUSCHE)

"Back-channels"	Backchannels weisen auf Zustimmung oder Ablehnung hin. Es ist sinnvoll, den jeweiligen Sinn in eckigen Klammern hinter den Backchannel zu schreiben.
	Beispiel: L Mhm [ja] L Hm-m [nein]

2.6 TRANSKRIPTION WÄHREND GRUPPENARBEIT

Während der Kleingruppenphase transkribiere nur **Schüler-Schüler-Interaktionen** innerhalb der videographierten Kleingruppe, sowie alle Interaktionen der Kleingruppenmitglieder mit anderen Personen.

3. Codiermanuale

3.1 Kategorisierungsmanual „Fokus“ (inhaltlich, organisatorisch, sonstiges)

Die Unterteilung in die drei genannten Kategorien soll dazu dienen, beurteilen zu können, wie groß der Anteil in den verschiedenen Arbeitsphasen ist. Es kann pro Turn nur eine Kategorie gewählt werden!

Kategorie „inhaltlich“ (1)

- Sobald sich in einem Turn ein inhaltlicher Aspekt findet, ist der Turn als inhaltlich zu setzen. Andere Kategorien innerhalb des Turns sind zu vernachlässigen!
- Alle fachlichen Äußerungen, wie Merksätze, Definitionen, Beiträge zum Sachgespräch, sachbezogene Fragen.
- Antworten auf sachbezogene Fragen ➔ auch Einwort-Antwort: Aha, ja, nein, mh...
- Inhaltliche Überlegungen zum Versuchsaufbau.
- Organisatorische Überlegungen zum Versuchsaufbau, wenn damit ein inhaltliches Problem bearbeitet wird (z.B.: „Wir können ja die Kreide in das Petroleum stellen“).
- Organisatorische Anweisungen, die einen inhaltlichen Lerneffekt haben (z.B.: „Nicht so, du musst das zufächern“).
- Zitieren von Aufgabenteilen zur Problemlösung ➔ Zitat für Widerlegung oder Bestätigung
- Genannte oder aufgezählte Problemelemente (vorausgreifend auf Kategorisierungsebene 2).

Kategorie „organisatorisch“ (2)

- Begrüßung
- Handlungsanweisungen
- Aufrufen / Herbeirufen / Anrufen von Schülern oder des Lehrers.
- Disziplinierende Maßnahmen
- Antworten auf organisatorische Fragen (➔ Achtung! Es können auch inhaltliche Antworten vorkommen! Diese sind dann auch inhaltlich zu werten!)
- Vorlesen / Diktieren von Aufgabenstellungen, Merksätzen, Protokollteilen...
- Mitsprechen beim Schreiben.
- Vorbereiten von Experimenten (Klären, wer Schutzbrille holt...).
- Diskussion über Rechtschreibung oder die Protokollgestaltung.
- Ausleihen von Stiften, Zetteln...
- Rückfragen, weil etwas akustisch nicht verstanden wurde (Die erneute Antwort ist in diesem Fall wieder der gleichen Kategorie zuzuordnen, wie zuvor).

