

Das Experiment als Lerngelegenheit

Vom interkulturellen Vergleich des
Physikunterrichts
zu Merkmalen seiner Qualität

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Johannes Börlin
aus Bubendorf Baselland

Berlin, 2012

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Christoph Bruder, Prof. Dr. Peter Labudde und
Prof. Dr. Andreas Müller (Genf)

Basel, den 27. März 2012

Prof. Dr. Martin Spiess



“Das Experiment als Lerngelegenheit. Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität” von Johannes Börlin steht unter einer Creative Commons Namensnennung-NichtKommerziell-KeineBearbeitung 3.0 Schweiz Lizenz.

Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung-NichtKommerziell-KeineBearbeitung 3.0 Schweiz zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <http://creativecommons.org/licenses/by-nc-nd/3.0/ch/> oder wenden Sie sich brieflich an Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Die Dissertation ist beim *Logos Verlag Berlin* (www.logos-verlag.de) in der Schriftenreihe *Studien zum Physik- und Chemielernen* als gedruckte Ausgabe erhältlich.

Zusammenfassung

Die vorliegende Arbeit ist Teil der Videostudie “Quality of Instruction in Physics” (QuIP), welche Merkmale gelingenden Physikunterrichts in den Ländern Deutschland, Finnland und der Schweiz untersucht. Die Studie verfolgt die Hauptziele: 1. Die Identifikation von Qualitätsaspekten des Physikunterrichts sowie 2. Die Erklärung der Leistungsunterschiede zwischen deutschen, finnischen und Schweizer Schülerinnen und Schüler auf der Unterrichtsebene (vgl. OECD, 2004, 2007, 2010).

Es wurden 99 Doppelstunden in Klassen des 9. und 10. Schuljahres aller Schultypen zum Thema “Zusammenhang zwischen elektrischer Energie und Leistung” videographiert. Zur Erhebung des Leistungszuwachses sowie der Veränderung der affektiven Variablen kamen Testinstrumente zum Einsatz (Pre-Posttest-Design). Die theoretische Basis bildet ein systemisches Modell des Unterrichts, welches Einflussgrößen, wie das pädagogische, fachliche Wissen der Lehrperson und den sozioökonomischen Hintergrund der Schülerinnen und Schüler kontrolliert (Lipowsky et al., 2005, Neumann et al., 2009, 2010).

Das experimentelle Handeln als charakteristisches Merkmal des Physikunterrichts bildet den Schwerpunkt des vorliegenden Teils der Studie. In einem ersten Analysedurchgang des Videomaterials wurden Sichtstrukturmerkmale experimentellen Handelns, wie die Vorbereitung des Experiments, die Durchführung und die Nachbereitung identifiziert. Das hierfür eingesetzte Kategoriensystem adaptierte man von Tesch (2005). Ein weiterer Analysedurchgang galt der Bewertung der zuvor identifizierten Experimentiereinheiten anhand von Qualitätskriterien (Tiefenstruktur). Diese sind aus dem Verständnis experimentellen Handelns als kontextorientierte, reflexive und theoriegeleitete Aktivität abgeleitet (Hodson, 1993, Millar et al., 1999) und bilden die Grundlage des eingesetzten, hoch-inferenten Ratingsystems.

Die Sichtstrukturanalyse experimentellen Handelns fördert signifikante Unterschiede zwischen den Ländern zutage: In Deutschland stehen zwei Drittel, in der Schweiz die Hälfte und in Finnland ein Drittel der Unterrichtszeit im Kontext experimentellen Handelns. Im deutschen Unterricht waren quantitative Experimente mehr als doppelt so häufig vertreten wie qualitative, wogegen in der Schweizer Stichprobe beide Arten des Experiments etwa gleich häufig auftraten. Im finnischen Unterricht überwogen qualitative Experimente leicht.

Eine Teilstichprobe von je 15 videographierten Doppelstunden pro Land wurde bezüglich seiner Tiefenstrukturmerkmale untersucht. Die Analyse zeigt, dass das

experimentelle Handeln im Unterricht kaum dem Verständnis einer kontextorientierten, reflexiven und theoriegeleiteten Aktivität entspricht. So sind die Experimentiereinheiten bei einer Mehrheit der untersuchten Doppelstunden nur teilweise eingebettet oder zielklar aufgebaut. Qualitative Vorstellungen zum Experiment werden selten thematisiert. Es dominiert die Erarbeitung quantitativer Zusammenhänge als Ziel experimentellen Handelns. In nur vier (von 68) Experimentiereinheiten hatten Schülerinnen und Schüler die Möglichkeit einzelne Aspekte des Experiments selbst zu gestalten.

In Fallanalysen wurden Lehrertranskripte der Experimentiereinheiten von vier Doppelstunden untersucht, welche sich durch hohe Qualität auszeichneten. Die Fallanalysen verfolgten das Ziel, die Ergebnisse der Tiefenstrukturanalyse exemplarisch zu illustrieren und das Ratinginstrument zu validieren. Folgende Gemeinsamkeiten konnten zwischen den vier Fällen hoher Qualität identifiziert werden. 1. Das Herstellen von Bezügen und Analogien, 2. Das Sichtbarmachen von Prozessen, 3. Die Auseinandersetzung mit Schülerbeiträgen sowie 4. Die Orientierung an Vorstellungen. Die Befunde der Fallanalysen decken sich überwiegend mit den Ergebnissen der Tiefenstrukturanalyse.

Naturwissenschaftsdidaktik, Physikdidaktik, Experiment, Experimentelles Handeln, Videostudie, Ländervergleich

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung und Ziele der Arbeit	1
1.2	Gliederung	2
I	Theoretische Grundlagen	5
2	Forschungsfragen	7
2.1	Verständnis von der Qualität experimentellen Handelns	8
2.2	Sichtstruktur experimentellen Handelns	8
2.3	Tiefenstruktur experimentellen Handelns	9
2.4	Bedeutung experimentellen Handelns für den Zuwachs an Fachwissenleistung	10
3	Experimentelles Handeln	11
3.1	Die fachdidaktische Dimension experimentellen Handelns	14
3.2	Die wissenschaftstheoretische Dimension experimentellen Handelns	18
3.3	Ziele und Vorstellungen von Lehrpersonen	22
3.4	Experimentelles Handeln als Kompetenz	27
3.5	Befunde der unterrichtlichen Praxis	29
3.5.1	Videostudien	29
3.5.2	Studien zur Effektivität experimentellen Handelns	32
3.6	Qualitätskriterien experimentellen Handelns im Physikunterricht	34
II	Methodik	39
4	Die Videostudie QuIP	41

4.1	Eine Übersicht über die Methodik des Gesamtprojektes	41
4.2	Datenerhebung	44
4.2.1	Datenerhebung in Deutschland und Finnland	44
4.2.2	Datenerhebung in der Schweiz	45
5	Schwerpunkt: Experimentelles Handeln	49
5.1	Sichtstruktur	49
5.1.1	Training der Kodierenden	51
5.1.2	Interraterreliabilität	53
5.2	Tiefenstruktur	56
5.2.1	Wahl der Teilstichprobe	59
5.2.2	Ablauf des Ratings	60
5.2.3	Interraterreliabilität	62
5.3	Das Instrument zur Analyse der Tiefenstruktur	64
5.3.1	Allgemeiner Teil	64
5.3.2	Experimentelles Handeln als kontextorientierte Aktivität	67
5.3.3	Experimentelles Handeln als reflexive Aktivität	70
5.3.4	Experimentelles Handeln als theoriegeleitete Aktivität	72
5.4	Fallanalysen	76
III	Ergebnisse	79
6	Analyse der Sichtstruktur experimentellen Handelns	81
6.1	Stichprobe	81
6.2	Analysemethodik	82
6.3	Ländervergleich	83
6.3.1	Unterricht im Kontext experimentellen Handelns	83
6.3.2	Die Phasen experimentellen Handelns	84
6.3.3	Art der Beschäftigung	88
6.3.4	Datenerfassung beim Experimentieren	88
6.3.5	Organisationsform und Arbeitsverteilung	91
6.3.6	Sachbegegnung	93
6.3.7	Die Experimentiereinheit	96
6.4	Schultypenvergleich	99

7	Analyse der Tiefenstruktur experimentellen Handelns	101
7.1	Analyse auf Klassenebene	102
7.2	Analyse auf der Ebene einzelner Experimentiereinheiten	109
7.3	Qualitative Analyse ausgewählter Experimentiereinheiten	113
7.3.1	Experimentiereinheit 1, S-1-05-01	114
7.3.2	Experimentiereinheit 1, F-0-16-01	121
7.3.3	Experimentiereinheit 1, D-1-29-01	130
7.3.4	Experimentiereinheit 1, S-3-31-01	137
7.3.5	Übersicht über die Kategorien der Transkriptanalyse	142
8	Experimentelles Handeln und der Zuwachs an Fachwissensleistung	147
8.1	Leistungsdaten	148
8.2	Der Zusammenhang zwischen der Sichtstruktur experimentellen Handelns und dem Leistungszuwachs	150
8.3	Der Zusammenhang zwischen der Tiefenstruktur experimentellen Handelns und dem Leistungszuwachs	152
IV	Diskussion und Ausblick	155
9	Diskussion und Ausblick	157
9.1	Zusammenfassung zentraler Ergebnisse der Sichtstruktur experimentellen Handelns	157
9.1.1	Unterrichtszeit im Kontext experimentellen Handelns	158
9.1.2	Phasen experimentellen Handelns	160
9.1.3	Experimentiereinheiten	161
9.1.4	Schultypen	162
9.2	Diskussion Sichtstruktur	162
9.3	Zusammenfassung zentraler Ergebnisse der Tiefenstruktur experimentellen Handelns	166
9.4	Diskussion Tiefenstruktur	171
9.5	Zusammenfassung zentraler Ergebnisse der Fallanalysen	175
9.6	Diskussion Fallanalysen	179
9.7	Diskussion zentraler Ergebnisse zum Zusammenhang zwischen experimentellem Handeln und Fachwissensleistung	179

9.8	Methodenreflexion	181
9.8.1	Gesamtstudie	181
9.8.2	Sichtstruktur	181
9.8.3	Tiefenstruktur	182
9.9	Ausblick	183
9.9.1	Implikationen für die Schulpraxis	183
9.9.2	Forschungsdesiderata	186
Literaturverzeichnis		188
Anhang		198
A Sichtstruktur experimentellen Handelns		199
A.1	Das Kategoriensystem (vgl. Tesch, 2005)	199
A.2	Beobachterübereinstimmung der Sichtstruktur	209
A.2.1	Übereinstimmungsmasse Phase 1	209
A.2.2	Übereinstimmungsmasse Phase 2	210
B Tiefenstruktur experimentellen Handelns		215
C Fallanalysen		217
Lebenslauf		220

Abbildungsverzeichnis

3.1	Verhältnis von Experiment und Theorie nach dem Verständnis von Duit & Tesch (in Vorb.).	20
3.2	Experimentelles Handeln im Unterricht als Verbindung der Wissensdomänen “domain of objects and observable things” und “domain of ideas” (Millar et al., 2002).	21
3.3	Das Kategoriensystem von Tesch (2005) (gekürzte Fassung).	30
4.1	Systemisches Modell für Unterrichtsqualität: Der durch die Sicht- und Tiefenstruktur beschriebene Unterricht wird durch den Lehrer und die Schüler beeinflusst. Als Output werden Veränderungen der Kompetenz sowie Veränderungen affektiver Variablen der Schülerinnen und Schüler erfasst.	43
4.2	Zeitlicher Ablauf der Datenerhebung: Pretest, Videoaufnahme einer Doppelstunde und Posttest. Das Pre-/Posttestdesign erlaubt die Ermittlung der Veränderung von Schülerkompetenz und -motivation über das Halbjahr.	44
5.1	Kategorien der Sichtstruktur experimentellen Handelns basierend auf Tesch (2005). E1: Identifizierung Experimentierphasen; G3: inhaltliche oder organisatorische Beschäftigung während der Vor- und Nachbereitung; K1, K2a, K2b, K3, K5: Kategorisierung der Durchführungsphase.	52
5.2	Die Struktur des Ratinginstruments bestehend aus den drei Perspektiven experimentellen Handelns und einem allgemeinen Teil (abgerundete Rechtecke). Den Subkategorien oder Facetten (Rechtecke) entsprechen Items, die auf einer vierstufigen Likert-Skala beurteilt werden.	57

6.1	Die Zeitdauer pro Doppelstunde, die im deutschen, finnischen und Schweizer Physikunterricht im Zusammenhang mit experimentellem Handeln steht, entspricht den Angaben unter <i>Experimentelles Handeln insgesamt</i> von Tabelle 6.2.	84
6.2	Der Anteil in Prozenten der <i>Vorbereitung</i> , <i>Durchführung</i> und <i>Nachbereitung</i> am experimentellen Handeln insgesamt, ausgewiesen für den deutschen, finnischen und den Schweizer Physikunterricht.	87
6.3	Zeitdauer pro Doppelstunde, während der qualitative (weiss) bzw. quantitative (hellgrau) Experimente durchgeführt werden. Die Kategorie <i>Unklar</i> (im Diagramm durch die Markierung des Medians bei null wiedergegeben; der Interquartilsabstand verschwindet hier) umfasst jenen Teil der <i>Durchführung</i> , bei der qualitative und quantitative Experimente gleichzeitig bearbeitet werden.	89
6.4	Zeitdauer, die pro Doppelstunde im Mittel für Demonstrationsexperimente der Lehrperson (weiss), für Schülerexperimente als Partnerarbeit (hellgrau), als Gruppenarbeit (dunkelgrau) aufgewendet wird.	92
6.5	Zeitdauer, die pro Doppelstunde im Mittel für die <i>Durchführung</i> von Experimenten mit Alltagsgeräten (weiss), mit physikalischen Geräten (hellgrau, verschwindender Interquartilsabstand) und am Bildschirm (dunkelgrau) aufgewendet wird.	94
6.6	Die Zeitdauer, die im deutschen, finnischen und Schweizer Physikunterricht pro Experimentiereinheit gesamthaft (weiss), als Lehrer-Demonstration (hellgrau) und als Experimentiereinheit in Gruppen (dunkelgrau) aufgewendet wird.	97
6.7	Die Zeitdauer pro Doppelstunde die in den Schultypen 0-3 für experimentelles Handeln gesamthaft (weiss), für die Vorbereitung (dunkelgrau), <i>Durchführung</i> (grau) und <i>Nachbereitung</i> (hellgrau) eingesetzt wird (Stichproben D und CH zusammengenommen). Typ 1, tiefstes; 2, mittleres und 3 höchstes Leistungsniveau (Typ 0, typengemischte Klasse).	99

7.1	Die Struktur des Ratinginstruments besteht aus den drei Perspektiven experimentellen Handelns und einem allgemeinen Teil (abgerundete Rechtecke). Den Subkategorien bzw. Facetten (Rechtecke) entsprechen Items, die auf einer 4-stufigen Likert-Skala beurteilt werden.	102
7.2	Überblick über die Ergebnisse der Tiefenstruktur: Zu den acht Facetten der drei Perspektiven <i>Kontextorientierung</i> , <i>Reflexivität</i> und <i>Theorieleitung</i> sind die Ergebnisse für jede Klasse zusammengefasst. <i>Weiss=trifft zu</i> , <i>hellgrau=trifft grösstenteils zu</i> , <i>dunkelgrau=trifft teilweise zu</i> und <i>schwarz=trifft nicht zu</i> . Die Klassen sind entsprechend ihres Rating-Summenwerts geordnet.	104
7.3	Kontextorientierung und assoziierte Facetten wie <i>Fachimmanente Einbettung</i> , <i>Lebensweltliche Einbettung</i> und <i>Zielklarheit</i> : Häufigkeitsverteilungen der Gesamtratings pro Land und Facette. Ausprägungen: <i>1=trifft zu</i> , <i>2=trifft grösstenteils zu</i> , <i>3=trifft teilweise zu</i> und <i>4=trifft nicht zu</i> . Median, strichpunktierte Linie; Grenzen des 95%-Konfidenzintervalls des Medians, gestrichelte Linien.	105
7.4	Reflexivität und assoziierte Facetten wie <i>Prozess Reflexion</i> und <i>Ergebnis Reflexion</i> : Häufigkeitsverteilungen der Gesamtratings pro Land und Facette. Ausprägungen: <i>1=trifft zu</i> , <i>2=trifft grösstenteils zu</i> , <i>3=trifft teilweise zu</i> und <i>4=trifft nicht zu</i> . Median, strichpunktierte Linie; Grenzen des 95%-Konfidenzintervalls des Medians, gestrichelte Linien.	106
7.5	Theorieleitung und assoziierte Facetten wie <i>Kognitiver Anspruch</i> , <i>Theoretische Fundierung</i> und <i>Begriffsbildung</i> : Häufigkeitsverteilungen der Gesamtratings pro Land und Facette. Ausprägungen: <i>1=trifft zu</i> , <i>2=trifft grösstenteils zu</i> , <i>3=trifft teilweise zu</i> und <i>4=trifft nicht zu</i> . Median, strichpunktierte Linie; Grenzen des 95%-Konfidenzintervalls des Medians, gestrichelte Linien.	107
7.6	Strukturdiagramm der Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, S-1-05-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.	119

7.7	Strukturdiagramm der Ebenen (Ideenebene IE, Objektebene OE), Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, F-0-16-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.	127
7.8	Strukturdiagramm der Ebenen (Ideenebene IE, Sozial-kommunikative Ebene SE), Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, D-1-29-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.	134
7.9	Strukturdiagramm der Ebenen (Ideenebene IE, Sozial-kommunikative Ebene SE), Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, S-3-31-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.	140
8.1	Das Diagramm aus Geller et al. (in Vorb.) zeigt die Mittelwerte und die 95%-Konfidenzintervalle der Schülerfähigkeiten beim Pre- und beim Posttest in Finnland, Deutschland und der Schweiz. <i>Students' Content Knowledge</i> : Der Wert Null entspricht dem Mittelwert der Aufgabenschwierigkeit.	149
9.1	Das Kategoriensystem der Sichtstruktur experimentellen Handelns basierend auf Tesch (2005) (entspricht Abbildung 5.1).	158
9.2	Die Zeitdauer pro Doppelstunde, die im deutschen, finnischen und Schweizer Physikunterricht im Zusammenhang mit experimentellem Handeln steht (entspricht Abbildung 6.1).	159
9.3	Die Struktur des Ratinginstruments, bestehend aus den drei Perspektiven experimentellen Handelns und einem allgemeinen Teil (abgerundete Rechtecke). Den Subkategorien bzw. Facetten (Rechtecke) entsprechen Items, die auf einer 4-stufigen Likert-Skala beurteilt werden (entspricht Abbildung 7.1).	167

- 9.4 Überblick über die Ergebnisse der Tiefenstruktur: Zu den acht Facetten der drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung* sind die Ergebnisse für jede Klasse zusammengefasst. *Weiss* = *trifft zu*, *hellgrau* = *trifft grösstenteils zu*, *dunkelgrau* = *trifft teilweise zu*, *schwarz* = *trifft nicht zu*. Die Klassen sind entsprechend ihres Rating-Summenwerts geordnet (entspricht Abbildung 7.2). . . . 168

Tabellenverzeichnis

3.1	Ziele experimentellen Handelns in absteigender Priorität als Ergebnis der Studien von Welzel et al. (1998) (links) und Swain et al. (2000), Beatty & Woolnough (1982) (rechts).	23
3.2	Ziele experimentellen Handelns aus der Sicht von Hofstein & Lunetta (2003) (links) und Hodson (1993) (rechts).	25
4.1	Anzahl der Fälle (Klassen) verteilt auf Schultypen (0-3) und Länder. Typ 3 entspricht dem höchsten, 2 dem mittleren und 1 dem tiefsten Leistungsniveau. Dem Typ 0 werden Schulklassen zugeordnet, dessen Schülerinnen und Schüler sich aus verschiedenen Leistungsniveaus (Schultyp 1-3) zusammensetzen.	45
4.2	Anzahl der Schülerinnen und Schüler nach Kantonen und Leistungsniveaus 1-3 (extrahiert aus Daten des Bundesamtes für Statistik 2005/06).	46
4.3	Verteilung der Stichprobe nach Kantonen und Leistungsniveaus 1-3. Die mit * versehenen Klassenzahlen entsprechen jenen Klassen, welche über den direkten Kontakt zur Lehrperson für die Teilnahme an der Studie gewonnen wurden. Sie gehören damit nicht zur Zufallsstichprobe.	48
5.1	Fälle, anhand derer die Beobachterübereinstimmung und das Cohens κ bestimmt wurden.	53
5.2	Cohens κ sowie prozentuale Übereinstimmung (PU) aller Kategorien für die Phase 1 der Sichtstrukturanalyse. Die Anzahl der analysierten Intervalle sind unter Anz. Int. aufgeführt.	54