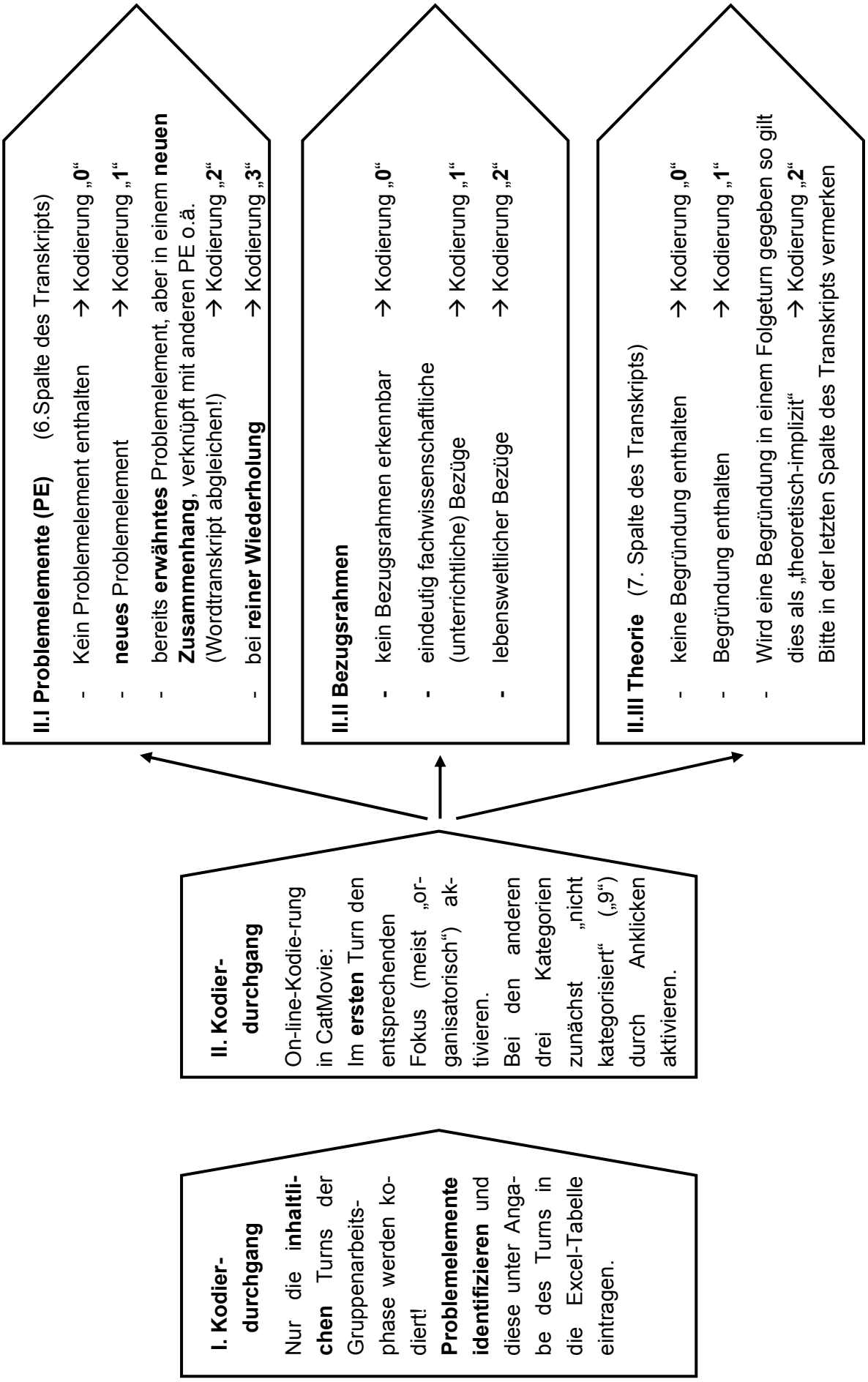
Kategorie „sonstiges“ (3)

- Privatgespräche ohne Bezug zum Unterricht
  ➔ bei allen anderen Merkmalen dieser Kategorie (s.u.) muss besonders darauf geachtet, werden, ob der betrachtete Turn nicht Verbindungen zu vorhergegangenen oder nachfolgenden Turns hat, welche diesen als theoretisch oder inhaltlich ausweisen.
- Äußerungen, die zwar aus dem Unterrichtsverlauf hervorgehen, aber keinen inhaltlichen oder organisatorischen Gehalt haben
  (z.B.: Ein Schüler hat Petroleum an den Fingern – „Du stinkst!“).
- Äußerungen, die dem ganzen Klassenverband zuzuordnen und nicht zu unterscheiden sind. (z.B.: Gemurmel in der Klasse)
- Einwürfe wie „Boah“, „Hä“, „Haha“, „Mist“... (➔ hier ist besonders auf Beziehungen zu anderen Turns zu achten!).
- Alle Äußerungen, die keiner anderen Kategorie zugeordnet werden können.