5.3	Cohens κ sowie prozentuale Übereinstimmung (PU) aller Kategorien ausgewiesen für die Phase 2 der Sichtstrukturanalyse. Die Anzahl der analysierten Intervalle sind unter Anz. Int. aufgeführt.	55
5.4	Zusammenfassung der verwendeten Kategorien/Perspektiven, Facetten mit Abkürzungen und dem entsprechenden Skalenniveau. Nominal: FU1 ... FU8, Binär: 0000 ... 1111 (0 Nein, 1 Ja), Ordinal: 1 (trifft zu), 2 (trifft grösstenteils zu), 3 (trifft teilweise zu), 4 (trifft nicht zu).	59
5.5	Die nach Land und Schultyp ausgewiesenen Fallzahlen (Anzahl Doppelstunden) der Teilstichprobe.	60
5.6	Intraklassenkoeffizienten (ICC3: two-way mixed model), F Statistik und das 95%-Konfidenzintervall, ausgewiesen für jede Facette. Für sämtliche Berechnungen des ICC3 gelten: $df1 = df2 = 47$ und $p < .001$. Cohens κ und das 95%-Konfidenzintervall für die Kategorie <i>Funktion</i> . Die Berechnungen basieren auf den Ratings der Fälle aus Deutschland und der Schweiz.	62
6.1	Verteilung der Fälle (Klassen) auf die verschiedenen Schultypen: Typ 1 entspricht dem tiefsten, 2 dem mittleren und 3 dem höchsten Leistungsniveau. Klassen des Typs 0 umfassen Schülerinnen und Schüler verschiedener Schultypen (1-3) oder es wird nicht nach Schultypen unterschieden.	82
6.2	Zur Kategorie E1 sind die Lage- und Streuungsparameter der Subkategorien <i>Keine</i> , <i>Vorbereitung</i> , <i>Durchführung</i> und <i>Nachbereitung</i> aufgeführt. Unter <i>Experimentelles Handeln insgesamt</i> wird die Summe aus Vorbereitungs-, Durchführungs- und Nachbereitungszeit ausgewiesen. Die Lage- und Streuungsparameter sind in Minuten pro Doppelstunde zu verstehen.	85
6.3	Mittelwertvergleiche zur Kategorie <i>Experimentierphasen</i> . H_0 -Hypothesen: Die Länderpaare (D-FIN, D-CH oder FIN-CH) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. sind in Minuten pro Doppelstunde zu verstehen.	86
6.4	Zur Kategorie G3 sind die Lage- und Streuungsparameter der Subkategorien <i>Inhaltliche Beschäftigung</i> und <i>Organisatorische Beschäftigung</i> aufgeführt. Die Lage- und Streuungsparameter sind in Prozenten der Vor- und Nachbereitungszeit zu verstehen.	88

6.5	Mittelwertsvergleiche zur Kategorie <i>Datenerfassung</i> . H_0 -Hypothesen: Die Stichprobenpaare (D-FIN, D-CH, FIN-CH bzw. Qualitativ-Quantitativ) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. werden in Minuten angegeben.	90
6.6	Kategorie <i>Datenerfassung</i> K1: Absolute und relative (in Klammern) Häufigkeiten der Subkategorien <i>Qualitativ</i> (QL), <i>Quantitativ</i> (QT) und <i>Unklar</i> (UN). Bei den relativen Häufigkeiten entspricht 1 der Anzahl Doppelstunden eines Landes, während denen Experimente durchgeführt werden (D: 43, FIN: 22, CH: 31).	91
6.7	Absolute und relative Häufigkeiten (in Klammern) der Subkategorien: <i>Demonstration Lehrperson</i> (DL), <i>Demonstration Schülerin/Schüler</i> (DS), <i>Schülerin/Schüler einzeln</i> (SE), <i>Partnerarbeit</i> (PA), <i>Gruppenarbeit</i> (GR), <i>Stationen</i> (ST), <i>Arbeitsteilig</i> (AT). Relative Häufigkeiten: 1 entspricht der Anzahl Doppelstunden eines Landes, bei denen Experimente stattfinden (D: 43, FIN: 22, CH: 31).	93
6.8	Mittelwertsvergleiche zur Kategorie <i>Sachbegegnung</i> . H_0 -Hypothese: Die Stichprobenpaare (Alltagsgeräte-Physikalische Geräte) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. werden in Minuten angegeben.	95
6.9	Kategorie <i>Sachbegegnung</i> K5: Absolute und relative (in Klammern) Häufigkeiten der Subkategorien Physikalische Geräte (PH), Alltagsgeräte (AL) und Bildschirmexperiment (BI). Relative Häufigkeiten: 1 entspricht der Anzahl Doppelstunden eines Landes, bei denen Experimente stattfinden (D: 43, FIN: 22, CH: 31).	95
6.10	Anzahl der Fälle, bei denen pro Doppelstunde insgesamt n ($n = 0, 1, 2, \dots, 7$) Experimentiereinheiten durchgeführt wurden.	96
6.11	Mittelwertsvergleiche zur Kategorie <i>Organisationsform</i> . H_0 -Hypothesen: Die Stichprobenpaare (D-FIN, D-CH, FIN-CH oder Lehrer Demonstration-Gruppenarbeit) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. sind in Minuten pro Experimentiereinheit zu verstehen.	98

7.1	Ergebnisse der Tiefenstrukturanalyse auf Klassenebene. Zu jeder Facette sind angegeben: Median (Med), Modalwert (Mod), Interquartilsabstand (IQR) sowie das 95%-Konfidenzintervall des Medians (Konf.).	105
7.2	Spannweite Δ der Ratings aller Experimentiereinheiten einer Klasse. Aufgetragen werden absolute und relative Häufigkeiten () (in % bezogen auf die 45 Klassen der Teilstichprobe).	110
7.3	Anzahl der Experimentiereinheiten n , die den Merkmalen FU1 bis FU8 zugeordnet sind.	111
7.4	Zusammenhang zwischen der <i>Funktion</i> der Experimentiereinheit und den Facetten <i>Prozess Reflexion</i> , <i>Kognitiver Anspruch</i> und <i>Theoretische Fundierung</i> . Die Daten wurden auf die Form einer 4-Feldertafel mit den Faktoren Facette und Funktion umkodiert. Alternativ-Hypothese H_1 : <i>odds ratio</i> ist grösser als 1. p gibt die Irrtumswahrscheinlichkeit für das Eintreten von H_1 an. $\alpha = .05$	111
7.5	Anzahl der Experimentiereinheiten n , die den Merkmalen OF1 bis OF4 zugeordnet sind.	112
7.6	Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall S-1-05-01: Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Die Durchführung findet in Gruppen statt (Abkürzungen der Facetten, vgl. Abbildung 7.1).	120
7.7	Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall F-0-16-01: Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Die Organisation der Durchführung teilt sich hälftig in eine Partnerarbeit (Pa) und eine Lehrer-Demonstration (DI) (Abkürzungen der Facetten, vgl. Abbildung 7.1).	129
7.8	Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall D-1-29-01: Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Das Experiment wird als Lehrerdemonstration durchgeführt (Abkürzungen der Facetten, vgl. Abbildung 7.1).	136

7.9	Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall S-3-31-01: Angegeben ist die Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Das Experiment wird als Lehrerdemonstration durchgeführt (Abkürzungen der Facetten, vgl. Abbildungen 7.1).	142
7.10	Kategorien und Subkategorien der Ideenebene: <i>Subkategorie</i> : Bezeichnung der Subkategorien (fett), Fall: Ursprüngliche Bezeichnung; <i>Beschreibung</i> : Kurze Beschreibung der Subkategorie; <i>Paraphrase</i> : Paraphrase zur entsprechenden Subkategorie.	143
7.11	Kategorien und Subkategorien der Objektebene: <i>Subkategorie</i> : Bezeichnung der Subkategorien (fett), Fall: Ursprüngliche Bezeichnung; <i>Beschreibung</i> : Kurze Beschreibung der Subkategorie; <i>Paraphrase</i> : Paraphrase zur entsprechenden Subkategorie.	144
7.12	Kategorien und Subkategorien der sozial-kommunikativen Ebene: <i>Subkategorie</i> : Bezeichnung der Subkategorien (fett), Fall: Ursprüngliche Bezeichnung; <i>Beschreibung</i> : Kurze Beschreibung der Subkategorie; <i>Paraphrase</i> : Paraphrase zur entsprechenden Subkategorie.	145
7.13	Relative Häufigkeiten der zu einer Kategorie zugeordneten Exzerpte. Angaben in Prozenten bezogen auf das Total der zugeordneten Exzerpte (gerundet auf eine signifikante Ziffer).	146
8.1	Ergebnis des Shapiro-Wilk-Tests für die drei Länderstichproben. Bei der deutschen Stichprobe lässt sich die Annahme der Normalverteilung nicht mehr aufrecht erhalten (H_0 -Hypothese: Es liegt eine Normalverteilung vor). n Stichprobengrösse, W Teststatistik, p Irrtumswahrscheinlichkeit.	151
8.2	Ergebnisse der Regressionsanalyse (Modell 1, 2 und 3) nach dem Standardverfahren (Methode der kleinsten Quadrate). b_i Koeffizienten, SE b_i Standardabweichung der Koeffizienten, Sign. Signifikanz der Koeffizienten: * $p < .05$, ** $p < .01$ und *** $p < .001$. Modell 1: $R^2 = .09$, $F(1, 97) = 9.7$, $p = 0.002$; Modell 2: $R^2 = .28$, $F(3, 95) = 12.4$, $p < 0.001$; Modell 3: $R^2 = .28$, $F(5, 93) = 7.3$, $p < 0.001$.	152

8.3	Kovarianzanalyse zwischen der abhängigen Variablen <i>Leistungszuwachs</i> und der unabhängigen Variablen (Facette) sowie der Kovariate <i>Land</i> . Spalten: Facette, F (F-Statistik), p (Irrtumswahrscheinlichkeit), Signifikanz (- ($p \geq .1$), . ($p < .1$), * ($p < .05$), ** ($p < .01$), *** ($p < .001$)), η^2 (Effektstärke).	153
9.1	Spezifische Chancen von Lehrer- und Schülerexperimenten.	185
A.1	Kategorie Experimentierphasen E1, Analyseprozess Phase 1	210
A.2	Nummerierung Experimentierphasen E2, Analyseprozess Phase 1	210
A.3	Art der Beschäftigung G3, Analyseprozess Phase 1	211
A.4	Datenerfassung K1, Analyseprozess Phase 1	211
A.5	Organisationsform K2a, Analyseprozess Phase 1	211
A.6	Organisationsform Schülerexperiment K2b, Analyseprozess Phase 1	211
A.7	Arbeitsverteilung beim Schülerexperiment K3, Analyseprozess Phase 1	212
A.8	Sachbegegnung K5, Analyseprozess Phase 1	212
A.9	Experimentierphasen E1, Analyseprozess Phase 2	212
A.10	Nummerierung Experimentierphasen E2, Analyseprozess Phase 2	212
A.11	Art der Beschäftigung G3, Analyseprozess Phase 2	213
A.12	Datenerfassung K1, Analyseprozess Phase 2	213
A.13	Organisationsform K2a, Analyseprozess Phase 2	213
A.14	Organisationsform Schülerexperiment K2b, Analyseprozess Phase 2	213
A.15	Arbeitsverteilung beim Schülerexperiment K3, Analyseprozess Phase 2	214
A.16	Sachbegegnung K5, Analyseprozess Phase 2	214
B.1	Die Identifikationsnummern (ID) der Fälle (Klassen) der gezogenen Teilstichprobe. Die ID enthält folgende Angaben: <i>Land-Schultyp-Schulnummer-Klassennummer</i>	215

C.1	Den Clustern I-IX zugeordnete Experimentiereinheiten. Innerhalb eines Clusters sind die Ratings der Facetten homogen bzw. ähnlich. EB1 Fachimmanente Einbettung, EB2 Lebensweltliche Einbettung, ZK Zielklarheit, PR Prozessreflexion, ER Ergebnisreflexion, KA Kognitiver Anspruch, TF Theoretische Fundierung und BB Begriffsbildung. 1=trifft zu, 2=trifft grösstenteils zu, 3=trifft teilweise zu, 4 trifft nicht zu.	219
-----	---	-----

Kapitel 1

Einleitung

1.1 Problemstellung und Ziele der Arbeit

Die vorliegende Arbeit ist Teil des trinationalen Projektes “Quality of Instruction in Physics” (QuIP). Auslöser für das Projekt waren die PISA¹-Studien von 2000 und 2003 (OECD, 2001, 2004). Dort konnte gezeigt werden, dass finnische Schülerinnen und Schüler international einen Spitzenplatz bei den Leistungen in Naturwissenschaften belegen². Im Vergleich zu ihren deutschen und Schweizer Kolleginnen und Kollegen erreichten diese ein jeweils signifikant besseres Testergebnis. Die Befunde wurden in den nachfolgenden PISA-Studien weiter bestätigt (OECD, 2007, 2010).

Infolge der erwähnten Erkenntnisse hat das finnische Bildungssystem international grosse Aufmerksamkeit erlangt – so auch die Unsere³. Mögliche Gründe für die Ausnahmeleistungen der finnischen Schülerinnen und Schüler wurden aus den Erkenntnissen der PISA-Studien beim Bildungssystem identifiziert (Domisch, 2009, Kobarg & Prenzel, 2009). Bisher existieren jedoch keine systematischen Untersuchungen, die das finnische Unterrichtsgeschehen analysieren. Hier setzt das Projekt QuIP ein, um auf der Ebene des Unterrichts nach Erklärungen für die erwähnten Länderunterschiede zu suchen. Ziel ist es dabei, die Praxis des Physikunterrichts in

¹PISA ist ein Akronym für “Programme for International Student Assessment”

²Auch bezüglich der mathematischen Kompetenzen und der Lesekompetenz erreichen finnische Schülerinnen und Schüler im internationalen Vergleich Spitzenergebnisse.

³Das Projekt “Quality of Instruction in Physics” und ihre Partnerorganisationen sind: Universität Duisburg-Essen (Deutschland), Mitarbeitende: Hans Fischer, Cornelia Geller, Melanie Keller, Knut Neumann und Jennifer Olszewski; Universität Jyväskylä (Finnland), Mitarbeitende: Jussi Helaaakoski, Sami Lehesvuori und Jouni Viiri; Pädagogische Hochschule Bern und Pädagogische Hochschule der Fachhochschule Nordwestschweiz (Schweiz), Mitarbeitende: Johannes Börlin, Carolin Junge, Peter Labudde, Judith Riegnig, Matthias von Arx.

Deutschland, Finnland und der Schweiz systematisch zu beschreiben (vgl. Kapitel 4). Hinter der Frage, wie die Länderunterschiede auf der unterrichtlichen Ebene erklärt werden könnten, verbirgt sich die Grundsatzfrage, welche Aspekte überhaupt Unterrichtsqualität und im Speziellen die Qualität des Physikunterrichts ausmachen. Das Projekt QuIP verfolgt somit zwei Hauptziele:

1. Die Identifikation von Qualitätsaspekten des Physikunterrichts.
2. Die Erklärung der Leistungsunterschiede zwischen deutschen, finnischen und Schweizer Schülerinnen und Schüler auf der Unterrichtsebene.

Das QuIP-Projekt verfolgt diese Ziele aus unterschiedlichen Gesichtspunkten. Untersucht werden bzw. wurden pädagogisches Wissen (Olszewski, 2010) und Motivation der Lehrperson (Keller, 2011), kognitive Aktivierung (Geller, in Vorb.), Klassenführung, Inhaltsstruktur des Unterrichts (Helaakoski, in Vorb.) und experimentelles Handeln (vorliegende Arbeit).

Für das Dissertationsvorhaben zum experimentellen Handeln hat sich die Gewichtung der beiden Hauptziele wie folgt ergeben: Das erste Ziel, also die Identifikation von Qualitätsaspekten experimentellen Handelns, steht klar im Vordergrund des Vorhabens, wobei hierzu erarbeitet wird, was unter solchen überhaupt zu verstehen ist. Dem zweiten Ziel wird insofern Rechnung getragen, als dass länderspezifische Unterschiede experimentellen Handelns diskutiert werden. Die Verbindung der Unterschiede mit den Schülerleistungen bzw. affektiven Variablen wird nur in Ansätzen diskutiert und muss an anderer Stelle erfolgen. Eine detaillierte Darstellung der Forschungsfragen des Dissertationsvorhabens findet sich in Kapitel 2.

1.2 Gliederung

Die vorliegende Untersuchung besteht aus vier Teilen, I Theoretische Grundlagen, II Methodik, III Ergebnisse sowie IV Diskussion und Ausblick. In jedem Teil wird ein spezifischer Aspekt der Untersuchung abgehandelt. In nachfolgender Übersicht sollen die Teile kurz beschrieben werden.

I Theoretische Grundlagen. In diesem Teil werden die Forschungsfragen vorgestellt sowie bisherige Erkenntnisse zum Experiment im Schulunterricht zusammengefasst. Die Zusammenfassung beginnt mit der fachdidaktischen und der

wissenschaftstheoretischen Dimension experimentellen Handelns. Daran anknüpfend, werden die mit experimentellem Handeln verbundenen Ziele und Vorstellungen präsentiert und diskutiert. Kontrastiert werden die Ziele und Vorstellungen durch eine Übersicht qualitativer oder quantitativer Studien, welche bisher Einblicke in die unterrichtliche Praxis experimentellen Handelns erlaubt haben. Als Resultat aus den erwähnten Befunden bisheriger Studien auf der einen und den Dimensionen experimentellen Handelns auf der anderen Seite werden drei bedeutsame Perspektiven experimentellen Handelns formuliert. Das auf den Perspektiven aufbauende Analyseinstrument wird im nachfolgenden Teil vorgestellt.

II Methodik. Gegenstand des zweiten Teils bildet die methodische Umsetzung der Untersuchung. Eingeleitet wird der Teil durch eine Übersicht über das Gesamtprojekt *Quality of Instruction in Physics*, dessen methodische Basis das Modell von Lipowsky et al. (2005) darstellt. Nachfolgend wird auf die Erhebung der Daten in den drei Ländern, im Speziellen auf die Besonderheiten der Erhebung in der Schweiz eingegangen. Die weiteren Abschnitte des Teils sind der Methodik bei der Analyse der Unterrichtsvideos gewidmet und umfassen die Analyse der Sichtstruktur, der Tiefenstruktur sowie die qualitativen Fallanalysen.

III Ergebnisse Dem methodischen Aufbau entsprechend, gliedert sich die Präsentation der Ergebnisse in die Analyse der Sichtstruktur und der Tiefenstruktur, wobei letztere auch die Fallanalyse mit einschliesst. Der Ergebnisteil wird mit der Analyse zur Frage, inwiefern unterrichtliche Merkmale experimentellen Handelns für die Leistungsentwicklung der Schülerinnen und Schüler bedeutsam sind, abgeschlossen.

IV Diskussion und Ausblick Die Ergebnisse des letzten Teils werden zusammengefasst, diskutiert und dabei mit den bisherigen Erkenntnissen der Forschung (Teil I) verbunden. Die Ableitung von Implikationen für die Unterrichtspraxis sowie Forschungsdesiderata schliessen die Untersuchung ab.

Teil I

Theoretische Grundlagen

Kapitel 2

Forschungsfragen

Ausgangspunkt des Projektes “Quality of Instruction in Physics” (QuIP) sind drei übergeordnete Forschungsfragen: 1) Welche Charakteristika des Physikunterrichts machen die Unterrichtsqualität aus? 2) Bestehen zwischen den Ländern bezüglich dieser Qualität Unterschiede? 3) Lassen sich die länderspezifischen Differenzen beim Zuwachs an Schülerleistung auf Unterschiede der Unterrichtsqualität zurückführen? Frage 1) bzw. Fragen 2) und 3) können dabei zwei grundsätzlichen Zielen zugeordnet werden: Die Identifikation von Qualitätsaspekten des Physikunterrichts sowie die Erklärung der Leistungsunterschiede¹ zwischen deutschen, finnischen und Schweizer Schülerinnen und Schüler auf der Ebene des Unterrichts.