Keine Kategorie (0)

- Äußerungen, die nicht transkribiert sind, weil sie akustisch unverständlich sind.
- Äußerungen, die aufgrund von Unvollständigkeit nicht zugeordnet werden können.

3.2 Inhaltliche-Progressions-Analyse (IPA) – Kodierschema



Was ist ein Problemelement?
Jeglicher praktischer Gebrauch, die Aufforderung hierzu oder jede theoretische Eigenschaftsumschreibung eines *Gegenstands der Aktionsbox* sowie hieraus abgeleitete Begriffe. Weiterhin zählen wir immaterielle *Substantive* aus dem Themenbereich Säure-Base wie pH-Wert (-Einheiten) oder Neutralisation zu den Problemelementen.

II.I Problemelemente:
D.h.:

- Gegenstände aus den Boxen werden mit einer Funktion versehen.
- Gegenstände aus den Boxen werden mit einer Eigenschaft versehen.

  Achtung: diese Eigenschaft kann zur Generierung eines neuen Problemelements führen, z. B.: WC-Reiniger, Rohrreiniger, Feinwaschmittel und Glasreiniger werden gemeinschaftlich als *Reinigungsmittel* kategorisiert. *Reinigungsmittel* wäre in diesem Fall ein neues Problemelement. (oder alternativ: *giftige Lösungen*)
- Mitschülern werden Handlungsanweisungen zur Benutzung oder Kombination von Gegenständen aus der Box gegeben.

1. Im ersten Kodierdurchgang werden die Problemelemente identifiziert und in der Excel-Liste vermerkt.
2. Im zweiten Kodierdurchgang werden die identifizierten Problemelemente kategorisiert.
   a. Inhaltliche Äußerungen *ohne Problemelement* erhalten die Kodierung ***„0“***.
   b. Bei der *erstmaligen Nennung* eines Problemelements ist die Zuordnung einfach, die Nennung ist neu und damit ***konstruktiv-differenziert (1)***.
   c. Immer dann, wenn die Excel-Tabelle die wiederholte Verwendung eines Problemelements anzeigt wird in die vorher laufenden Turns des Word-Transkripts geschaut, die ebenfalls dieses Problemelement enthalten. Wird das Problemelement in einem *neuen Kontext* (neue Argumentation, Kombination mit anderen Problemelementen, Bezug auf bereits getätigte Erkenntnisse) benutzt, so gilt dieser Beitrag als ***konstruktiv-integriert (2)***. In der vorletzten Spalte des Transkripts wird ebenfalls eine „2“ notiert und die Nummer des bezugbildenden Turns notiert.

      Achtung: Wird einem Problemelement eine *neue Eigenschaft* zugeschrieben und ist ferner *kein Bezug* zu bereits vorher getätigten Aussagen zu erkennen, so gilt dies ebenfalls als ***konstruktiv-differenziert (1)***!!!
   d. Problemelemente, die ein zweites Mal im *gleichen Kontext*, mit der gleichen Eigenschaftszuschreibung versehen werden, gelten als reine *Wiederholung* und sind somit als ***redundant (3)*** zu kodieren.

II.II Bezugsrahmen:

1. Ist der Bezugsrahmen der Äußerung nicht zu erkennen, so erfolgt die wiederum die Kodierung ***„0“***.
2. Wird in der Aussage ein fachwissenschaftlicher Bezug erkennbar, so erfolgt die Kodierung ***fachwissenschaftlich (1).***
3. Wird in der Aussage ein lebensweltlicher/ alltagsweltlicher Bezug erkennbar, so erfolgt die Kodierung ***lebensweltlich (2).***

Entscheidungshilfen für die Frage lebensweltlich vs. fachwissenschaftlich.