Als Teil des Gesamtprojektes orientiert sich die vorliegende Dissertation an den erwähnten Fragen und Zielen: Im Grundsatz geht es dabei um die Entwicklung eines Qualitätsverständnisses experimentellen Handelns im Unterricht und der anschließenden Identifizierung der Qualität in der unterrichtlichen Praxis. Wie jedes Qualitätsverständnis ist auch dasjenige des experimentellen Handelns nicht ableitbar oder objektivierbar. Es ist vielmehr ein Produkt der Gewichtung verschiedener Einflüsse², die ihrerseits einem laufenden Wandel unterworfen sind. Eine Qualität experimentellen Handelns wird daran gemessen werden, inwiefern sie die relevanten Einflüsse einzubeziehen vermag und sich verträglich mit etablierten Vorstellungen der Unterrichtsqualität erweist. Den Anforderungen wird wie folgt Rechnung getragen: Die fachdidaktische Forschung zum Experiment im Physikunterricht wird aufgearbeitet (Forschungsfragen Abschnitt 2.1). Daraus werden Qualitätskriterien festgelegt. Diese

¹Vgl. PISA-Studien 2000-2009 (OECD, 2001, 2004, 2007, 2010)

²Z.B. aus der fachdidaktischen Forschung, der naturwissenschaftlichen Forschung, der Praxis der Lehrenden, u.a.m.

gehen in die Entwicklung eines Beurteilungsinstruments (Ratingsystem) ein, anhand welchem man Einheiten experimentellen Handelns bewertet (Abschnitt 2.3). Zuvor müssen die Einheiten experimentellen Handelns identifiziert werden (Abschnitt 2.2). Die empirischen Befunde zur Qualität experimentellen Handelns werden sodann mit dem Zuwachs an Schülerleistung in Relation gesetzt (Abschnitt 2.4).

2.1 Verständnis von der Qualität experimentellen Handelns

Die erste Forschungsfrage (im Folgenden mit Q1 bezeichnet) bezieht sich auf die Entwicklung eines Verständnisses von der Qualität experimentellen Handelns. Dieses ist im Hinblick auf die im Verlauf des Projektes vorgenommene Bewertung von Experimentiereinheiten von besonderer Bedeutung. Beitragen zur Entwicklung des Verständnisses soll einerseits die Sicht der Lehrenden auf das experimentelle Handeln, wobei einschlägige Forschungsergebnisse einbezogen werden. Auf der anderen Seite ergeben sich wichtige Ergebnisse zum Qualitätsverständnis experimentellen Handelns aus bisherigen Befunden der unterrichtlichen Praxis, sowie aus der theoretischen Auseinandersetzung mit dem Experiment als Lerngelegenheit.

Q1 Welche Qualitätsverständnisse experimentellen Handelns im Physikunterricht lassen sich in der Literatur identifizieren?

2.2 Sichtstruktur experimentellen Handelns

Forschungsfragen S1-S3 beziehen sich auf die Sichtstrukturanalyse experimentellen Handelns. Diese identifiziert Einheiten experimentellen Handelns und charakterisiert sie bezüglich wichtiger Eigenschaften, wie Dauer der Experimentiereinheit, Phasen experimentellen Handelns und Organisationsform (vgl. Abschnitt 5.1). Eine solche Analyse ist in zweierlei Hinsicht bedeutsam: Die Identifikation der Experimentiereinheiten im videographierten Unterricht bildet die Voraussetzung für die weiterführende Tiefenstrukturanalyse. Die Sichtstrukturanalyse liefert wichtige Grundinformationen der experimentellen Praxis, die in der Ergänzung zur qualitätsbezogenen Tiefenstrukturanalyse zur Beschreibung der Experimentiereinheiten beitragen.

- S1 Welche experimentellen Handlungsmuster lassen sich im videographierten Unterricht identifizieren?
- S2 Inwiefern bestehen zwischen den Ländern Unterschiede bei den experimentellen Handlungsmustern?
- S3 Inwiefern sind zwischen den verschiedenen Schultypen Unterschiede bei den experimentellen Handlungsmustern auszumachen?

2.3 Tiefenstruktur experimentellen Handelns

Forschungsfragen T1-T7 geben die Richtung für die Untersuchung der Tiefenstrukturmerkmale experimentellen Handelns vor. Dabei geht das im Rahmen des ersten Teils entwickelte theoretische Qualitätsverständnis in die hoch-inferente Analyse der Experimentiereinheiten ein (vgl. Abschnitte 3.6 und 5.2). Hierfür wird ein Ratinginstrument eingesetzt, welches auf den drei Perspektiven *Experimentelles Handeln als kontextorientierte Aktivität* (Kontextorientierung), *als reflexive Aktivität* (Reflexivität) und *als theoriegeleitete Aktivität* (Theorieleitung) basiert, wobei jede Perspektive durch mehrere Items (Facetten) beschrieben wird.

Die Forschungsfragen lassen sich vier Bereichen zuordnen: die Identifikation von Qualitätsmerkmalen und die Untersuchung der diesbezüglichen Länderspezifika auf Klassenebene (T1, T2), das Herstellen des Zusammenhangs zwischen der Bewertung einzelner Experimentiereinheiten und der summativen Bewertung aller Einheiten einer Klasse (Gesamtrating, vgl. Abschnitt 5.2.2) (T3), die deskriptive Analyse der Kategorien *Funktion* und *Offenheit* auf der Ebene einzelner Experimentiereinheiten (T4) sowie die Beschreibung einzelner Fälle zur Validierung des Ratinginstruments (vgl. Abschnitt 5.2) (T5-T7).

- T1 Findet sich die Qualität experimentellen Handelns, aufgespannt durch die drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung*, in der untersuchten Teilstichprobe wieder?
- T2 Lassen sich bezüglich der Perspektiven oder Facetten Unterschiede zwischen den Ländern nachweisen?
- T3 Wie verhalten sich die Ratings einzelner Experimentiereinheiten innerhalb einer Klasse zu ihrer summativen Beurteilung (Gesamtrating)?

- T4 Welche Merkmale der Kategorien *Funktion* und *Offenheit* lassen sich in den Experimentiereinheiten identifizieren?
- T5 Was zeichnet Experimentiereinheiten hoher Qualität aus? Welche Eigenschaften der Lehrperson sind damit verbunden?
- T6 Inwiefern decken sich die Befunde der Fallanalysen mit den Ergebnissen aus dem Ratingverfahren?
- T7 Welche neuen oder verfeinerten Kategorien ergeben sich aus den Fallanalysen?

2.4 Bedeutung experimentellen Handelns für den Zuwachs an Fachwissensleistung

Mit Forschungsfragen Z1 und Z2 werden die unterrichtlichen Merkmale aus der Sicht- und Tiefenstrukturanalyse auf den Zuwachs an Fachwissensleistung (Geller et al., in Vorb.) bezogen.

Es soll dabei die Vermutung von Tesch (2005) geprüft werden, dass die insgesamt im Kontext experimentellen Handelns stehende Unterrichtszeit positiv mit dem Zuwachs an Schülerleistung korreliert. Zudem wird untersucht, inwiefern die heuristisch abgeleiteten Perspektiven und deren Facetten relevant für die Entwicklung der Schülerleistung sind.

- Z1 Lässt sich zwischen der im Kontext experimentellen Handelns stehenden Unterrichtszeit und dem Leistungszuwachs ein Zusammenhang nachweisen?
- Z2 Lassen sich zwischen Qualitätsmerkmalen experimentellen Handelns und dem Leistungszuwachs Zusammenhänge nachweisen?

Kapitel 3

Experimentelles Handeln

Experimentelles Handeln im Physikunterricht ist bestimmt eines der meistbeachteten und am besten untersuchten Themenbereiche der Physikdidaktik. Wichtige Metaanalysen zum Thema finden sich in Harlen & Wake (1999), Hodson (1993), Hofstein & Lunetta (2003), Lunetta et al. (2007), Lunetta (1998), Jenkins (1999), Trumper (2003), Watson (2000), White (1996) u.a.m. Für einen umfassenden Überblick sei an dieser Stelle auf die genannte Literatur verwiesen.

Das vorliegende Theoriekapitel widmet sich der Frage, was die Qualität experimentellen Handelns im Physikunterricht ausmacht. Voraussetzung für die Entwicklung eines Qualitätsverständnisses ist dabei die Festlegung dessen, was unter experimentellem Handeln überhaupt verstanden werden soll. Es stellt sich heraus, dass der Begriff des experimentellen Handelns¹, trotz der erwähnten fachdidaktischen Bedeutung, mit einer grossen Unschärfe behaftet ist. Das mag damit zusammenhängen, dass mit experimentellem Handeln gleich mehrere unterrichtliche Dimensionen verbunden werden: Als wichtiger Teil im Methodenrepertoire des Physikunterrichts haben Experimente eine fachdidaktische Dimension. Es sind Lerngelegenheiten, die dazu beitragen können, physikalische Sachverhalte besser zu verstehen, Phänomene kennen zu lernen, den Austausch zwischen Lernenden zu fördern u.a.m. Eine Klassifikation experimentellen Handelns nach diesem fachdidaktischen Gesichtspunkt findet sich z.B. bei Kircher et al. (2001).

Experimente haben also eine ganz spezifisch unterrichtsmethodische Funktion. Die so gearteten Experimente sind meist keine im engeren Sinne wissenschaftlichen

¹Im englischen Sprachraum sind für das experimentelle Handeln im Schulunterricht die Termini ‘practical work’, ‘practical activity’, ‘experiment’, für Schülerexperimente (aber nicht konsequent auf diese beschränkt) die Begriffe ‘laboratory work’ und ‘laboratory activity’ gebräuchlich.

Experimente². Jedoch knüpfen solche Aktivitäten im Unterricht meist stark an Fachinhalte und Konzepte an. Das hat historisch-epistemologische Gründe: Fachinhalte und Konzepte der Physik sind untrennbar mit paradigmatischen experimentellen Anordnungen³ verwoben. Den Anordnungen kommt bei der Entwicklung physikalischer Erkenntnis eine Schlüsselfunktion⁴ zu. Hier zeigt sich die wissenschaftstheoretische Dimension experimentellen Handelns: Das Experiment bzw. die experimentelle Repräsentation ist ein Schlüsselement bei der Gewinnung physikalischer Erkenntnis. Auf diese Dimension reduziert, können Experimente Aufschluss geben darüber, welche Mechanismen bei der Erkenntnisgewinnung in den Naturwissenschaften bedeutsam sind. Man könnte in dem Zusammenhang auch von den Methoden der Naturwissenschaft sprechen. Es zeigt sich jedoch – und das ist bestimmt ein zweiter Grund für die Unschärfe des Begriffs Experiment bzw. experimentelles Handeln –, dass es bisher nicht gelungen ist zu beschreiben, welche diese Methoden sein könnten:

[There is] no well confirmed picture of how science works, no theory of science worthy of general assent. We did once have a well developed and historically influential philosophical position, that of positivism or logical empiricism, which has by now been effectively refuted. We have a number of recent theories of science which, while stimulating much interest, have hardly been tested at all. And we have specific hypotheses about various cognitive aspects of science, which are widely discussed but wholly undecided. If any extant position does provide a viable understanding of how science operates, we are far from being able to identify what it is. (Laudan et al., 1986, zitiert in Jenkins (1999))

Trotz der schwierigen Ausgangslage, die im Wesentlichen auch heute noch besteht, muss es das Ziel sein, die wissenschaftstheoretische Dimension bei den Überlegungen zur Qualität experimentellen Handelns mit einzubeziehen. Als anerkannter Teil der

²Hier ist ein Verständnis experimentellen Handelns angesprochen, das sich stark an den methodischen Belangen bei der Durchführung des Experiments und weniger an den damit verbundenen Denkprozessen orientiert. Beispielhaft dafür ist die Definition von Kircher et al. (2001, Hervorh. im Original): “Bei einem Experiment werden von einem Experimentator in einem realen System bewusst gesetzte und ausgewählte natürliche Bedingungen *verändert, kontrolliert und wiederholt beobachtet*.”

³Z.B. das Faden- oder Federpendel für das Konzept des harmonischen Oszillators, die Fallröhre für das Konzept des freien Falls, der Permanentmagnet mit Eisenfeilspänen für das Feldkonzept.

⁴Man denke z.B. an die Bedeutung des harmonischen Oszillators für die Entwicklung der Quantenmechanik.

scientific literacy wird ihr bei der Frage, was naturwissenschaftliche Bildung ausmacht, in vielen Ländern eine wachsende Bedeutung zuerkannt. So betrachtet wird das Wissen über experimentelles Handeln selbst zum Fachwissen.

Aufgrund der eben entwickelten Gedanken zum experimentellen Handeln strukturiert sich das vorliegende Kapitel wie folgt: Im Abschnitt 3.1 werden die Merkmale der fachdidaktischen Dimension experimentellen Handelns ausgeführt. Dazu gehört auch die Auseinandersetzung mit den beiden wichtigsten Organisationsformen experimentellen Handelns, dem Schüler- und dem Lehrerexperiment. Es folgt Abschnitt 3.2 zur wissenschaftstheoretischen Dimension experimentellen Handelns. Zentral wird hierbei die Rolle des Experiments zur Gewinnung von Erkenntnis sein sowie die Frage, welche Modelle damit assoziiert werden können. Die Auffassung einer Qualität experimentellen Handelns muss sich jedoch auch danach richten, welche Ziele und Vorstellungen die Akteure mit experimentellem Handeln verbinden. Ihnen wird im Abschnitt 3.3 nachgegangen. In Abschnitt 3.4 wird der Handlungsaspekt *Fragen und Untersuchen* des Schweizer Kompetenzmodells HarmoS vorgestellt und diskutiert. Einen Überblick über wichtige Forschungsarbeiten zur Erfassung experimentellen Handelns im Unterricht bietet Abschnitt 3.5. Ausführliche Erwähnung finden hier die Arbeiten von Tesch (2005), Tesch & Duit (2004), Tesch & Gerber (2005), die Experimente im Rahmen der IPN-Videostudie untersucht haben sowie von Abrahams & Millar (2008), in der Unterrichtsbeobachtungen und Interviews Aufschluss über die experimentelle Praxis geben.

Vor dem aufgezeigten Hintergrund wird ein Modell experimentellen Handelns für den Unterricht erarbeitet (Abschnitt 3.6), das als Grundlage für die Tiefenstrukturanalyse der im Rahmen der Gesamtstudie erhobenen Unterrichtsvideos dient.

Eine genauere Definition, was unter experimentellem Handeln im Schulunterricht verstanden werden soll, erfolgt in den Abschnitten 3.1 und 3.2. Trotzdem sei bereits an dieser Stelle auf verwandte bzw. gleichbedeutende Begriffe hingewiesen: Im deutschen Sprachraum sind die vorgängig verwendeten Begriffe Experiment bzw. experimentelles Handeln gebräuchlich. Im Folgenden werden weiterhin beide Termini verwendet, wobei der Terminus des experimentellen Handelns dann gegenüber dem Begriff Experiment vorgezogen wird, wenn die Aktivität betont werden soll. Auch im englischen Sprachraum kommt der Begriff ‘experiment’ vor. Eher noch gebräuchlicher ist der Begriff ‘practical work’. Damit intendiert man eine Abgrenzung zum wissenschaftlichen Experimentieren. Alle bisher genannten Begriffe werden so-

wohl für das Schüler- als auch für das Lehrerexperiment verwendet. Des Weiteren existieren, vorwiegend in der Chemiedidaktik die Termini ‘laboratory work’, ‘laboratory activities’ oder kurz ‘labwork’. Meist sind damit Schülerexperimente gemeint, je nach Kontext aber auch Lehrerexperimente.

3.1 Die fachdidaktische Dimension experimentellen Handelns

Experimentelles Handeln ist das charakteristische Merkmal des Physikunterrichts. Als Methode nimmt es im Unterricht die herausragende Stellung ein (Tesch, 2005). Begründen lässt sich diese jedoch kaum ausschliesslich aus unterrichtsmethodischen Überlegungen. Vielmehr hat der Einsatz von Experimenten im Unterricht eine lange Tradition. Waren in der Mitte des neunzehnten Jahrhundert noch fast ausschliesslich Lehrerexperimente im Einsatz, so ist bereits gegen Ende des Jahrhunderts ein erster Wechsel zu Schülerexperimenten zu verzeichnen (DeBoer, 1991). Die zunächst stark angeleiteten, formalen Schülerexperimente wurden im Zug der Reformpädagogik von offenen, aus heutiger Sicht radikal konstruktivistischen Schülerexperimenten abgelöst.

[...] the beginner not only may but *must be put absolutely* in the position of an original discoverer. (Armstrong, 1898, Hervorh. J.B.)

Die radikale Haltung wurde mit dem Aufkommen des Projektansatzes Anfang des zwanzigsten Jahrhunderts etwas gemildert (DeBoer, 1991). Mit dem Thompson-Report 1918 beginnt dann die (vorläufige) Abkehr vom Schülerexperiment (DeBoer, 1991). Dieses wird als Zeitverschwendung angesehen und der Einsatz von Lehrerexperimenten⁵ empfohlen. Die Kritik am Schülerexperiment nehmen andere wichtige Autoren auf und verstärken sie noch (vgl. Hodson, 1993, S.86). Die Debatte über den Einsatz von Schüler- bzw. Lehrerexperimenten bleibt bis heute bestehen, da es nicht gelungen ist, die Überlegenheit einer der beiden Methoden wissenschaftlich zu belegen.

Der geschichtliche Exkurs zeigt, dass experimentelles Handeln eigentlich zwei Methoden (Organisationsformen) umfasst: das Schüler- und das Lehrerexperiment. Er

⁵Die Begriffe Lehrerexperiment und Demonstrationsexperiment werden synonym verwendet.

zeigt aber auch, dass in der Geschichte des Physikunterrichts alternative Organisationsformen zum Experiment keine Rolle spielen. Dabei ist die Bedeutung des Experiments im Prozess der naturwissenschaftlichen Erkenntnisgewinnung umstritten. Kircher et al. (2001) rechtfertigen die herausragende Stellung des Schulexperiments vor dem Hintergrund einer wissenschaftstheoretischen Relativierung wie folgt:

- weil die Schulphysik sich vorwiegend mit der anschaulicheren klassischen Physik befasst,
- weil mit Schülerexperimenten eine Reihe relevanter Unterrichtsziele verknüpft sind,
- weil Experimente ein unverzichtbarer Bestandteil der physikalischen Methodologie sind,
- weil Experimente den Physikunterricht erlebnisreicher und zufriedenstellender machen können.

(Kircher et al., 2001, S.168)

Ohne im Detail auf die einzelnen Punkte einzugehen, zeigt sich: Die Bedeutung des Experiments im Unterricht wird vorrangig fachlich begründet.

Experimente werden vorrangig fachlich begründet. Überlegungen dazu, welche Bedeutung der Einsatz von Experimenten für das Lernen haben könnte, sind nicht zu finden. Hodson (1993), Jenkins (1999) u.a. weisen darauf hin, dass experimentelles Handeln sich in Bezug auf das Lernen der Schülerinnen und Schüler gegenüber anderen Methoden in keiner Weise besonders auszeichnet. Dennoch werden kaum alternative Methoden in Erwägung gezogen. Sie stossen gar bei vielen Lehrenden und Lernenden auf Widerstand.

They don't expect reading and discussion or drama and role play – they do expect Bunsen burners and practical work. They don't want to find out that science is not a set of facts, that theories change, and that science does not have all the answers – they want the security of a collection of truth which are indisputable. (Lakin & Wellington, 1991, zitiert in Hodson (1993))

Offenbar ist mit dem Experiment ein Verständnis der Naturwissenschaften verbunden. Eine Loslösung von der Dominanz des Experiments im Unterricht ist, wie obiges Zitat illustriert, mit dem Gefühl von Unsicherheit verbunden.

Aus der Sicht der fachdidaktischen Dimension experimentellen Handelns ergeben sich aus der bisherigen Diskussion zwei Forderungen: Experimente müssen als Lernangebote wahrgenommen und begründet werden. Experimentelles Handeln sowie die Physikalische Bildung als Ganzes umfassen ein breites inhaltliches Spektrum⁶ – dem Umstand ist durch ein angemessen vielfältiges Methodenrepertoire im Unterricht Rechnung zu tragen. Mit der Forderung verbunden ist einerseits die Prüfung von Alternativen zum Experiment als Unterrichtsmethode, andererseits die Anerkennung einer offenen Auslegung experimentellen Handelns (‘practical work’):

In my view, any learning method that requires learners to be *active*, rather than passive, accords with the belief that students learn best by direct experience and so could be described as ‘practical work’. (Hodson, 1993, Hervorh. im Original)

Hier bewertet Hodson die Aktivität der Lernenden als ein, für das experimentelle Handeln konstitutives Element. Dabei ergeben sich Parallelen zur aktiven Konstruktion von Wissen im Sinne eines moderaten Konstruktivismus (Labudde, 2000).