Besondere Kennzeichen eine lebensweltlichen Bezugsrahmen sind:

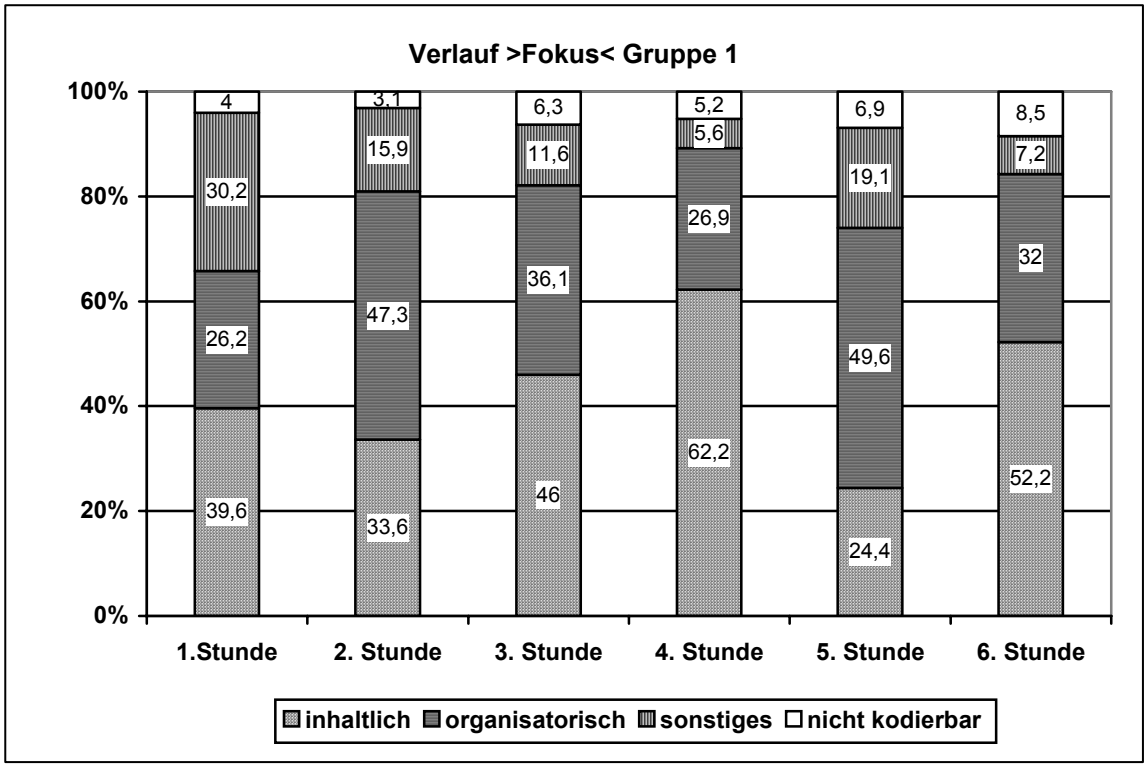
- Eigenschaften von Säuren: prinzipiell gefährlich, ätzend (unabhängig von der Konzentration), großer pH-Wert
- Eigenschaften von Basen: eher ungefährlich, können Säuren „süßen“, sind synonym mit Wasch- (Reinigungsmitteln), haben kleinen pH-Wert.
-
-
-

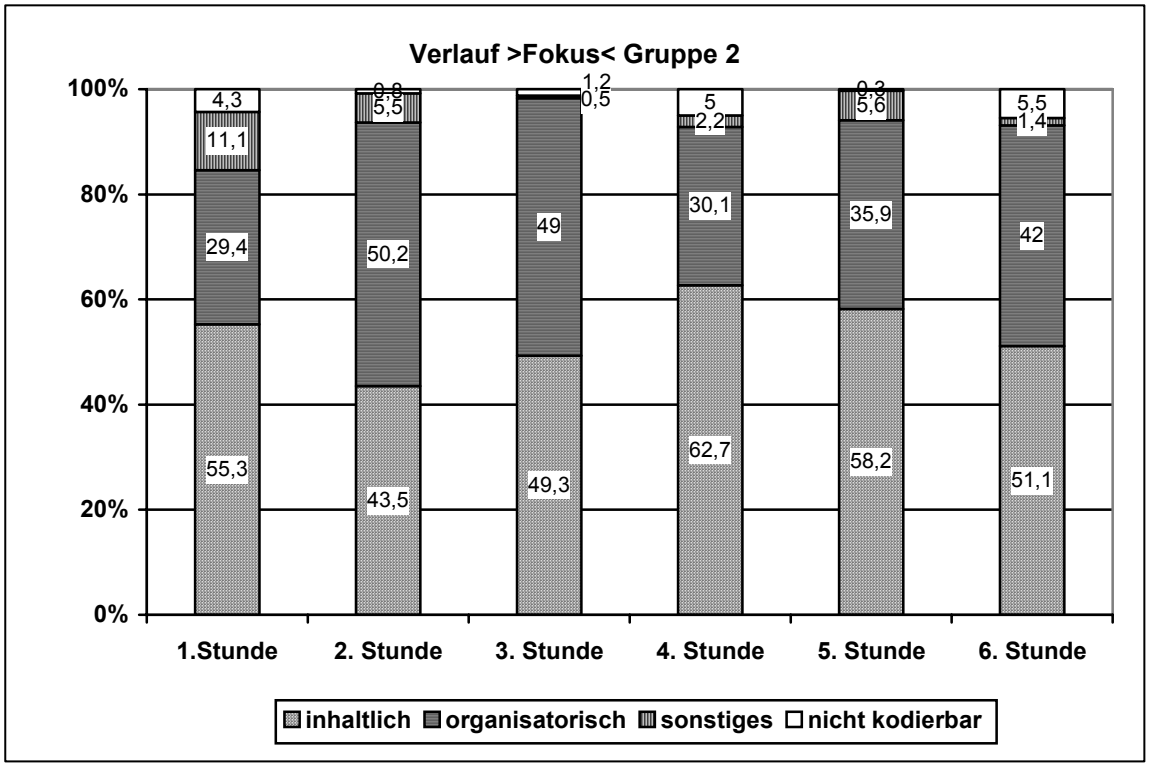
Platz für Ergänzungen

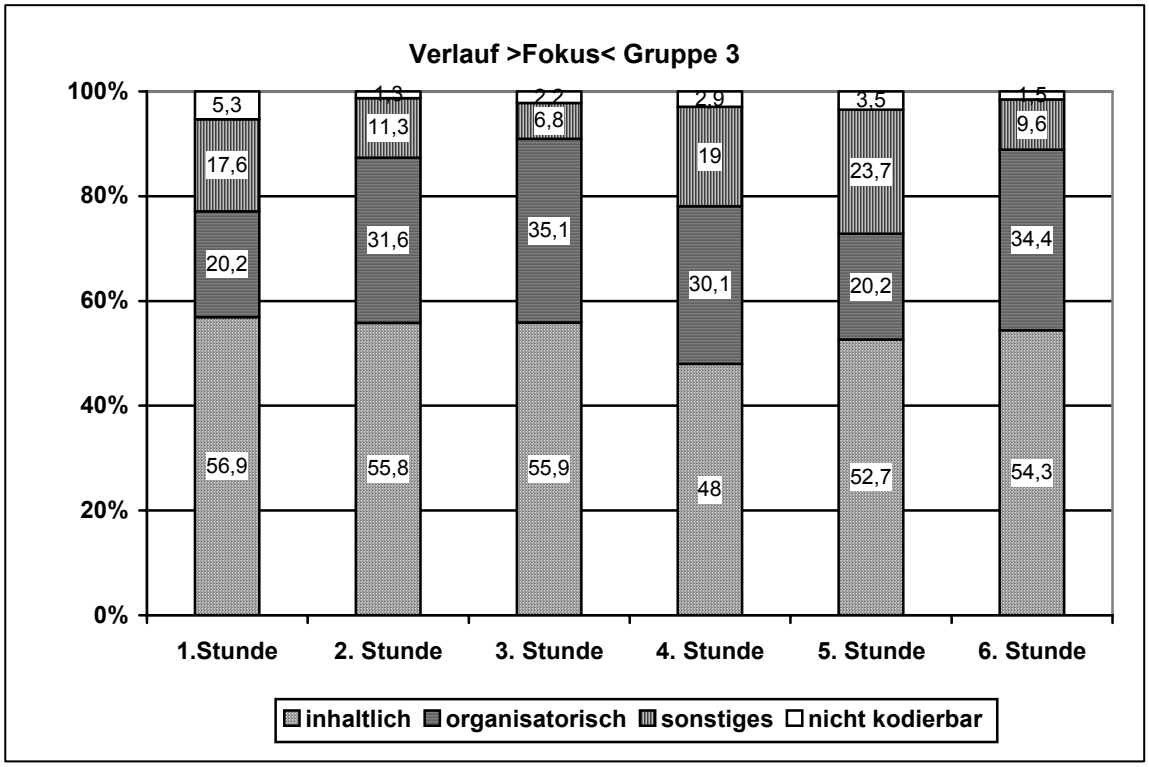
II.III Theorie:

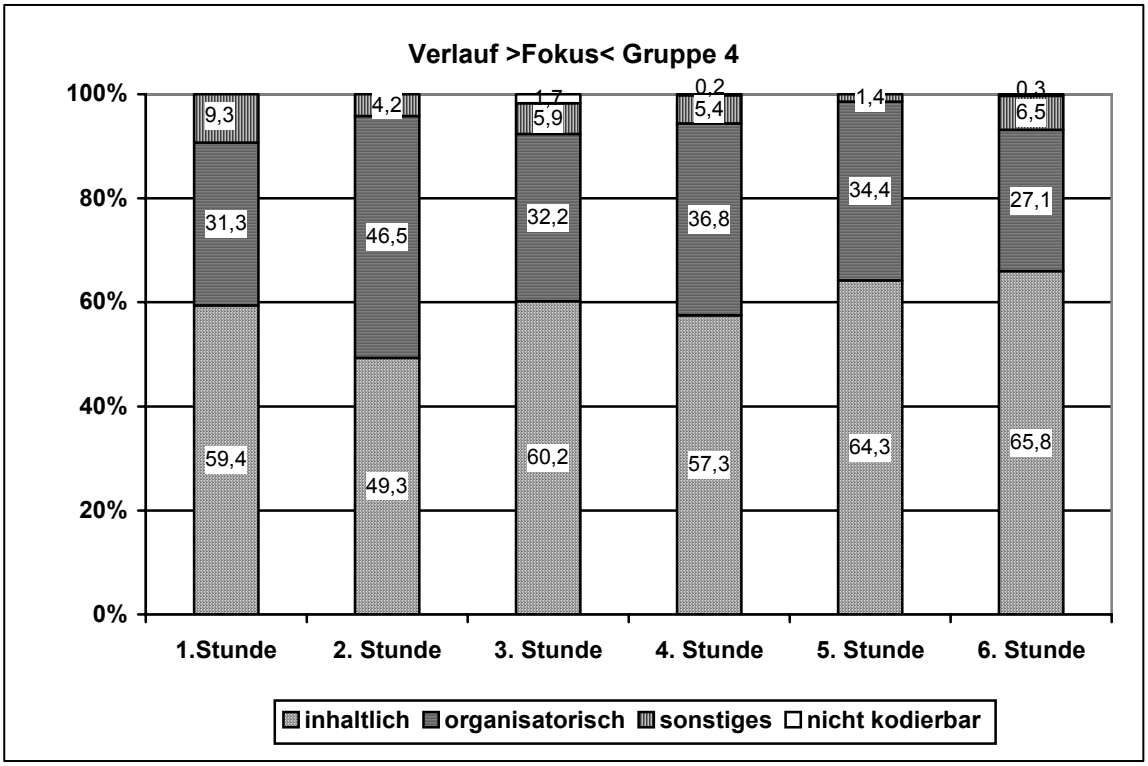
1. Ist *keine theoretische Fundierung* in der Äußerung zu erkennen (= reine Behauptung), so wird der Turn als ***atheoretisch (0)*** kodiert.
2. Ist der Äußerung eine *Begründung* oder auch nur ein *Bezug auf* eine – auch unterrichtsunabhängige – *theoretische Fundierung* zu entnehmen, so wird dies als ***theoretisch-explizit (1)*** kodiert.
3. Sollte eine als *atheoretisch* identifizierte Äußerungen in einem *Folgeturn* von einem (Mit)schüler *theoretisch untermauert oder mit einer Begründung versehen* werden, so wird dies als ***theoretisch-implizit (2)*** kodiert. In diesem Fall bitte in der letzen Spalte des Word-Transkripts eine „2“ unter Verweis auf den Bezugsturn angeben.
   Anm:. Insbesondere zur Identifikation von theoretisch-impliziten Äußerungen ist es sinnvoll, das Word-Transkript passagenweise zu lesen.