Für Labudde (2000) gehören Schüler- und Lehrerexperimente zu den konstruktivistischen Elementen des naturwissenschaftlichen Unterrichts. Sie sollen – sofern verschiedene Bedingungen erfüllt sind – “ein aktives Konstruieren von Wissen” ermöglichen. Begründet wird die Einschätzung durch Chancen, die mit dieser Methode assoziiert werden: Die Verbindung von Handlungen und Operationen, die Herstellung von Intersubjektivität sowie, bezogen auf die Lehrenden, als Element der Orientierung im Unterricht. Welche Bedingungen müssen nun erfüllt sein, damit Experimente im Unterricht als konstruktivistische Elemente angesehen werden? Bezogen auf das Schülerexperiment weist Labudde darauf hin, dass sowohl rezeptartige als auch sehr offene Schülerexperimente im Hinblick auf die aktive Konstruktion von Wissen problematisch sein können. Während rezeptartig organisierte Schülerexperimente ein passives Verhalten der Lernenden begünstigten, könnten offene Experimente zu einer Überforderung führen. Aus der Perspektive des moderaten Konstruktivismus, die Labudde vertritt, kommt es bei der Umsetzung von Schüler-

⁶Hodson (1993) teilt naturwissenschaftliche Bildung in drei Hauptaspekte ein: *learning science* – das Erarbeiten und Entwickeln von konzeptuellem und theoretischem Wissen, *learning about science* – das Entwickeln eines Verständnisses von den Methoden der Naturwissenschaft und das Bewusstsein über das komplexe Zusammenspiel zwischen Wissenschaft und Gesellschaft sowie *doing science* – das Durchführen von eigenen Untersuchungen zur Entwicklung von experimentellen Fähigkeiten und Fertigkeiten.

experimenten offenbar auf die Ausgewogenheit zwischen der Führung der Lehrperson und den Gestaltungsmöglichkeiten der Lernenden an. Das gilt eigentlich auch für das Lehrereperiment. Einblicke in die unterrichtliche Praxis zeigen allerdings, dass sowohl beim Schüler- als auch beim Lehrereperiment eine ausgeprägte Engführung des Experiments vorherrschend ist (vgl. Tesch (2005) und Abschnitt 3.5).

Als weiteres konstruktivistisches Element experimentellen Handelns kann die von Labudde hervorgehobene Intersubjektivität gelten und damit verbunden die Kommunikation: Experimente sind Anlässe für Kommunikation. Lernende tauschen sich über Beobachtungen und die damit verbundenen Vorstellungen aus. Ideen werden durch gegenseitige Zustimmung bestärkt oder durch Widerspruch heraus gefordert (Millar, 2004). Dennoch sollten solche Anlässe der Kommunikation in Bezug auf das Erlernen von Fachinhalten nicht überstrapaziert werden: Zu recht weisen Driver et al. (1994), Millar (2004) darauf hin, dass viele naturwissenschaftliche Erklärungen der Alltagserfahrung zuwider laufen (für Beispiele vgl. Muckenfuss, 2004, Röhneck, 1986). Hinter jedem Erkenntnisprozess verbergen sich Entwicklungsgeschichten, die im Unterricht kaum nachvollzogen werden können: “Scientific knowledge is more than personally-held belief reinforced by personally-gathered observational confirmation” (Hodson, 1996). Dies bedeutet, dass Lehrpersonen tragfähige Vorstellungen zu physikalischen Phänomenen oder Vorgängen anbieten und etablieren müssen. Sie können nicht erwarten, dass die wissenschaftlich gefestigten Vorstellungen von Schülerinnen und Schülern “wiederentdeckt” werden. Gleichzeitig sind die Lehrenden herausgefordert, Schülervorstellungen durch ein geschicktes *scaffolding* im Sinne einer aktiven Konstruktion von Wissen mit einzubeziehen.

Zur Frage, was die Qualität experimentellen Handelns im Physikunterricht ausmacht lässt sich aus Sicht der fachdidaktischen Dimension Folgendes zusammenfassen: Die Dominanz experimentellen Handelns im Unterricht ist historisch bedingt und lässt sich nicht durch die Überlegenheit der Methoden gegenüber anderen erklären. Die unklaren Befunde bezüglich der unterschiedlichen Lernwirksamkeit von Schüler- und Lehrereperiment lassen vermuten, dass die Methode bzw. die Organisationsform kein relevanter Indikator für die Unterrichtsqualität darstellt. Das geht auch aus Labudde (2000) hervor, der Experimenten nur unter der Erfüllung bestimmter Voraussetzungen die Funktion einer “aktiven Konstruktion von Wissen” zuschreibt. Offenbar sind es die Voraussetzungen, welche die Qualität experimentellen Handelns ausmachen. Als solche können aus bisherigen Überlegungen genannt

werden: der Einbezug der Lernendenperspektive bei der Gestaltung, die Ausgewogenheit zwischen Führung und Offenheit sowie die Nutzung von Experimenten als Kommunikationsanlässe.

3.2 Die wissenschaftstheoretische Dimension experimentellen Handelns

Experimentelles Handeln im Unterricht wird oft über seine wissenschaftstheoretische Dimension begründet. Für die Auseinandersetzung mit der Dimension innerhalb der fachdidaktischen Diskussion sprechen mehrere Argumente:

Die Einschätzung der erkenntnistheoretischen Stellung des Experiments wurde innerhalb der letzten 50 Jahre grundsätzlich revidiert (Feyerabend, 1976, 1983, Kuhn, 1962, 1970). Jedoch existiert jenseits der empiristischen, positivistischen Tradition⁷ keine klare Vorstellung davon, wie die Naturwissenschaften funktionieren und welche Rolle darin das Experiment einnimmt (siehe S. 12, wo auf Laudan et al. (1986) verwiesen wurde). Der Mangel an Orientierung auf der einen Seite und die immer noch weit verbreitete empiristische, positivistische Auffassung vieler Forschender auf der anderen Seite (Pomeroy, 2006) mögen dazu führen, dass auch im Unterricht die aus heutiger Sicht nicht mehr adäquaten Auffassungen über die Naturwissenschaften dominieren.

Das Experiment hat in den Naturwissenschaften eine grundsätzliche Wandlung vollzogen (Jenkins, 1999). Experimente werden in der Regel nicht mehr von einzelnen Wissenschaftlern durchgeführt. Nicht selten arbeiten Teams aus mehreren Forschergruppen an einem Experiment. Dabei hat die Komplexität der Experimente enorm zugenommen, so dass einzelne Wissenschaftler meist nur Teile davon wirklich überblicken können. Aus Gründen der Komplexität können Experimente jeweils nur von wenigen Institutionen auf der Welt nachvollzogen werden. Die hier skizzierte Veränderung der naturwissenschaftlichen Praxis sind sowohl aus gesellschaftlicher wie unterrichtlicher Perspektive von grosser Tragweite.

Die starke fachinhaltliche Verbundenheit des Experiments (Harlen & Wake, 1999, Hodson, 1993, Millar, 2004, u.a.) sowie der Anspruch, mit dem Experiment den Erkenntnisprozess der Naturwissenschaften im Unterricht zu illustrieren⁸, führen dazu,

⁷Für eine Definition dieser Traditionen sei auf Monk & Dillon (2000) verwiesen.

⁸Vgl. den Aspekt *learning about science* der naturwissenschaftlichen Bildung (Hodson, 1993)

dass experimentelles Handeln selbst zum Fachinhalt wird: So, wie der Begriff der *Energie* im Hinblick auf seine fachinhaltliche Bedeutung im Unterricht grosse Beachtung genießt, wäre es folgerichtig, dem Begriff *Experiment* eine mindestens ebensolche Aufmerksamkeit zukommen zu lassen. Aufgrund der eben erwähnten grossen Komplexität naturwissenschaftlicher Experimente kann diese jedoch nicht durch die Herstellung von Authentizität und der Entwicklung eines impliziten Verständnisses bei den Lernenden erreicht werden. Naheliegender ist die Explizierung von Teilspekten experimenteller Praxis.

Ausgehend von den erwähnten Argumenten, die für einen angemessenen Stellenwert des Experiments in Bezug auf den naturwissenschaftlichen Erkenntnisgewinn sprechen, werden im Folgenden verschiedene Modelle experimentellen Handelns vorgestellt.

Duit & Tesch (in Vorb.) berufen sich auf die Arbeit von Koponen & Mäntylä (2006), wenn sie feststellen, dass Experimente untrennbar mit theoretischen Vorstellungen verbunden sind. Sie verdeutlichen dies durch die in Abbildung 3.1 dargestellte wechselseitige Beziehung: “theory is a factor in experimental design” auf der einen Seite, “experimentation is a factor in theory construction” auf der anderen (van Fraassen, 1980, zitiert in Koponen & Mäntylä (2006)). Damit geben sie zwei aus insgesamt sechs bei Koponen & Mäntylä dargelegten wissenschaftstheoretischen Standpunkten ein besonderes Gewicht: *Observations are theory-laden and recognition of phenomena is guided by theory, experiments and theory are intertwined.*

Die didaktische Rekonstruktion der wissenschaftstheoretischen Aspekte experimentellen Handelns wird bei Koponen & Mäntylä (2006) noch weiter verfeinert. Dabei unterscheiden die Autoren unterschiedliche Niveaus der Konzeptualisierung. Auf dem ersten Niveau, jenem der *Qualitäten*, haben Experimente explorativen Charakter. Sie werden dazu eingesetzt, Phänomene kennen bzw. besser kennen zu lernen. In einem nächst höheren Niveau (*Quantitäten*) werden qualitativ gewonnene Zusammenhänge quantitativ untersucht, wobei bereits Hypothesen bestehen. Beispielsweise untersuchen die Lernenden im Rahmen eines Schülerexperiments den Zusammenhang zwischen elektrischer Spannung und Stromstärke. Schliesslich werden die so etablierten Gesetze im höchsten Niveau der Konzeptualisierung (*Theorie*) miteinander verbunden. Bezogen auf das Beispiel würde etwa das Verhalten eines einfachen Stromkreises mit einem Wasserkreislauf verglichen (Wasseranalogie des elektrischen Stromkreises). Beim Prozess der Erkenntnisgewinnung, so die Autoren,

werden unterschiedliche Niveaus der Konzeptualisierung durchlaufen, wobei jeweils vom höchsten Niveau, also dem Niveau der Theorie ausgegangen wird. Weiter wird unterschieden, ob experimentelles Handeln im Kontext der Entdeckung (context of discovery) oder dem Kontext der Bestätigung (context of justification) statt findet.

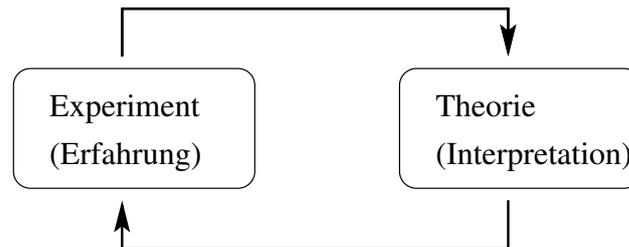


Abbildung 3.1: Verhältnis von Experiment und Theorie nach dem Verständnis von Duit & Tesch (in Vorb.).

Auch Millar (2004) argumentiert aus der Perspektive der Wissenschaftstheorie, wenn er den *process approach*⁹ kritisiert, welcher in den 1960er Jahren durch das Nuffield Science Project initiiert wurde und bis heute Anklang findet: “‘Discovery learning’ is based on a empiricist view of science and an inductive view of the ‘scientific method’”. Die Anziehung, die von der Methode ausgeht, ist für Millar mit dem Versprechen selbstgesteuerten Lernens verbunden. Das Versprechen lässt sich scheinbar ideal mit einem konstruktivistisch¹⁰ orientierten Unterricht verbinden.

In seiner Kritik verfolgt Millar zwei Argumente: Der *process approach* suggeriere, dass die Methoden der Naturwissenschaften erlernt werden könnten und dass sie konstituierend für naturwissenschaftliches Arbeiten seien. Dabei, so Millar (1989), seien viele der so genannten Prozesse nicht spezifisch naturwissenschaftlich und liessen sich auf allgemeine kognitive Fähigkeiten zurückführen, die alle Menschen im Laufe ihres Lebens entwickelten. Der Autor bezweifelt, dass solche allgemeinen kognitiven Fähigkeiten durch den Prozessansatz gezielt gefördert werden können. Zudem – so das zweite Argument – könne es nicht gelingen, die Lernenden in die Rolle des Wissenschaftlers (*pupil as a scientist*) zu versetzen, wie das der Prozessansatz vorsieht, denn anerkannte Ideen und Theorien ergäben sich nicht einfach aus den Messungen.

⁹Synonym zu dem Begriff werden die Termini *inquiry-based approach* und *discovery learning* verwendet.

¹⁰Der konstruktivistische Ansatz als wichtiges Element der fachdidaktischen Dimension weist “bedeutende Parallelen” (Labudde, 2000) zu wissenschaftstheoretischen Überlegungen und damit zur wissenschaftstheoretischen Dimension auf. Eine strikte Trennung der beiden Dimensionen lässt sich somit nicht vollziehen.

Eher sei dazu ein Zusammenspiel vieler Faktoren notwendig, von denen das Experiment und die daraus gewonnenen Ergebnisse lediglich einen Teil ausmachten. Mit Verweis auf die bedeutende Rolle der Kreativität beim Erkenntnisprozess hält Nersessian (1991, zitiert in Ntombela (1999)) den Terminus ‘discovery’ für irreführend. Adäquater umschreibe der Begriff ‘invention’ den erwähnten Prozess.

In Anlehnung an ein Modell von Giere (1991) zur Beschreibung wissenschaftlichen Argumentierens schlagen Millar et al. (1999) die Unterscheidung zweier Wissensdomänen vor (vgl. Abbildung 3.2 sowie Millar et al. (2002), Millar (2004, 2009), wo das Modell weiter ausgearbeitet und angewandt wird): die Domäne der Objekte (*domain of objects and observable things*) und die Domäne der Ideen (*domain of ideas*). Experimentelles Handeln¹¹ im Unterricht, so Millar (2004), entstehe durch die Verbindung der beiden Domänen:

The role of practical work, then, in the teaching and learning of science content is to help students make links between two ‘domains’ of knowledge: the domain of objects and observable properties and events on the one hand, and the domain of ideas on the other. (Millar, 2004)

Die Modelle in Abbildungen 3.1 und 3.2 scheinen praktisch identisch zu sein. Dennoch finden sich interessante Unterschiede: Während beim Modell von Duit & Tesch die Interaktion zwischen dem Experiment und der Theorie hervorgehoben wird, ist bei Millar et al. die Interaktion selbst kennzeichnend für experimentelles Handeln. An die Stelle der typischen Gegensätzlichkeit zwischen Theorie und Experiment tritt bei Millar et al. die Gegensätzlichkeit der beiden Wissensdomänen.

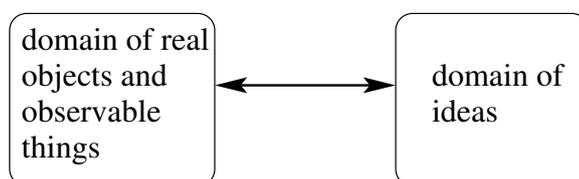


Abbildung 3.2: Experimentelles Handeln im Unterricht als Verbindung der Wissensdomänen “domain of objects and observable things” und “domain of ideas” (Millar et al., 2002).

¹¹Millar et al. (2002) verwenden die Begriffe *labwork* und *practical work* als Sammelbegriffe für experimentelles Handeln im Unterricht. Sie stehen für Schüler- sowohl als auch Lehrerexperimente. Millar (2004) zieht hingegen den Terminus ‘practical work’ vor: “[...] practical work is used in preference to ‘laboratory work’ because location is not a salient feature in characterising this kind of activity.”

Zur Frage, was die Qualität experimentellen Handelns im Physikunterricht ausmacht, lässt sich aus den Überlegungen zur wissenschaftstheoretischen Dimension also Folgendes zusammenfassen: Schulexperimente sollten danach beurteilt werden, inwiefern sie dazu beitragen, relevante Aspekte des Erkenntnisprozesses zu illustrieren. Als besonders wichtig wird in dem Zusammenhang die Verbindung zwischen den physischen und mentalen Handlungen erachtet (Duit & Tesch, in Vorb., Millar et al., 2002). Experimente sollten immer mit Ideen, Modellen und Theorien verknüpft werden.

3.3 Ziele und Vorstellungen von Lehrpersonen

Welche Erwartungen und Vorstellungen verbinden Lehrpersonen mit dem experimentellen Handeln? Im Rahmen eines Internationalen Projektes zur Erforschung der Rolle des Experiments (Psillos & Niedderer, 2003) wurde diese Frage mit einer Delphi-Studie untersucht (Welzel et al., 1998). An der Befragung nahmen 400 Lehrende aller naturwissenschaftlichen Disziplinen und Ausbildungsstufen aus sechs europäischen Ländern teil. Als Resultat der ersten Runde mit offenen Fragen wurden fünf übergeordnete Ziele identifiziert (vgl. Tabelle 3.1, linke Spalte). Darauf aufbauend entwickelte man für die nachfolgende Runde ein Instrument mit geschlossenem Antwortformat.

Swain et al. (2000) untersuchten die gleiche Frage mit Hilfe eines auf Kerr (1964) basierenden Instruments. Die Stichprobe umfasste 66 Lehrpersonen naturwissenschaftlicher Fächer aus 58 verschiedenen Schulen in Grossbritannien. Über die Frage nach den Zielen des Experimentierens hinaus wurde untersucht, inwiefern sich diese im Laufe der Zeit bei den Lehrpersonen verändert hatten. Dazu bezog man die Daten von Beatty & Woolnough (1982) und Kerr (1964) ein. Die bei Swain et al. (2000) und Beatty & Woolnough (1982) herausgearbeiteten wichtigsten vier Ziele sind mit jenen von Welzel et al. (1998) vergleichbar (Tabelle 3.1, rechte Spalte).

Zwischen den Ergebnissen von Swain et al., Beatty & Woolnough und jenen von Welzel et al. lassen sich jedoch auch interessante Unterschiede erkennen, die hauptsächlich auf die Entwicklung des angelsächsischen Bildungssystems zurückzuführen sind (vgl. Abschnitt 3.2, *process approach*): Währenddem der Aspekt der Theorie bei Welzel et al. gleich an erster Stelle erscheint, tritt er bei Swain et al., Beatty & Woolnough unter den wichtigsten fünf Zielen nicht auf. Hingegen findet sich bei

Ziele experimentellen Handelns	
Welzel et al. (1998)	Swain et al. (2000), Beatty & Woolnough (1982)
1 Theorie und Praxis miteinander verbinden	I to encourage accurate observation and description
2 Experimentelle Fähigkeiten erwerben	II to promote a logical reasoning method of thought
3 Methoden wissenschaftlichen Denkens kennen lernen	III to arouse and maintain interest
4 Motivation/Soziales	IV to make phenomena more real
5 Wissen überprüfen	

Tabelle 3.1: Ziele experimentellen Handelns in absteigender Priorität als Ergebnis der Studien von Welzel et al. (1998) (links) und Swain et al. (2000), Beatty & Woolnough (1982) (rechts).

beiden Untersuchungen der Aspekt der Methodik experimentellen Handelns¹², des Kennenlernens bzw. des Förderns von wissenschaftlichem Denken und der Motivation. Generell lässt sich aus den Ergebnissen ablesen, dass Lehrende in Grossbritannien dem Experimentieren als Methode zum Erlangen wissenschaftlicher Erkenntnis eine grössere Bedeutung beimessen als ihre Kolleginnen und Kollegen aus der Studie von Welzel et al. (1998). Demgegenüber gewichten Lehrpersonen aus der Studie von Welzel et al. den Fachwissensaspekt stärker (Tabelle 3.1, linke Spalte, Ziele 1 und 5).