E Fokusverläufe der vier Kleingruppen






Verlauf >Fokus< Gruppe 3
100%
80%
60%
40%
20%
0%
5,3
17,6
20,2
56,9
1,3
11,3
31,6
55,8
2,2
6,8
35,1
55,9
2,9
19
30,1
48
3,5
23,7
20,2
52,7
1,5
9,6
34,4
54,3
1.Stunde
2. Stunde
3. Stunde
4. Stunde
5. Stunde
6. Stunde
inhaltlich
organisatorisch
sonstiges
nicht kodierbar


Verlauf >Fokus< Gruppe 4
100%
80%
60%
40%
20%
0%
9,3
31,3
59,4
4,2
46,5
49,3
1,7
5,9
32,2
60,2
0,2
5,4
36,8
57,3
1,4
34,4
64,3
0,3
6,5
27,1
65,8
1.Stunde
2. Stunde
3. Stunde
4. Stunde
5. Stunde
6. Stunde
inhaltlich
organisatorisch
sonstiges
nicht kodierbar

F. Liste der Problemelemente der sechs Unterrichtsstunden

			turn-Nr.	turn-Nr.	turn-Nr.
1	·	WC-Reiniger			
2	·	Salzsäure			
3	·	Zitronensaft			
4	·	Speisessig / Essig(säure)			
5	·	Limonade (Sprudel.etc.)			
6	·	Mineralwasser			
7	·	Dest. Wasser			
8	·	Wasser			
9	·	Zuckerwasser			
10	·	Feinwaschmittel			
11	·	Glasreiniger			
12	·	Kalkwasser			
13	·	Natriumhydroxid(lsg.) / Natronlauge			
14	·	Rohrreiniger			
15	·	Saurer Regen			
16	·	Rotkohlsaft			
17	·	Bromthymolblau (o.ä.)			
18	·	pH-(Test)Stäbchen (o.ä. Formulierungen)			
19	·	Reagenzglas / Stopfen / Ständer			
20	·	Becherglas			
21	·	Zucker			
22	·	Kochsalz			
23	·	Calciumoxid			
24	·	Magnesiumoxid			
25	·	Metalloxid			
26	·	Schwefel(di)oxid			
27	·	Stick(stoff)oxid (~monoxid ~dioxid)			
28	·	Nichtmetalloxid			
29	·	Indikator			
30	·	Farbumschlag (-wechsel etc.)			
31	·	saure Lösung (Säure)			
32	·	basische Lösung (Base)			
33	·	neutrale Lösung			
34	·	Neutralisation			
35	·	pH-Einheiten			
36	·	Lösung			