Die erwähnten Ziele der Lehrpersonen sind eng mit deren Vorstellungen über experimentelles Handeln verbunden. Jonas-Ahrend (2004) untersuchte die Vorstellungen im Rahmen einer internationalen Studie. Dabei wurden Lehrervorstellungen¹³ mit einem Fragebogen erhoben. Ergänzend fanden Interviews mit deutschen Lehrpersonen statt. Jonas-Ahrend findet, dass die Vorstellungen über die Rolle des Experiments im Physikunterricht weit gefächert seien: So erwähnen Lehrpersonen auch historische, gesellschaftliche, didaktische, psychologische und organisatorische Aspekte des Experiments. Die Autorin hebt hervor, dass meist fachliche und unterrichtsmethodische Aspekte vor psychologischen, erkenntnis- und wissenschaftstheoretischen Aspekten genannt werden. Für die Lehrenden sei das Experiment ein unverzichtbarer Teil des Unterrichts. Gleichzeitig bekundeten sie Mühe zu begrün-

¹²Die bei Welzel et al. erwähnten experimentellen Fähigkeiten beziehen sich hauptsächlich auf methodische Ziele des Experimentierens

¹³Jonas-Ahrend befragte Lehrende aus den GUS-Staaten, den USA und Deutschland.

den, worin die Bedeutung des Experiments für den Unterricht liege. Die Unsicherheit werde auch deutlich bei der Frage, welchen Beitrag Experimente zum Erkenntnisgewinn der Schülerinnen und Schülern leisten können. Es stelle sich heraus, dass Lehrpersonen Experimente fast ausschliesslich aus ihrer Sicht als Lehrende beurteilten und dabei übersähen, dass Schülerinnen und Schüler eine andere Sicht auf Experimente haben könnten als sie selbst. Experimente würden von Lehrpersonen als wichtiges Hilfsmittel zur Planung und Strukturierung des Unterrichts angesehen.

Es ist anzunehmen, dass sich die geäusserten Vorstellungen und Ziele zum Experiment als Produkt vieler Einflüsse ergeben, wie z.B. der wissenschaftlichen und unterrichtlichen Tradition und Sozialisation, didaktisch-pädagogischen Beliefs sowie Commonsense-Vorstellungen über die Naturwissenschaften. Auch wenn sich aus dem Verständnis und den Zielen experimentellen Handelns Parallelen zur unterrichtlichen Praxis ergeben, ist nicht von einem unmittelbaren Zusammenhang auszugehen. Das zeigen auch die empirischen Befunde zur Unterrichtspraxis (vgl. Abschnitt 3.5 und Tesch (2005), Tiberghien et al. (2001), Abrahams & Millar (2008)).

Einige der Vorstellungen und Ziele experimentellen Handelns wurden in der fachdidaktischen Forschung aufgenommen und zusammen mit Befunden aus der unterrichtlichen Praxis reflektiert. Auf der Grundlage empirischer Untersuchungen sowie der fachdidaktischen Diskussionen der letzten Jahrzehnte fassen Hofstein & Lunetta (2003), Hodson (1993) die in Tabelle 3.2 aufgeführten Ziele und Vorstellungen zum experimentellen Handeln zusammen. Der Vergleich der Ergebnisse von Hofstein & Lunetta und Hodson zeigt, von den unterschiedlichen Formulierungen abgesehen, überwiegende Einigkeit. Auch im Vergleich mit Tabelle 3.1 ergibt sich weitgehende Übereinstimmung. Offenbar gelangen Forschende und Lehrende zu ganz ähnlichen Vorstellungen und Zielen experimentellen Handelns.

Es fällt auf, dass die mit experimentellem Handeln verbundenen Ziele ambitioniert und umfassend sind. Sie decken sich in wesentlichen Teilen mit den Zielen der naturwissenschaftlichen Bildung insgesamt und lassen sich den Aspekten *learning science*, *learning about science* und *doing science* zuordnen (Hodson, 1993). Darüber hinaus gehören auch affektive Ziele, z.B. das Wecken von Interesse, dazu.

Die dadurch implizierte grosse Bedeutung experimentellen Handelns passt insofern zu Befunden der unterrichtlichen Praxis, als die für die Aktivitäten eingesetzte Zeit mehr als die Hälfte der gesamten Unterrichtsdauer ausmacht (Tesch (2005), vgl. Abschnitt 3.5). Hingegen zeigt sich, dass viele der erwähnten Ziele durch die

Ziele experimentellen Handelns	
Hofstein & Lunetta (2003)	Hodson (1993)
1 Understanding of scientific concepts	I To motivate by stimulating interest and enjoyment.
2 Interest and motivation	II To teach laboratory skills.
3 Scientific practical skills and problem solving abilities	III To enhance the learning of scientific knowledge.
4 Scientific habits of mind (more recent)	IV To give insight into scientific method and to develop expertise in using it.
5 Understanding of the nature of science (more recent)	V To develop certain 'scientific attitudes', such as open-mindedness, objectivity and willingness to suspend judgement.

Tabelle 3.2: Ziele experimentellen Handelns aus der Sicht von Hofstein & Lunetta (2003) (links) und Hodson (1993) (rechts).

Aktivitäten kaum erreicht werden (Tesch, 2005, Abrahams & Millar, 2008), wobei die Gründe dafür unklar sind.

Educationists need to consider the situation carefully: is our faith in laboratories misplaced and our theory about their value incorrect, or is practice poor? It is easier, and causes less emotional upheaval to theorists, to criticise and suggest improvements to practice than to change theory. (White, 1996)

Im Sinne der Aufforderung von White (1996), hat Hodson (1993) Ziele und Vorstellungen experimentellen Handelns vor dem Hintergrund einer *scientific literacy* kritisch hinterfragt. Darin kommt er zum Schluss, dass experimentelles Handeln nicht generell als motivierend erlebt wird (vgl. Tabelle 3.2, Ziel 2 bzw. I). Vielmehr sind es auch hier verschiedene Einflussfaktoren, welche sich auf das Erleben der Aktivität auswirken: Erwähnt werden die Übernahme von Verantwortung für und die Vertrautheit in die Aktivität. Während in den unteren Jahrgangsstufen noch offene Experimente vorkämen, dominierten bei den oberen eher geschlossene, rezeptartige Experimente mit entsprechend negativen Auswirkungen auf die Motivation. Ähnlich wie bei der Motivation verhalte es sich mit dem Ziel, fachinhaltliches Lernen durch Experimente zu unterstützen bzw. die Methoden der Naturwissenschaft kennen zu lernen und sie anzuwenden (vgl. Tabelle 3.2, Ziele 1 & 5 bzw. III & IV):

It seems that all that can be safely concluded from the several research studies [...] is that *some* teachers are able to use practical work successfully, with *some* students, to achieve *some* of their goals. (Hodson, 1993, Hervorh. im Original)

Für die unbefriedigenden Ergebnisse machen Hofstein & Lunetta (2003) verschiedene forschungsmethodischen Schwächen verantwortlich, auf die im Folgenden jedoch nicht weiter eingegangen wird.

Deutlicher wird Hodson in seiner Kritik am Ziel, experimentelle Fertigkeiten (laboratory skills) (vgl. Tabelle 3.2, Ziel 3 bzw. II) zu erlernen. Legitimiert werden die Fertigkeiten einerseits durch die Annahme der Übertragbarkeit auf andere Situationen und Gebiete, andererseits als eine Propädeutik der naturwissenschaftlichen Disziplinen. Während Hodson grundsätzliche Zweifel an der Übertragbarkeit bekundet, zumal sich eine solche nur schwer überprüfen liesse, hält er das Argument der naturwissenschaftlichen Propädeutik für “[...] ethically dubious [...] and hopelessly over-ambitious [...]”. Aus seiner Perspektive kann somit das Ziel der Förderung experimenteller Fertigkeiten nicht aufrechterhalten werden. Umso wichtiger ist es abzuwägen, inwiefern spezifische Fertigkeiten zur Erreichung inhaltlicher Ziele wichtig sind, oder wie es Harlen & Wake (1999) ausdrücken: “Practical work should be seen as means to various ends and not as an end itself”.

Kritisch hinterfragt werden muss auch das Ziel, wissenschaftliche Einstellungen (*scientific attitudes* bzw. *habits of mind*¹⁴, vgl. Tabelle 3.2, Ziel 4 bzw. V) zu fördern. Einerseits zeigt sich, dass sich Naturwissenschaftler bezüglich der Einstellungen nicht grundsätzlich von anderen Berufsgruppen unterscheiden (Wong & Hodson, 2009, Campanario, 2002). Somit sind Offenheit, Objektivität und das Bilden eines unvoreingenommenen Urteils, die gewöhnlich zu diesen Einstellungen gezählt werden, nicht allein den wissenschaftlichen Arbeitsweisen zuzuordnen. Vielmehr gehören sie zu den allgemeinen kognitiven Fähigkeiten, von denen unklar ist, ob sie sich überhaupt gezielt fördern lassen (Millar, 1989).

Wie soeben ausgeführt, sind nicht alle Ziele und Vorstellungen, die mit experimentellem Handeln verbunden werden, gleichermaßen angemessen. Auf der anderen Seite sind diese Ziele nur unter bestimmten Voraussetzungen erreichbar. Sie werden nicht *automatisch* durch Experimente befördert. Harlen & Wake (1999) formulieren

¹⁴Der Begriff *scientific habits of mind* geht zurück auf John Dewey. Er ist mit den *scientific attitudes* verwandt.

auf der Basis ihrer kritischen Reflexion eigene Ziele, die den Potentialen experimentellen Handelns besser entsprechen sollen (als die in Tabelle 3.2 zusammengestellten):

- *Providing first-hand experience*, so that pupils can ‘see it for themselves’ and in some cases do it themselves, although this purpose is often best served by a good demonstration or a field trip rather than ‘hands on’ practical work.
- *Testing ideas* by making predictions, setting up a valid test, collecting reliable evidence and relating what is found to the original idea. This practical work should be theory-based.
- *Experience of ‘doing science’* through carrying out an investigation which has a degree of open-endedness.

(Harlen & Wake, 1999, Hervorh. J.B.)

3.4 Experimentelles Handeln als Kompetenz

Die grosse Bedeutung, die experimentellem Handeln beigemessen wird, schlägt sich auch bei der Konzeption von Kompetenzmodellen nieder, wie die Entwicklungen in Deutschland und der Schweiz zeigen (KMK, 2004, Hammann et al., 2007, Wissenschaftliches Konsortium HarmoS, 2008, EDK, 2011, Hammann et al., 2007). Eine ausführliche Diskussion dieser und weiterer Kompetenzmodelle würde den Rahmen der Arbeit sprengen. Für eine Einführung in den Kompetenzbegriff sei z.B. auf die grundlegende Arbeit von Weinert (2001) verwiesen. Im Folgenden wird nur auf den Handlungsaspekt experimentellen Handelns im Schweizer Kompetenzmodell genauer eingegangen. Für das Ende der Sekundarstufe I lautet der mit *Fragen und untersuchen* (FU) bezeichnete Aspekt wie folgt:

Die Schülerinnen und Schüler können

- 1 *Situationen und Phänomene mit mehreren Sinnen wahrnehmen, beobachten und beschreiben und dazu verschiedenartige Fragen, Problemstellungen und einfache Hypothesen formulieren sowie Variablen für deren Überprüfung bestimmen* (insbesondere im Zusammenhang mit Kraft und Gegenkraft, Energieerhaltung und -umwandlung, Stromkreisen, Stoffumwandlungen, dem Aufbau von Zellen, dem Verhalten von Tieren, Biodiversität);

- 2 *angeleitet Erkundungen, Untersuchungen und Experimente planen, durchführen und dabei gezielt Schätzungen und Messungen vornehmen, Daten sammeln und auswerten und dabei zu Fragen und Hypothesen sachgemäss Stellung nehmen* (insbesondere Kraftumwandlungen, mechanische und elektrische Leistung, chemische Reaktionen, Körperfunktionen, Bestimmung von Tieren und Pflanzen in Lebensräumen);
- 3 *beim Erkunden, Untersuchen und Experimentieren sowie beim technischen Konstruieren geeignete Werkzeuge, Instrumente und Materialien auswählen und einsetzen* (insbesondere Instrumente zum Messen von Kraft, Stromstärke und Spannung; Instrumente zum Beobachten wie Mikroskop und Stereolupe);
- 4 *Ergebnisse aus Erkundungen, Untersuchungen und Experimenten in verschiedenen Formen darstellen* (insbesondere als Skizze, Bericht, Protokoll, Tabelle, Diagramm, Graph, Plan);
- 5 *Erfolge und Mängel ihrer Planung, Durchführung und Auswertung einschätzen und Verbesserungen vorschlagen.*

(EDK, 2011, S.33)

Die fünf Teilaspekte umfassen wichtige Phasen experimentellen Handelns. Sie orientieren sich sowohl an der mentalen als auch der physischen Aktivität der Lernenden. Mit Ausnahme von Teilaspekt 2 sind ganz unterschiedliche Organisationsformen experimentellen Handelns denkbar, d.h. je nach Inhalt können sich Schüler- oder Lehrerexperimente zur Förderung der Teilaspekte eignen.

Interessant ist der Umstand, dass auch Fachinhalte Eingang in die Teilaspekte gefunden haben, die konsequenterweise den Themenbereichen zugeordnet werden müssten. Der konzeptionelle Mangel offenbart hier eine grundsätzliche Eigenschaft experimentellen Handelns. Experimentelles Handeln ist eine holistische Aktivität (Hodson, 1993, Millar, 2004), die nur in einer engen Verbindung zu den Inhalten überhaupt sinnvoll ist. Das wird auch in EDK (2011) anerkannt: "Kompetenzen zeigen sich in der Verbindung von Handlungsaspekten und Themenbereichen." Bemerkenswert erscheint auch die Tatsache, dass von den insgesamt fünf Teilaspekten, die *Fragen und untersuchen* ausmachen, drei (1, 4, 5) vorwiegend mentale Handlungen betreffen. In der Hinsicht wird der kognitive Aspekt dieser Aktivität angemessen

berücksichtigt. Allerdings lässt die Formulierung der Teilaspekte vermuten, dass die Herausforderungen im Hinblick auf die mentalen Handlungen noch zu wenig in die Ausgestaltung des Handlungsaspekts eingegangen sind: Während bei Teilaspekt 2 die Rede von angeleiteten Erkundungen ist, fehlt eine analoge Formulierung beispielsweise beim Formulieren von Hypothesen oder beim Beobachten (Teilaspekt 1). Die Unterschätzung mentaler Handlungen im Zusammenhang mit dem Experiment entspricht dabei, wie in nächstem Abschnitt noch ausgeführt wird, der gängigen unterrichtlichen Praxis.

3.5 Befunde der unterrichtlichen Praxis

3.5.1 Videostudien

Mit der videobasierten Unterrichtsforschung ist ein bestimmtes Methodenrepertoire verbunden, dessen Hauptmerkmal die wissenschaftliche Videographierung des Unterrichts ist. Ohne auf Details einzugehen, seien an dieser Stelle charakteristische Elemente des Methodenrepertoires genannt (eine detaillierte Beschreibung findet sich in Seidel et al. (2005)): Das Unterrichtsgeschehen wird meist durch zwei Videokameras aufgezeichnet. Eine Kamera ist dabei auf die Lehrperson gerichtet (Aktionskamera), die andere fest auf die Klasse und steht typischerweise neben der Wandtafel (Totalenkamera).

Der so aufgezeichnete Unterricht wird anhand bestehender Kategoriensysteme analysiert bzw. kodiert. Man unterscheidet hierbei die Analyse auf der Basis eines Zeitstichprobenplanes (zeit-basierte Kodierung) und die Analyse auf der Basis von Vorkommnissen (event-basierte Kodierung). Im ersten Fall wird der videographierte Unterricht in Intervalle fester Dauer unterteilt. Dabei werden die Kategorien für jedes Intervall neu zugeordnet (kodiert). Im zweiten Fall sind die Einheiten der Analyse nicht feste Intervalle, sondern ergeben sich aus dem Kontext des videographierten Unterrichts. Einheiten sind dabei durch Vorkommnisse wie z.B. eine Frage der Lehrperson ans Plenum oder die Äusserung einer Schülerin/eines Schülers definiert. Den so festgelegten Zeitintervallen ordnet man wiederum Kategorien zu. Kategoriensysteme werden entsprechend der für die Beurteilung nötigen Interpretationsleistung als niedrig bzw. hoch-inferent charakterisiert.

Erste gross angelegte Projekte videobasierter Unterrichtsforschung (Videostudien), welche der skizzierten Methodik folgen, sind TIMSS 1995 Video und die Nachfolgestudie TIMSS 1999 (Stigler et al., 2000, Hiebert et al., 2003) für den Mathematikunterricht. Der Physikunterricht wurde erstmals im Rahmen der IPN Videostudie in Deutschland (Seidel et al., 2005) und der Parallelvideostudie Lehr-Lernkultur Physik in der Schweiz (Gerber, 2007, Knierim, 2008) untersucht. Als Teil der Studien entwickelte man ein Kategoriensystem zur Erfassung experimentellen Handelns im Unterricht (Tesch, 2005, Tesch & Duit, 2004, Tesch & Gerber, 2005), worauf im Folgenden noch ausführlicher eingegangen wird. Schliesslich gibt die CPV Videostudie Physik Einblick in den tschechischen Physikunterricht und erlaubt den Vergleich mit Ergebnissen von TIMSS 1995 Video und der IPN Videostudie (Janik et al., 2006).

Ein Kategoriensystem zur Analyse experimentellen Handelns im Physikunterricht wurde von Tesch (2005) im Rahmen der IPN Videostudie in einer ersten Projektphase entwickelt und in einer zweiten Phase in Deutschland und der Schweiz eingesetzt (Tesch & Gerber, 2005). Der Autor der vorliegenden Arbeit hat keine Kenntnis von weiteren Kategoriensystemen zur videobasierten Analyse experimentellen Handelns. Im Folgenden werden die beiden Projektphasen, die Entwicklung und der Einsatz des Instruments, beschrieben.

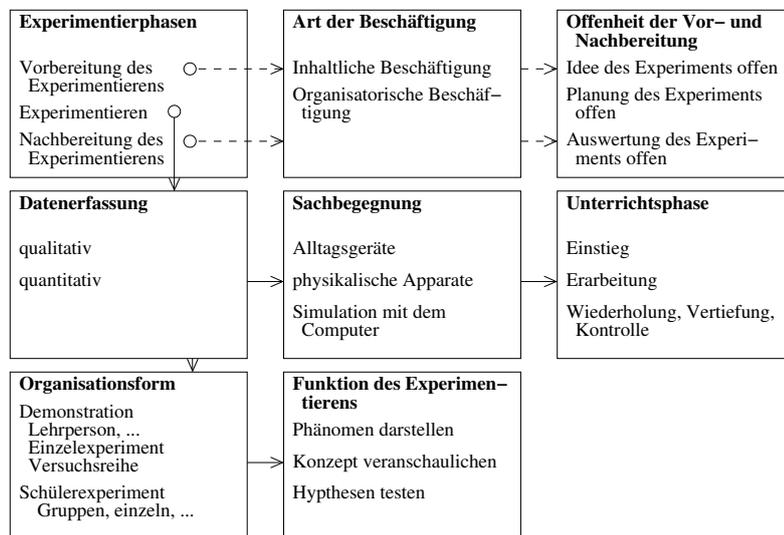


Abbildung 3.3: Das Kategoriensystem von Tesch (2005) (gekürzte Fassung).

Während der ersten Phase wurde ein Kategoriensystem experimentellen Handelns entwickelt (Tesch, 2005). Als Ausgangspunkt dienten die Funktionen experimentellen Handelns von Kircher et al. (2001). In einem ersten Analysedurchgang werden

hierbei die identifizierten Zeitintervalle den Phasen *Vorbereitung des Experimentierens*, *Experimentieren*, *Nachbereitung des Experimentierens* oder *keine* (Experimentierphase) zugeordnet (vgl. Abbildung 3.3, *Experimentierphasen*).

Im nachfolgenden Analysedurchgang werden den Zeitintervallen phasenspezifische Unterkategorien zugeordnet. Eine Gruppe von Unterkategorien betrifft die Gestaltung der Vor- und Nachbereitungsphasen. Hierbei kann der Grad der Offenheit sowie die Art der Beschäftigung – inhaltlich oder organisatorisch – kodiert werden (Abbildung 3.3, *Offenheit der Vor- und Nachbereitung* und *Art der Beschäftigung*). Eine weitere Kategoriengruppe dient zur Klassifikation der Phase *Experimentieren* (*Datenerfassung*, *Sachbegegnung*, *Unterrichtsphase* und *Organisationsform*, Abbildung 3.3). Dabei ist ein Teil der Funktionen von Kircher et al. (2001) in der Facette *Funktion des Experimentierens* integriert.

Das so weit beschriebene System wurde für die zeitbasierte Kodierung (10s-Intervall) vorgesehen. Eine Pilotierung fand im Rahmen der ersten Projektphase mit 13 Lehrpersonen (344 Schülerinnen und Schülern des 7. und 8. Schuljahres) aus Schleswig-Holstein statt¹⁵ (Tesch, 2005, Tesch & Duit, 2004). Dabei wurden pro Lehrperson zwei dreistündige Unterrichtssequenzen zur Einführung in den elektrischen Stromkreis und den Kraftbegriff aufgezeichnet. Zu Beginn und Ende des Schuljahres fanden ergänzend Fragebogenerhebungen statt. Hierbei erfasste man affektive und leistungsbezogene Variablen.

Die Ergebnisse der ersten Phase wurden in Tesch & Duit (2004), Tesch (2005) ausgewiesen. Sie deuten darauf hin, dass experimentelles Handeln bzw. das Experiment eine bedeutende Rolle im untersuchten Physikunterricht einnimmt: Während die eigentlichen Handlungsphasen der Experimente (*Experimentieren*) nur 28% der Unterrichtszeit ausmachen, steht immerhin 64% im Kontext zu einem Experiment. Allerdings findet Tesch grosse Unterschiede zwischen den unterrichteten Fachgebieten: Während die Einführung in den elektrischen Stromkreis üblicherweise mit längeren Schülerexperimentierphasen verbunden ist, kommen beim Kraftbegriff mehrere kurze Demonstrationsexperimente, oft unter Einsatz von Alltagsmaterialien, vor.

Die Verbindung der Ergebnisse aus der Videoanalyse mit den begleitenden Fragebogenerhebungen lassen einen positiven Zusammenhang zwischen dem Leistungszuwachs und der Gesamtdauer des Experiments erahnen (Tesch, 2005). Hingegen

¹⁵Die Schulen bzw. Lehrpersonen wurden nicht zufällig ausgewählt. Sie nahmen am SINUS-Modellversuch teil (SINUS: Steigerung der Effizienz des mathematischen-naturwissenschaftlichen Unterrichts).

konnte zwischen der blossen Dauer der Handlungsphasen der Experimente (Phasen *Experimentieren*) und der Leistungsentwicklung kein Zusammenhang festgestellt werden. Dies interpretiert Tesch dahingehend, dass die in der Vor- und Nachbereitung der Experimente enthaltene Einbettung von besonderer Bedeutung für die Unterrichtsqualität sei.

Von der zweiten Projektphase existiert bislang keine abschliessende Analyse. Angaben zum Design finden sich in Tesch & Duit (2004). Erste Ergebnisse werden in Tesch & Gerber (2005) vorgestellt. Insgesamt 90 vorwiegend zufällig ausgewählte Klassen des 9. Schuljahres aus Deutschland und der Schweiz¹⁶ (vgl. Gerber, 2007) nahmen an der Studie teil. Aufgezeichnet wurden Einführungsdoppelstunden in die Optik und in den Kraftbegriff. Analog zur ersten Projektphase fanden zudem am Anfang und am Ende des Schuljahres begleitende Fragebogenerhebungen statt.

Im Fokus der Untersuchung steht dabei neu die Identifikation länderspezifischer Unterschiede. Die in Tesch & Gerber (2005) beschriebenen Ergebnisse bestätigen einen Teil der in der ersten Phase gefundenen Unterrichtsmuster, insbesondere die Dauer der im Kontext zu einem Experiment stehenden Unterrichtszeit und den geringen Grad an Offenheit der Experimentiereinheiten. Nach Tesch & Gerber unterrichten Schweizer Lehrpersonen, im Vergleich zu ihren deutschen Kolleginnen und Kollegen, eher schülerorientiert. Die Autoren machen, entsprechend dieser Aussage, bei der Organisation der Experimente in der Schweiz einen grösseren Grad an Offenheit aus als in Deutschland. Länderspezifische Unterschiede bezüglich der im Kontext mit Experimenten stehenden Unterrichtszeit können keine nachgewiesen werden (Dalehefte et al., 2009). Tendenziell wird im schweizerischen Physikunterricht jedoch mehr Zeit für die eigentliche Handlungsaktivität (*Experimentieren*) aufgewendet als im deutschen Unterricht. Dagegen werden Experimente in Deutschland eher länger vor- und nachbereitet als in der Schweiz.

3.5.2 Studien zur Effektivität experimentellen Handelns

Neben dem Kategoriensystem von Tesch (2005), welches für die Videoanalyse entwickelt wurde, existieren weitere Versuche, experimentelles Handeln zu kategorisieren. Millar et al. (1999) schlagen ein Kategoriensystem vor, welches bei verschiedenen Untersuchungen zum Einsatz kam: Tiberghien et al. (2001) verwenden es für die Analyse und den Vergleich von Experimentieranleitungen auf der Sekundar- und

¹⁶50 Klassen aus vier Bundesländern sowie 40 Klassen aus der Deutschschweiz.

der Teritärstufe in mehreren Ländern. Abrahams & Millar (2008) nutzen das System als Basis für eine qualitative Fallanalyse experimentellen Handelns im englischen Naturkundeunterricht 11- bis 16-jähriger Schülerinnen und Schüler. Auf die Untersuchung wird nachfolgend vertieft eingegangen. Eine Weiterentwicklung des Systems findet sich bei Millar (2009).

Der Frage, wie effektiv der Einsatz von Experimenten als Lehr-Lern-Strategie im Unterricht ist, wird in der Untersuchung von Abrahams & Millar (2008) nachgegangen. Dabei wurden sowohl leistungsbezogene als auch affektive Variablen untersucht, wobei der erwähnte Artikel sich ausschliesslich auf die Analyse von Variablen der Leistung bezieht. Die Studie basiert auf dem Modell von Millar et al. (1999). Darin wird zwischen zwei verschiedenen Effektivitäten unterschieden: Das erste Niveau der Effektivität ist als Passung zwischen den von der Lehrperson intendierten, praktischen und mentalen Handlungen und den von den Lernenden tatsächlich vollzogenen Handlungen zu verstehen. Dagegen bezieht sich das zweite Niveau der Effektivität auf die Passung zwischen den von der Lehrperson intendierten und den von den Schülerinnen und Schülern erreichten Lernzielen bzw. Kompetenzen. Die so definierten Effektivitätsniveaus experimentellen Handelns werden mit einer weiteren Dimension verbunden. Hierbei beziehen sich die Autoren auf Millar et al. (1999):

The fundamental purpose of practical work in school science is to help students make links between the real world of objects, materials and events, and the abstract world of thought and ideas. (Millar et al., 1999)

Millar et al. sprechen auch davon, Beziehungen zwischen der Domäne der Observablen und der Domäne der Ideen herzustellen. Das verwendete Modell wird somit durch eine 2x2 Matrix aufgespannt: (Effektivität Niveau 1, Effektivität Niveau 2)x(Domäne der Observablen, Domäne der Ideen). Die Autoren bemerken, dass die Elemente der Matrix nicht unabhängig sind und weiter, dass selbst einfache Beobachtungen theoriegeladen sind. Sie zitieren Feyerabend (1976, 1983), für den Beobachtungen "fully theoretical" sind. Trotzdem kann, wie auch Feyerabend vorschlägt, eine pragmatische Unterscheidung zwischen den beiden Domänen getroffen werden.

An der Studie nahmen 25 Klassen aus acht Schulen teil. Die Erhebung umfasste die Beobachtung einer oder mehrerer Naturkundestunden des nationalen Curriculums Key Stage 3 oder 4 (11-14 bzw. 15-16-jährige Schülerinnen und Schüler). Dabei sollten die Stunden experimentelles Handeln beinhalten. Zudem plante man (gemäss

Ankündigung) Interviews mit den betroffenen Lehrpersonen und mit einzelnen Schülerinnen und Schülern zu den Stunden. Es wurden keine Vorgaben bezüglich dem unterrichteten Thema gemacht. Jedoch achtete man darauf, dass die drei Naturwissenschaften Biologie, Chemie und Physik angemessen vertreten waren. Die geringe Fallzahl bei der Biologie erklären die Autoren mit den weniger häufig stattfindenden Experimenten im entsprechenden Fach.

Abrahams & Millar stellen fest, dass eine deutliche Mehrheit der beobachteten Experimente das Lernen von Fachwissen zum Ziel hatten. Aspekte des experimentellen Aufbaus, der Datenerhebung oder der Interpretation der Ergebnisse wurden fast nie thematisiert. Das bedeutet gemäss den Autoren, dass Schülerinnen und Schüler ein implizites Verständnis davon erlangen, was experimentelles Handeln ausmacht. Oft verbinden Lehrende das Experiment mit der Entwicklung einer bestimmten wissenschaftlichen Idee bei den Schülerinnen und Schülern. Gleichzeitig wird auf der Ebene der Planung des Experiments nicht ersichtlich, wie Schülerinnen und Schüler ihr Handeln mit der erwähnten Idee verbinden sollten. Hier fehlt es an Überlegungen der Lehrperson dazu. Grundsätzlich wird sehr wenig Zeit darauf verwendet, die Entwicklung von Ideen bzw. Konzepten bei Schülerinnen und Schülern zu fördern. Die meisten Lehrpersonen scheinen dabei eine empiristische Sichtweise auf die Rolle des Experiments zu vertreten. Somit erwarten sie, dass sich die relevanten Ergebnisse aus dem Experiment, also den Beobachtungen und Messungen, ergeben.

Die Autoren vertreten die Ansicht, dass die Effektivität experimentellen Handelns deutlich verbessert werden könnte, wenn Lehrpersonen sich darüber bewusst wären, dass wissenschaftliche Ideen sich nicht aus Beobachtungen oder Messungen ergeben. Sie plädieren dafür, wissenschaftliche Ideen im Unterricht sorgfältig einzuführen. Dabei reiche es nicht, die Ideen vor bzw. nach dem Experiment zu etablieren. Vielmehr müssten diese schon während der Aktivität eine wichtige Rolle spielen. Nur durch die Ideen liesse sich das Experiment überhaupt erst verstehen.

3.6 Qualitätskriterien experimentellen Handelns im Physikunterricht

In den vorangehenden Abschnitten dieses Kapitels wurden zunächst die fachdidaktischen und wissenschaftstheoretischen Dimensionen experimentellen Handelns aufgezeigt, sodann ihre Ziele ausgeführt und schliesslich die Befunde aus der Unterrichts-

forschung dargestellt. Gegenstand des Abschnitts ist nun das Zusammenführen der verschiedenen Aspekte zu einem ausformulierten Qualitätsverständnis experimentellen Handelns.

Experimentelles Handeln kann aus verschiedenen Perspektiven beurteilt werden. Nachfolgend werden drei Perspektiven postuliert, die aus den bisherigen Überlegungen bedeutsam erscheinen. Sie basieren auf den Arbeiten von Millar et al. (1999), Hodson (1993), Harlen & Wake (1999) u.a. Weitere Perspektiven sind denkbar. Es wird somit hier kein Anspruch auf Vollständigkeit erhoben. Der Entscheid, die Perspektiven auf die genannten Arbeiten abzustützen, hat hauptsächlich normativen Charakter. Millar et al., Hodson, Harlen & Wake verbindet, dass sie sich dem Experiment im Physikunterricht aus einer klar fachdidaktischen Position annähern. Ihre Position ist durch die Curriculumsreform¹⁷ in Grossbritannien aus den 60er Jahren geprägt, bei der problemorientierte Schülerexperimente eine bedeutende Rolle ausmachten. Nach Auffassung von J.B. wird ihrer reflektierten und teilweise kritischen Haltung zum *discovery-learning* im deutschen Sprachraum noch zuwenig Beachtung geschenkt, woraus sich die Auswahl der Autoren begründet.

Die Perspektiven P1-P3: *Experimentelles Handeln* wird als eine

P1 kontextorientierte,

P2 reflexive,

P3 und theoriegeleitete

Aktivität aufgefasst. Durch die Perspektiven wird ein Qualitätsverständnis postuliert, welches nachfolgend genauer beschrieben wird.

P1 Experimentelles Handeln als kontextorientierte Aktivität. Experimente sind Lerngelegenheiten, die den Lehrenden helfen, ihren Unterricht zu strukturieren (Jonas-Ahrend, 2004, Labudde, 2000). Eine Strukturierung ist jedoch nur dann möglich, wenn es zwischen den Einheiten experimentellen Handelns und dem verbleibenden Unterricht Bezüge gibt. Je mehr Bezüge aufgezeigt werden können, desto eher können Schülerinnen und Schüler das Experiment als Lerngelegenheit wahrnehmen und Verantwortung für ihr Lernen übernehmen. Gleichzeitig sind dies die Voraussetzungen, damit experimentelles Handeln als motivierende Aktivität erlebt

¹⁷Damit verbunden das Nuffield Science Project (vgl. Abschnitt 3.2).

werden kann (Hodson, 1993). Das Aufzeigen eines Beziehungsnetzes zwischen dem Experiment und inhaltlichen bzw. lebensweltlichen Aspekten des Unterrichts wird als wesentliches Element der *Kontextorientierung* aufgefasst. Mit dem Verlauf der Experimentiereinheit werden Bezüge gefestigt bzw. neue hergestellt. Lehrende sollen im Sinne eines selbstverantwortlichen Lernens deutlich machen, welche Intentionen mit der Aktivität verbunden werden (Hart et al., 2000).

The learning that is intended from a particular piece of practical work should be clear and it should be tailored in order that it can serve that purpose effectively [...] Both teachers and pupils need to be aware of the purpose of the practical work in a particular case and prepare for it and follow it up in relation to the specific purpose. (Harlen & Wake, 1999)

Wissenschaftliche Experimente stehen kaum für sich alleine. Sie sind in einem Kontext situiert (Jenkins, 1999). Für Hodson (1996) ist experimentelles Handeln ohne einen Kontextbezug nicht denkbar. Hodson (1996), Millar (1989) argumentieren, dass die für experimentelles Handeln wichtigen *scientific attitudes* bzw. *habits of mind* nicht exklusiv forschenden Tätigkeiten zugeordnet werden können. Diese sind nach Auffassung der Autoren allgemeine kognitive Fähigkeiten, die für verschiedene, auch nicht-wissenschaftliche Tätigkeiten, relevant sind. Erst die Verbindung der kognitiven Fähigkeiten mit naturwissenschaftlichen Inhalten macht experimentelles Handeln aus.

Scientific classification [...] is not just a matter of noting similarities and differences – or would it be sufficient in science lessons to classify banknotes and postage stamps [...]? (Hodson, 1996)

P2 Experimentelles Handeln als reflexive Aktivität. Experimente sind Lerngelegenheiten, die in ihrem Verlauf eine Entwicklung zeigen. Somit lassen sich die meisten Experimente in Phasen einteilen, bei Tesch (2005) etwa in die drei Phasen *Vorbereitung*, *Experimentieren*¹⁸ und *Nachbereitung*. Eine Entwicklung lässt sich sowohl zwischen den Phasen als auch innerhalb der Phasen beobachten. Deutlich wird dies bei offenen Schülerexperimenten, wie sie dem Ziel *Experience of ‘doing science’*

¹⁸Die Phase wird später auch als Durchführungsphase bezeichnet.

in Harlen & Wake (1999) entsprechen. Hierbei sind Anlässe der Reflexion unumgänglich, denn erst durch sie wird eine Entwicklung für Schülerinnen und Schüler sichtbar: Lernende erleben dabei den Erfolg oder Misserfolg des Experiments. Sie werden gefordert, ihren Entwicklungsstand immer wieder neu zu bestimmen und ausgehend davon das weitere Vorgehen zu planen. Für Hodson ist experimentelles Handeln eine reflexive Aktivität:

[...] current knowledge and expertise informs and determines the conduct of the inquiry and, simultaneously, involvement in inquiry (and, crucially, reflection on it) refines knowledge and sharpens expertise. (Hodson, 1996)

P3 Experimentelles Handeln als theoriegeleitete Aktivität. Experimente sind Lerngelegenheiten, die naturwissenschaftliche Ideen, Vorstellungen und Konzepte transportieren und das Denken in ihnen fördern. In dem Zusammenhang wird von einer *Theorieleitung* des Experiments gesprochen (Millar et al., 1999). Sie ist ein konstruktivistisches Merkmal experimentellen Handelns im Unterricht (Labudde, 2000). Auch Schülerinnen und Schüler beurteilen Experimente im Licht ihrer Ideen, Vorstellungen und Konzepte. Sie sind dabei nicht notwendigerweise verträglich mit den etablierten naturwissenschaftlichen Theorien. Ein Zusammenführen der individuellen Präkonzepte mit tragfähigen naturwissenschaftlichen Ideen ist dabei der zentrale Aspekt des theoriegeleiteten experimentellen Handelns.

Die Theorieleitung des Experiments wird ausserdem durch neuere wissenschaftstheoretische Erkenntnisse unterstützt. (Koponen & Mäntylä, 2006, Duit & Tesch, in Vorb., Giere, 1991). Darunter wird verstanden, dass Theorien experimentellen Untersuchungen vorangehen. Für Kosso (1992) geben Theorien Antworten auf eine Frage von der Art: “what’s going on here?” In dem Sinne umfassen Theorien, wie sie im Unterrichtsverlauf zur Anwendung kommen, ganz unterschiedliche Niveaus der Entwicklung und beinhalten sowohl einfache qualitative Erklärungshilfen der Lernenden als auch elaborierte wissenschaftliche Vorstellungen. Theorien und Experimente sind ferner untrennbar miteinander verbunden, wie dies in den Modellaspekten “theory is a factor in experimental design” und “experimentation is a factor in theory construction” (van Fraassen, 1980, zitiert in Koponen & Mäntylä (2006)) in Abschnitt 3.2 bereits diskutiert wurde.

Teil II

Methodik

Kapitel 4

Die Videostudie QuIP

4.1 Eine Übersicht über die Methodik des Gesamtprojektes

Inwiefern unterscheidet sich der deutsche, finnische und Schweizer Physikunterricht im 9. Schuljahr? Welche Merkmale gelingenden Unterrichts lassen sich in den Ländern identifizieren? Diese beiden Forschungsfragen bilden den Ausgangspunkt des Forschungsprojektes “Quality of Instruction in Physics” (QuIP), welches gemeinsam mit Partnern der Universität Duisburg-Essen und der Universität Jyväskylä durchgeführt wurde (Labudde, 2006).

Das Projekt setzt beim Befund von PISA 2003 an, wonach finnische Schülerinnen und Schüler bei den Leistungstests in Naturwissenschaften signifikant besser abschneiden als ihre deutschen und Schweizer Kolleginnen und Kollegen (OECD, 2004). Aufgrund der Studienanlage kann PISA die Unterschiede jedoch nicht erklären. Verbunden mit den Forschungsfragen ist damit das Ziel, Zusammenhänge zwischen den erwähnten Leistungsunterschieden und dem jeweiligen länderspezifischen Unterricht zu identifizieren und zu beschreiben. Dafür wurde Unterricht zu vergleichbaren Bedingungen in den jeweiligen Ländern videographiert. Aus Erfahrungen anderer Videostudien ist bekannt, dass der Unterricht sich bezüglich seiner länderspezifischen Muster, der sog. *cultural scripts* unterscheidet (vgl. Stigler et al., 2000). Somit sind auch Unterschiede zwischen Finnland, dessen Physikunterricht noch nie im Rahmen einer Videostudie untersucht wurde, und den übrigen beiden Ländern zu erwarten.

Allerdings ist es bisher nicht gelungen, Zusammenhänge zwischen den *cultural scripts* und den Schülerleistungen aufzuzeigen. Es ist anzunehmen, dass sich die Ursachen für die Leistungsunterschiede eher durch tiefer gehende Analysen (hochinferente Verfahren) von Unterricht erschliessen lassen. Man spricht in dem Zusammenhang auch von der Tiefenstrukturebene des Unterrichts. Im Gegensatz dazu sind mit der Sichtstrukturebene Merkmale des Unterrichts verbunden, für die eine eher geringe Interpretationsleistung des Beobachters nötig ist, wie das z.B. bei den *cultural scripts* der Fall ist. Im Projekt dient die Sichtstrukturanalyse hauptsächlich der Erfassung von Grundinformationen des Unterrichts sowie der Identifikation bestimmter Unterrichtsphasen für die anschliessende Tiefenstrukturanalyse. Letztere erlaubt in der Ergänzung zur Sichtstrukturanalyse eine Bewertung des Unterrichts oder spezifischer Unterrichtsphasen nach Gesichtspunkten der Qualität (vgl. Fischer et al., 2010).

Dem Forschungsprojekt liegt ein systemisches Modell für Unterrichtsqualität zugrunde (Abbildung 4.1, vgl. Lipowsky et al. (2005), Neumann et al. (2009, 2010)). Zentrales Element ist das Unterrichtsgeschehen, welches durch die erwähnte Sicht- und Tiefenstruktur des Unterrichts beschrieben wird. Darauf Einfluss haben die Schülerinnen und Schüler sowie die Lehrperson. Diese werden durch Konstrukte wie Motivation, Ausbildung (Lehrperson) und familiären Hintergrund (Schülerinnen und Schüler) u. a. charakterisiert. Als Ausgangsgrössen (Output) des Unterrichtsgeschehens werden Veränderungen der Schülerkompetenz und -motivation erfasst.

Die Erhebung der Daten gliedert sich in einen Pretest, die Videographierung einer Doppelstunde zum Thema “Zusammenhang zwischen elektrischer Energie und Leistung” und einen Posttest (Abbildung 4.2). Sie erfolgte grösstenteils im Wintersemester 2008/2009 und wurde mit dem Ende des Sommersemesters 2009 abgeschlossen. Die Eignung des Themas ging aus einer Curriculumsanalyse der Schulsysteme in den drei Ländern hervor. Die Themenvorgabe gewährleistete zwischen den verschiedenen Fällen vergleichbare Bedingungen. Der Gefahr einer zu grossen Beeinflussung der Lehrpersonen durch die Themenvorgabe wurde mit einer sorgfältigen Kommunikation begegnet: Die Lehrpersonen wurden aufgefordert, aus ihrer Unterrichtseinheit Elektrizität, die sie wie gewöhnlich abhalten sollten, jene Doppelstunde auszuwählen, welche am besten zum Thema “Zusammenhang zwischen elektrischer Energie und Leistung” passe.

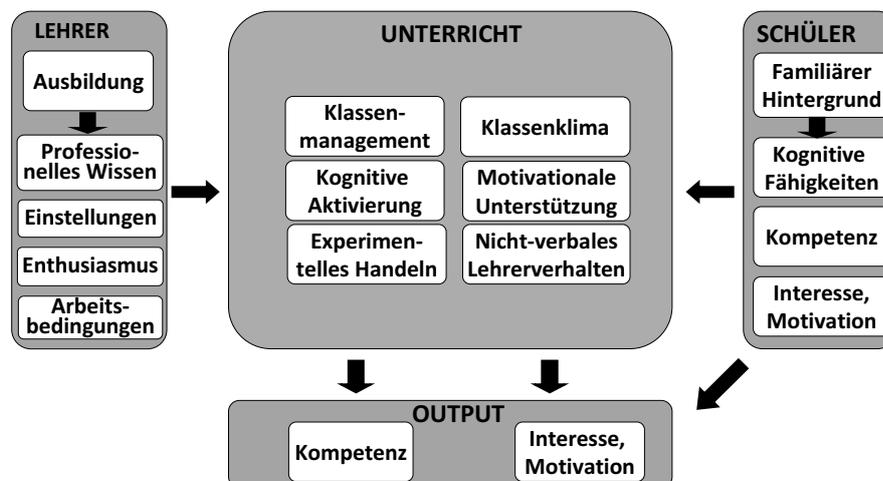


Abbildung 4.1: Systemisches Modell für Unterrichtsqualität: Der durch die Sicht- und Tiefenstruktur beschriebene Unterricht wird durch den Lehrer und die Schüler beeinflusst. Als Output werden Veränderungen der Kompetenz sowie Veränderungen affektiver Variablen der Schülerinnen und Schüler erfasst.

Der Zeitpunkt der videographierten Doppelstunde ergab sich damit aus der individuellen Konzeption der Unterrichtseinheit der Lehrperson. Die Zeitpunkte des Pre- und Posttests wurden wie folgt festgelegt: Die Durchführung des Pretests erfolgte unmittelbar vor dem Beginn der Unterrichtseinheit Elektrizität. Schliesslich wurde der Posttest am Ende der Unterrichtseinheit Elektrizität durchgeführt, falls zwischen diesem Zeitpunkt und jenem der videographierten Doppelstunde mindestens ein Monat verstrichen war. Andernfalls wurde der Posttest einen Monat nach der videographierten Doppelstunde durchgeführt. Typischerweise verstrich zwischen dem Pre- und dem Posttest ein Halbjahr.

Sowohl beim Pre- als auch beim Posttest wurden Instrumente zur Erfassung der Schülerkompetenz und Motivation eingesetzt. Dazu kamen beim Pretest Instrumente zur Erhebung des Lehrer- und Schülerhintergrundes, beim Posttest Instrumente zur Erfassung der pädagogischen und physikalischen Kenntnisse der Lehrperson sowie ein Experimentiertest für Schülerinnen und Schüler. Letzterer wurde aufgrund des grossen organisatorischen und materiellen Aufwandes ausschliesslich im Rahmen des Posttests durchgeführt. Der Experimentiertest wird derzeit im Rahmen einer Masterarbeit analysiert (Zberg, in Vorb.).

Die Typikalität der videographierten Doppelstunde erfasste man durch ein direkt nach der aufgezeichneten Stunde eingesetztes Fragebogeninstrument. Die Ty-

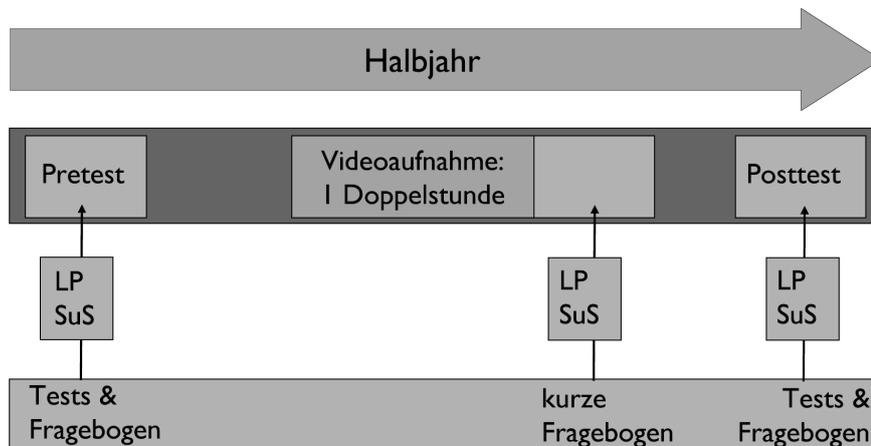


Abbildung 4.2: Zeitlicher Ablauf der Datenerhebung: Pretest, Videoaufnahme einer Doppelstunde und Posttest. Das Pre-/Posttestdesign erlaubt die Ermittlung der Veränderung von Schülerkompetenz und -motivation über das Halbjahr.

pikalität als Mass gibt dabei an, wie typisch der aufgezeichnete Unterricht von der Lehrperson, von den Schülerinnen und Schülern empfunden wurde. Das Herstellen eines Zusammenhangs zwischen dem Unterrichtsgeschehen, repräsentiert durch die videographierten Doppelstunden, und der Veränderung der Schülerkompetenz bzw. der Motivation setzt für den jeweiligen Unterricht typische Doppelstunden voraus.

4.2 Datenerhebung

Das gesamte Stichprobe umfasst Klassen des 9. und 10. Schuljahres, abhängig davon, in welchem Schuljahr das Thema “Zusammenhang zwischen elektrischer Energie und Leistung” unterrichtet wurde¹. Da das finnische Schulsystem als Gesamtschule organisiert ist und damit keine Aufgliederung nach Leistungsniveaus kennt, war es für die deutsche und schweizerische Länderstichprobe wichtig Schülerinnen und Schüler des gesamten Leistungsspektrums zu erfassen. Die Tabelle 4.1 zeigt die Gesamtstichprobe nach Leistungsniveaus (Schultypen) und Land dargestellt.

4.2.1 Datenerhebung in Deutschland und Finnland

In Deutschland werden drei Leistungsniveaus unterschieden (nach Leistungsniveau absteigend): Gymnasium (Typ 3), Realschule (2), Hauptschule (1). Des weiteren

¹Deutschland: 10. Schuljahr, Schweiz und Finnland 9. Schuljahr

Land	Schultypen				Σ
	0	1	2	3	
D	13	7	9	14	43
FIN	25	0	0	0	25
CH	4	9	11	7	31
Σ	42	16	20	21	99

Tabelle 4.1: Anzahl der Fälle (Klassen) verteilt auf Schultypen (0-3) und Länder. Typ 3 entspricht dem höchsten, 2 dem mittleren und 1 dem tiefsten Leistungsniveau. Dem Typ 0 werden Schulklassen zugeordnet, dessen Schülerinnen und Schüler sich aus verschiedenen Leistungsniveaus (Schultyp 1-3) zusammensetzen.

existieren Gesamtschulen, welche Schülerinnen und Schüler aus allen Niveaus umfassen. Einschränkend muss erwähnt werden, dass keine Hauptschulklassen berücksichtigt werden konnten. Das deshalb, weil das vorgegebene Thema im erwähnten Schultyp nicht unterrichtet wurde. Anstelle der Hauptschulklassen wurden Klassen des tiefsten Leistungsniveaus der Gesamtschulen (sog. G-Kurse) in die Ziehung der Stichprobe mit einbezogen. Aufgrund des grossen Aufwandes, die mit der Erhebung einer für Deutschland repräsentativen Stichprobe verbunden wäre, wurde die Ziehung auf das Land Nordrhein-Westfalen (NRW) eingeschränkt. Damit ist die Stichprobe repräsentativ für NRW, nicht aber für Deutschland (vgl. Olszewski, 2010).

Im finnischen Schulsystem wird nicht nach Leistungsniveaus unterschieden (vgl. Abschnitt 4.2). Für die Ziehung der Stichprobe wurden Schulen aus Mittelfinnland² berücksichtigt. Damit ist Stichprobe auch für die Region repräsentativ. Da das Schulsystem in Finnland national organisiert ist, kann davon ausgegangen werden, dass die Stichprobe weitgehend die Verhältnisse von ganz Finnland wiedergibt.

4.2.2 Datenerhebung in der Schweiz

Das Schulsystem der Schweiz liegt aufgrund der föderalen Struktur des Landes mehrheitlich in der Verantwortung der einzelnen Kantone. Analog zu Deutschland weisen die meisten Kantone ein dreigliedriges Schulsystem auf. Die unterschiedlichen Leistungsniveaus bzw. Schultypen werden je nach Kanton unterschiedlich bezeichnet. Die Bezeichnungen für die Leistungsniveaus im Kanton Bern (absteigende Reihenfolge) sind: Gymnasium (Typ 3), Sekundarschule (2), Realschule (1). Zusätzlich

²Region im Umkreis von 200km um Jyväskylä.

treten auch Schulklassen mit Schülerinnen und Schülern aus unterschiedlichen Leistungsniveaus auf. Diese werden im Folgenden dem Schultyp 0 zugeordnet.

Um die Komplexität der Studie einzugrenzen, wurden ausschliesslich Schulklassen der deutschsprachigen Schweiz bei der Ziehung berücksichtigt. Aufgrund der kantonal unterschiedlichen Curricula konnten selbst die Kantone der Deutschschweiz nicht alle bei der Erhebung berücksichtigt werden. Zur Prüfung der Verträglichkeit des Themas “Zusammenhang zwischen elektrischer Energie und Leistung” mit den kantonalen Curricula wurde eine Curriculumsanalyse erstellt. Sie basiert auf Szlovák (2005) und stützt sich auf weiterführende Recherchen der kantonalen Lehrpläne für das 9. Schuljahr aller Schultypen. Aufgrund der Analysen wurden folgende Kantone für eine Teilnahme an der Studie angefragt: Bern BE, Basel-Landschaft BL, Basel-Stadt BS, Luzern LU, Solothurn SO, Thurgau TG, Zürich ZH. Die Erziehungsdirektionen der Kantone wurden daraufhin um Erlaubnis für die Durchführung der Studie gebeten. Die angeschriebenen Erziehungsdirektionen stimmten alle der Durchführung der Studie in ihrem Kanton zu.

Stichprobe. Ausgehend von der gesamten Stichprobe, welche rund 40 Klassen umfassen sollte, wurden für jeden Kanton und Schultyp Unterstichprobengrößen bestimmt. Die Unterstichprobengrößen wurden proportional zur Schülerzahl in den jeweiligen Kantonen und Schultypen festgelegt (siehe Tabelle 4.2).

Kantone	Schultypen			Σ
	1	2	3	
Basel-Landschaft	2496	3383	2597	8476
Basel-Stadt	1096	1627	1715	4438
Bern	12667	14493	2297	29457
Luzern	3854	7397	2554	13805
Solothurn	1300	5873	1479	8652
Thurgau	3196	4364	293	7853
Zürich	15432	17113	5832	38377
Σ	40041	54250	16767	111058

Tabelle 4.2: Anzahl der Schülerinnen und Schüler nach Kantonen und Leistungsniveaus 1-3 (extrahiert aus Daten des Bundesamtes für Statistik 2005/06).

Für jeden Kanton und jedes Niveau wurde ein Inventar aller Klassen des 9. Schuljahres erstellt. Die Informationen dafür extrahierte man aus folgenden Quellen: Angaben des Bundesamtes für Statistik (Stand: 2005/06), Informationen der

Erziehungsdirektionen und Internetrecherchen. Aus den Inventaren wurden zufällig Klassen gezogen.

Die Schulleitungen der Klassen wurden für eine Teilnahme an der Studie angefragt. Verteilt auf 10 Ziehungen schrieb man insgesamt 603 (267 Typ 1 und 0, 309 Typ 2, 27 Typ 3) Schulen an. Von diesen willigten 27 Schulleitungen, betroffene Lehrpersonen und die Eltern von 30 Klassen ein, an der Studie teilzunehmen. Das entspricht einer Zusagequote von 5.0%. Die Gründe für die ersten 171 Absagen wurden analysiert. Als Gründe nannten Schulleitende und Lehrpersonen: Unvereinbarkeit mit dem Lehrplan (23%), (zu) viele Studienanfragen (22%), Fluktuationen/Vakanzen (15%), ohne Angabe von Gründen (12%), Arbeitsbelastung (11%), Reformen/Umstrukturierungen (6%), andere Gründe (6%), kein Interesse (5%).

Aufgrund der grossen Schwierigkeiten die geplante Stichprobengrösse zu erreichen, wurden Kontakte zu interessierten Lehrpersonen direkt hergestellt. In den meisten Fällen geschah dies über Fachdidaktikdozierende der betreffenden Kantone. So konnten weitere 7 Klassen für die Teilnahme gewonnen werden. Im August 2008 waren damit 37 Klassen und deren 34 Lehrpersonen für die Hauptstudie vorgesehen. Im Verlauf der Erhebung sagten leider sechs weitere Lehrpersonen ihre Teilnahme ab. Die Absagen wurden wie folgt begründet: längere Krankheit (2), Angst vor Videoaufnahmen (1), Unvorhergesehenes (2), unbekannte Gründe (1). Damit konnte die Erhebung mit 31 Klassen und deren 28 Lehrpersonen abgeschlossen werden. Die Verteilung der Klassen auf Schultypen und Kantone ist in Tabelle Tabelle 4.3 dargestellt.

Prozentual verteilen sich die Klassen wie folgt auf die verschiedenen Schultypen: typengemischte Klassen (Typ 0): 13%, Typ 1: 29%, Typ 2: 35%, Typ 3: 23% (vgl. Tabelle 4.3). Unter der Annahme, dass sich typengemischte Klassen hälftig aus Schülerinnen und Schülern des Typs 1 und 2 zusammensetzen, ergibt sich folgende prozentuale Verteilung: 35%(36%) (Typ 1), 42%(49%) (2), 23%(15%) (3). Die Angaben in Klammern entsprechen den tatsächlichen Anteilen der Schülerinnen und Schüler auf den drei Leistungsniveaus (vgl. Tabelle 4.2). Gegenüber der Grundgesamtheit ist damit der Typ 3 überrepräsentiert, währenddem Typ 2 unterrepräsentiert ist.

Kantone	Schultypen				Σ
	0	1	2	3	
Appenzell-Ausserrhoden	0	0	0	1*	1*
Basel-Landschaft	0	1	0	0	1
Basel-Stadt	0	0	0	0	0
Bern	0	1*	3	1 + 1*	4 + 2*
Fribourg	0	0	0	1*	1*
Glarus	0	0	0	1*	1*
Luzern	1	1	1	1	4
Solothurn	2	1	0	0	3
Thurgau	1*	3	3	0	6 + 1*
Zürich	0	1 + 1*	4	1	6 + 1*
Summe	3 + 1*	7 + 2*	11	3 + 4*	24 + 7*

Tabelle 4.3: Verteilung der Stichprobe nach Kantonen und Leistungsniveaus 1-3. Die mit * versehenen Klassenzahlen entsprechen jenen Klassen, welche über den direkten Kontakt zur Lehrperson für die Teilnahme an der Studie gewonnen wurden. Sie gehören damit nicht zur Zufallsstichprobe.

Kapitel 5

Schwerpunkt: Experimentelles Handeln

5.1 Sichtstruktur

Unter dem Begriff der Sichtstrukturmerkmale werden jene Merkmale des Unterrichts zusammengefasst, welche ohne grösseren Interpretationsaufwand feststellbar sind (vgl. Abschnitt 4.1). Beim experimentellen Handeln sind das z.B. die Organisationsform des Experiments, die Identifikation wichtiger Phasen wie der Vorbereitung, Durchführung und Nachbereitung des Experiments oder die Anzahl Experimente pro untersuchter Zeiteinheit. Diese Charakteristika experimentellen Handelns sind aus zwei Gründen wichtig: 1) Sie entsprechen einer wichtigen Terminologie der Unterrichtspraxis und erlauben damit Vergleiche zu anderen Studien bzw. zur eigenen Unterrichtspraxis. 2) Sie bilden die Grundvoraussetzung für weiterführende hoch-inferente Beurteilungsmethoden experimentellen Handelns. Auf der Basis der Sichtstrukturanalyse werden folgende Forschungsfragen untersucht (vgl. Abschnitt 2.2):

- S1 Welche experimentellen Handlungsmuster lassen sich im videographierten Unterricht identifizieren?
- S2 Inwiefern bestehen zwischen den Ländern Unterschiede bei den experimentellen Handlungsmustern?
- S2 Inwiefern sind zwischen den verschiedenen Schultypen Unterschiede bei den experimentellen Handlungsmustern auszumachen?

Die Analyse basiert auf dem von Tesch (2005) entwickelten Kategoriensystem (vgl. Abbildung 3.3). Das für die vorliegende Untersuchung angepasste Kategoriensystem von Tesch findet sich im Anhang, Kapitel A (vgl. Abbildung 5.1). Im Folgenden werden die wichtigsten Änderungen kurz zusammengefasst:

Hoch-inferente Kategorien werden weggelassen. Dazu werden folgende Kategorien gezählt: *Offenheit der Idee und Planung G1*, *Offenheit der Auswertung G2* sowie *Unterrichtsphase K6*. Sie werden teilweise in der Kodierung der Tiefenstruktur integriert (vgl. Abschnitte 5.2 und 5.3).

Das Kategoriensystem wird vereinfacht. Zur Vereinfachung des Kategoriensystems werden folgende Kategorien weggelassen: *Wiederholungsexperiment E3* sowie *Ablaufform K4*. Sie haben im Hinblick auf die zu untersuchenden Forschungsfragen eine untergeordnete Bedeutung.

Mit Ausnahme der Kategorie K1 wird die Subkategorie *Unklar* weggelassen. Bei K1 wird *Unklar* kodiert, wenn qualitative und quantitative Experimente gleichzeitig durchgeführt werden (z.B. bei Schülerexperimenten).

Die Subkategorie *Keine* übernimmt im System von Tesch (2005) meist eine Schalterfunktion. So werden während der Vorbereitung eines Experiments alle Kategorien K1 bis K5 auf *Keine* gesetzt. Die Wahl der Subkategorie *Keine* ist damit nicht eine Einschätzung der Kodierenden, sondern eine direkte Folge davon, dass für E1 Vorbereitung Experiment kodiert wurde. Zugunsten der Übersichtlichkeit des Systems, werden in der vorliegenden Arbeit jene Subkategorien mit Schalterfunktion in der Darstellung weggelassen (siehe Abbildung 5.1).

Die Kategorie *Organisationsform K2* wird auf zwei Teilkategorien erweitert. Im bestehenden Kategoriensystem lassen sich die Subkategorien nicht in jedem Fall disjunkt kodieren. So kann es vorkommen, dass Schülerinnen und Schüler in Gruppen Experimente an Stationen durchführen. Mit den beiden neu eingeführten Kategorien K2a und K2b lässt sich das auf eindeutige Weise kodieren. Zur Kategorie K2a wird zudem die Subkategorie Partnerarbeit hinzugefügt. Partner- und Gruppenarbeit sollen unterschieden werden, da das Arbeiten in Gruppen (bestehend aus mindestens drei Lernenden) wesentlich höhere Ansprüche an die Schülerinnen und Schüler stellt. Dazu gehören das Einnehmen einer Rolle bzw. Aufgabe in der Gruppe sowie die Aufteilung der anfallenden Arbeiten.

Zeitbasierte Kodierung in Intervallen von 20 statt 10 Sekunden. Bei der Pilotierung hat sich gezeigt, dass der überwiegende Teil der Unterrichtsphasen *Vorbereitung Experiment*, *Experimentieren* sowie *Nachbereitung Experiment* deutlich länger als 10 Sekunden dauern. Damit rechtfertigt sich die Verlängerung des Analyseintervalls auf 20 Sekunden.

Experimentiereinheit anstelle von Experimentierphase. Bei Tesch bezeichnet eine Experimentierphase ein Experiment inklusive Vor- und Nachbereitung. Von dieser Terminologie wird hier zur Erhöhung der Begriffsschärfe abgewichen: Die Phasen eines Experiments oder kurz Experimentierphasen bezeichnen die Bestandteile *Vorbereitung*, *Durchführung* und *Nachbereitung* eines Experimentes. Dagegen soll der Begriff der Experimentiereinheit für ein Experiment inklusive der Vor- und Nachbereitung verwendet werden.

In der Abbildung 5.1 sind die Kategorien und ihre Subkategorien dargestellt. Eine Besonderheit des Systems ist die hierarchische Ordnung der Kategorien. Die oberste Ebene der Hierarchie wird durch die Kategorie Experimentierphasen E1 gebildet. Dabei wird der Unterricht in die Subkategorien *Vorbereitung*, *Durchführung*, *Nachbereitung* und *Keine* eingeteilt. Jede Experimentiereinheit wird mit einer Laufnummer versehen (Kategorie *Nummerierung der Experimentiereinheiten* E2). Für die Subkategorien Vor- und Nachbereitung wird zusätzlich zwischen *inhaltlicher* und *organisatorischer Beschäftigung* unterschieden (*Art der Beschäftigung* G3). Mit den Kategorien *Datenerfassung*, *Organisationsform* und *Sachbegegnung* wird die Durchführung des Experiments genauer spezifiziert. Bei Schülerarbeitsphasen (*Organisationsform* Subkategorien 3, 4 oder 5) wird zusätzlich zwischen *Stationenlernen* (ja oder nein) sowie zwischen arbeitsgleicher und arbeitsteiliger Verteilung der Arbeit unterschieden.

5.1.1 Training der Kodierenden

An der Sichtstrukturanalyse experimentellen Handelns arbeiteten zwei Kodierende, wovon eine deutscher, die andere finnischer und deutscher Muttersprache war. Sie teilten sich die Stichprobe etwa hälftig auf. Zum Nachweis einer hinreichend grossen Interraterreliabilität wurden zusätzlich 10% der Fälle (11 Doppelstunden, siehe Tabelle 5.1) von beiden Kodierenden bearbeitet. Die entsprechenden Doppelstunden wurden zufällig aus der deutschen und schweizerischen Länderstichprobe gezogen.

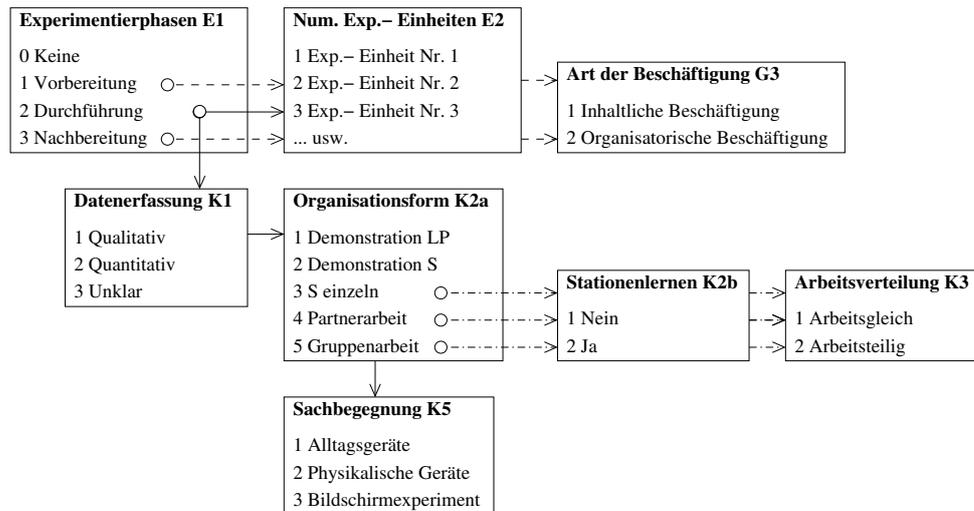


Abbildung 5.1: Kategorien der Sichtstruktur experimentellen Handelns basierend auf Tesch (2005). E1: Identifizierung Experimentierphasen; G3: inhaltliche oder organisatorische Beschäftigung während der Vor- und Nachbereitung; K1, K2a, K2b, K3, K5: Kategorisierung der Durchführungsphase.

Somit konnte die Übereinstimmung bzw. die Interraterreliabilität zwischen den beiden Kodierenden gemessen werden. Dabei wurde für die von beiden Kodierenden analysierten 10% der Stichprobe ein Cohens κ (Wirtz & Caspar, 2002, vgl.) von $\kappa > .8$ gefordert.

Um die angestrebte hohe Interraterreliabilität zu erreichen, kam dem sorgfältigen dem sorgfältigen Training der Kodierenden eine besondere Bedeutung zu. Dazu verwendete man Unterrichtsaufzeichnungen der Vorgängerstudie *Lehr-Lern-Kultur* (Gerber, 2007, Knierim, 2008). Der Ablauf des Trainings gliederte sich in drei Phasen: Einführung in das Kodiersystem, gemeinsames Kodieren und paralleles Kodieren. Gegenstand der Einführung war das Vertrautmachen mit dem Kodiersystem und das Klären von Fachbegriffen. Dazu wurden Sequenzen aus Unterrichtsvideos der Vorgängerstudie vorgeführt und besprochen. Anschliessend kodierte man gemeinsam Unterrichtsvideos, wobei Unstimmigkeiten direkt diskutiert werden konnten. Ziel dieser Phase war das Erreichen eines möglichst ähnlichen Verständnisses der Kategorien. Die Kodierenden konnten danach erste Unterrichtsvideos unabhängig voneinander, d.h. parallel kodieren. Die im Anschluss durchgeführte Interraterreliabilitätsanalyse zeigte allfällige Differenzen beim Verständnis der Kategorien und war Grundlage für die weiteren Trainingszyklen.

Phase	Fälle
1	S-1-05-01, S-2-13-01, S-2-28-01, S-3-09-01, D-1-25-01, D-2-42-01
2	D-0-33-01, D-1-29-01, D-2-23-01, D-3-32-01, S-1-33-01

Tabelle 5.1: Fälle, anhand derer die Beobachterübereinstimmung und das Cohens κ bestimmt wurden.

5.1.2 Interraterreliabilität

Die Interraterreliabilität wurde bei insgesamt 11 Doppelstunden aus Deutschland und der Schweiz analysiert (ca. 10% der Gesamtstichprobe, vgl. Tabelle 5.1). Die Kodierung der erwähnten Doppelstunden verteilte man innerhalb des Analyseprozesses auf zwei Phasen: Unmittelbar nach Abschluss des Trainings (Phase 1), nachdem ein Drittel der Gesamtstichprobe kodiert worden war (Phase 2). So konnten Veränderungen im Verständnis der Kategorien über den Analyseprozess abgeschätzt werden.

Eine Besonderheit bei der Berechnung der Interraterreliabilität (ausgedrückt durch die prozentuale Übereinstimmung sowie Cohens κ) stellt die Abhängigkeit der Kategorien dar. Beide Übereinstimmungsmasse setzen Kategorien voraus, welche unabhängig voneinander sind (vgl. Wirtz & Caspar (2002)). Das führt dazu, dass die Übereinstimmungsmasse bei bestimmten Kategorien nur für einen Teil aller Intervalle berechnet werden können. Im Folgenden werden die Abhängigkeiten zwischen einzelnen Kategorien und die damit verbundene Einschränkung bei der Berechnung der Übereinstimmungsmasse für jede Kategorie kurz erläutert.

E1: Die Kategorie ist unabhängig. Die Übereinstimmungsmasse beziehen sich auf sämtliche Intervalle der 11 Doppelstunden.

E2: Die Kategorie hängt von E1 ab. E2 wird nur dann eingeschätzt, wenn E1 $\neq 0$ ist. Zudem hängt jede Subkategorie von E2 von der Wahl der vorhergehenden ab. Aus diesem Grund werden die Übereinstimmungsmasse ausschliesslich für die ersten beiden Subkategorien (Experimentiereinheiten 1 und 2) betrachtet.

G3: Ist abhängig von der Kategorie E1. G3 wird nur dann eingeschätzt, wenn E1=1 oder E1=3 ist. Die Übereinstimmungsmasse werden damit nur für jene Intervalle berechnet, bei denen beide Kodierende E1=1 oder E1=3 gewählt haben.

K1, K2a, K2b, K3, K5: Hängen von Kategorie E1 ab. K1, K2a, K2b, K3 und K5, werden nur dann eingeschätzt, wenn E1= 2 ist. Das Mass der Übereinstimmung bezieht sich damit ausschliesslich auf jene Intervalle, für die E1= 2 zugeordnet wurde. Weiter hängen die Kategorien K2b und K3 auch noch von K2a ab. Die Masse der Übereinstimmung der Kategorien K2b und K3 beziehen sich somit auf jene Intervalle, in denen für K2a der Wert 3, 4 oder 5 gesetzt wurde.

In Tabelle 5.2 sind die Ergebnisse für die Phase 1 ausgewiesen. Bis auf die Kategorie *Organisationsform* K2a erreichen die Kodierenden eine sehr gute Übereinstimmung. Die Übereinstimmung bei der Kategorie K2a kann noch als akzeptabel bezeichnet werden.

Kategorie	Anz. Int.	κ	PU [%]
Experimentierphasen E1	1684	.86	90
Num. Exp.-Einheiten E2	701	.97	98
Art d. Beschäftigung G3	531	.82	95
Datenerfassung K1	246	.84	93
Organisationsform K2a	246	.67	78
Stationenlernen K2b	171	1.00	100
Arbeitsverteilung K3	171	.99	99
Sachbegegnung K5	246	.83	96

Tabelle 5.2: Cohens κ sowie prozentuale Übereinstimmung (PU) aller Kategorien für die Phase 1 der Sichtstrukturanalyse. Die Anzahl der analysierten Intervalle sind unter Anz. Int. aufgeführt.

In Tabelle 5.3 sind die Ergebnisse für die Phase 2 ausgewiesen. Bei der *Organisationsform* der Schülerexperimente (Stationenlernen) K2b und bei der *Arbeitsverteilung* K3 kann κ nicht berechnet werden. Das liegt daran, dass das Mass κ ein etwa gleich häufiges Auftreten von mindestens zwei Subkategorien voraussetzt. Wie ein Vergleich mit Tabellen A.14 und A.15 zeigt, ist diese Voraussetzung nicht gegeben. Bei allen 161 Intervallen ordneten beide Kodierenden konstant die Subkategorien *Nein* (K2b, Subkategorie 1 (Kein Stationenlernen)) und *Arbeitsgleich* (K3, Subkategorie 1) zu. Somit ergibt sich eine prozentuale Übereinstimmung von 100% bei beiden Kategorien, wobei jeweils nur eine Subkategorie kodiert wurde.

Mehrheitlich liegt die Übereinstimmung auch bei der Phase 2 in einem zufriedenstellenden Bereich. Die kleineren Unterschiede im Vergleich zu den Ergebnissen der Phase 1 lassen sich durch die Vielfalt experimenteller Handlungsmuster und den

damit verbundenen Schwierigkeiten bei der Kategorisierung erklären. Erfreulich ist, dass die Kategorie K2a deutlich besser abschneidet als bei der Phase 1, wo sich nur ein akzeptabler Wert ergab.

Kategorie	Anz. Int.	κ	PU [%]
Experimentierphasen E1	1337	.87	91
Num. Exp.-Einheiten E2	353	.69	92
Art d. Beschäftigung G3	295	.82	95
Datenerfassung K1	281	.78	92
Organisationsform K2a	281	1.00	100
Stationenlernen K2b	161	-	100
Arbeitsverteilung K3	161	-	100
Sachbegegnung K5	281	-.22	31

Tabelle 5.3: Cohens κ sowie prozentuale Übereinstimmung (PU) aller Kategorien ausgewiesen für die Phase 2 der Sichtstrukturanalyse. Die Anzahl der analysierten Intervalle sind unter Anz. Int. aufgeführt.

Als ungenügend muss die Kategorie *Sachbegegnung* K5 mit $\kappa = -.22$ gelten. Die ungenügende Interraterreliabilität in dieser Kategorie ist auf eine grundsätzliche Unstimmigkeit bei der Doppelstunde D-3-32-01 zurückzuführen, wie eine nachträgliche Befragung ergab: Im Rahmen einer Schülerexperimentiereinheit soll Wasser mit einem Tauchsieder erhitzt werden. Als Tauchsieder dient ein Widerstand, welcher mit einem Netzgerät verbunden wird. Die unterschiedlichen Positionen bezüglich der Sachbegegnung können wie folgt zusammengefasst werden:

Position 1: Aus elementaren elektrischen Bauteilen wird etwas hergestellt. Die Bauteile werden nicht für den Physikunterricht hergestellt. Daher, so wird argumentiert, handelt es sich bei den eingesetzten Materialien um Alltagsgeräte (und nicht um physikalische Geräte). Unterstützt wird die Position durch die Praxis bei S-1-05-01. Hier werden aus Nägeln, Blechstreifen, Brettern und Glühlampen etc. elektrische Stromkreise mit Schalter hergestellt. Diese Materialien wurden als Alltagsgeräte eingestuft.

Position 2: Die Materialien haben einen stark technischen Charakter. Unterstützt wird die Einschätzung durch die Sprechweise der Lehrperson. So ist von Betonwiderständen und einem Netzgerät die Rede. Das führt zur Auffassung, dass es sich bei den Materialien um physikalische Geräte handelt.

Bei der Gegenüberstellung beider Positionen fällt der Entscheid zu Gunsten der Position 2. Der technische Charakter der eingesetzten Materialien überwiegt. Gegen-

über S-1-05-01 lässt sich D-3-32-01 abgrenzen: Bei letzterer Doppelstunde kommen als technische Bauteile Widerstände zum Einsatz. Sie werden im Vergleich mit den in S-1-05-01 verwendeten Materialien als deutlich weniger alltäglich eingestuft.

Zusammenfassend kann festgestellt werden, dass mit Ausnahme der Kategorie K5 eine zufriedenstellende bis sehr gute Übereinstimmung erzielt werden konnte. Wie ausgeführt wurde, ergaben sich bei der Kategorie K5 Schwierigkeiten beim Verständnis von Alltagsgeräten und physikalischen Geräten. Hierbei variieren die Ergebnisse beider Phasen stark ($\kappa_1=.83$, $PU_1=96$; $\kappa_2=-.22$, $PU_2=31$). Die Befunde zu dieser Kategorie müssen mit der entsprechenden Vorsicht interpretiert werden (vgl. Abschnitt 9.1).

5.2 Tiefenstruktur

Im Gegensatz zur Sichtstrukturanalyse fokussiert die Tiefenstrukturanalyse auf Aspekte des Unterrichtsgeschehens, welche sich nicht unmittelbar aus einzelnen beobachtbaren Merkmalen erschliessen lassen. In den Vordergrund tritt dabei die Interpretation der Beobachter, welche das Unterrichtsgeschehen aus fachdidaktischer und pädagogischer Perspektive beurteilen. Man spricht auch von hoch-inferenten Beurteilungsmethoden. Bisherige Videostudien lassen vermuten, dass sich hoch-inferente Methoden im Vergleich zu niedrig-inferenten eher dazu eignen, qualitätsrelevante Aspekte des Unterrichts zu entschlüsseln (Pauli & Reusser, 2006, Petko et al., 2003, Stigler et al., 2000).

Zur hoch-inferenten Analyse experimentellen Handelns entwickelte man ein Ratinginstrument, dessen Kategorien bzw. Perspektiven bereits in Abschnitt 3.6 vorgestellt wurden. Das Ratinginstrument ist in nachfolgendem Abschnitt 5.3 zu finden. Eine Übersicht über die Perspektiven und den allgemeinen Teil des Instruments ist in Abbildung 5.2 dargestellt.

Den theoretischen Rahmen des Kategoriensystems bilden die Perspektiven *Experimentelles Handeln als kontextorientierte Aktivität*, *als reflexive Aktivität* und *als theoriegeleitete Aktivität*, auf die auch mit den Abkürzungen *Kontextorientierung*, *Reflexivität* und *Theorieleitung* verwiesen wird. Den Perspektiven sind Subkategorien, sog. Facetten zugeordnet (Tabelle 5.4). Von ihnen wird angenommen, dass sie die Hauptaspekte der betreffenden Perspektiven ausmachen. Die genaue Bedeutung

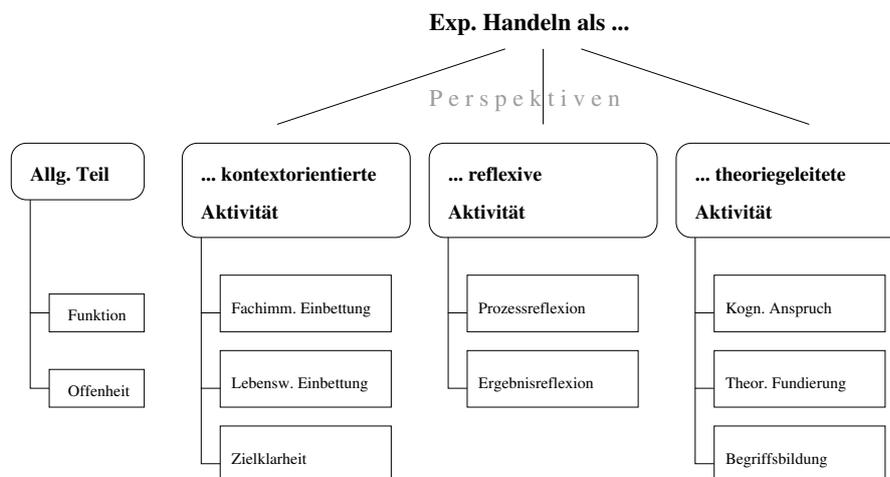


Abbildung 5.2: Die Struktur des Ratinginstrumentes bestehend aus den drei Perspektiven experimentellen Handelns und einem allgemeinen Teil (abgerundete Rechtecke). Den Subkategorien oder Facetten (Rechtecke) entsprechen Items, die auf einer vierstufigen Likert-Skala beurteilt werden.

der Facetten geht aus dem Instrument hervor (vgl. Abschnitt 5.3). Auszugsweise wird an dieser Stelle die Beschreibung der Facette *Ergebnisreflexion* wiedergegeben:

Ergebnisreflexion. Ergebnisse des Experiments werden zu einer übergeordneten griffigen Aussage zusammengefasst. Sofern vorhanden, wird die Problemstellung mit den Ergebnissen verglichen.

Beschreibung: Die entsprechende Situation tritt in der Regel im Rahmen der Nachbereitungsphase einer Experimentiereinheit auf. Dabei geht die Lehrperson auf die gefundenen Ergebnisse des Experiments ein und würdigt sie im Hinblick auf die Problemstellung, sofern von einer solchen ausgegangen wurde.

Meist werden Ergebnisse zusammengefasst und Vereinfachungen vorgenommen. Gelegentlich werden mit dem Experiment verbundene Anschlussfragen aufgeworfen und diskutiert.

Die notwendigen Bedingungen für eine Ergebnisreflexion sind: Fast alle Schülerinnen und Schüler können das Angebot nutzen. Die Lehrperson ist für die Qualität der Reflexion besorgt.

Konstitutive Elemente der Facette werden jeweils direkt nach dem Facettentitel in einer Kurzbeschreibung gegeben. Für die *Ergebnisreflexion* lässt sich entnehmen, dass es darum geht, Ergebnisse zu einer griffigen Aussage zusammenzufassen und die Problemstellung des Experiments mit den Ergebnissen zu vergleichen. In der

Beschreibung wird weiter aufgeführt, bei welcher Phase die Facette typischerweise auftritt, welche unterrichtlichen Merkmale dafür charakteristisch sind und an welche Bedingungen das Vorhandensein der Facette geknüpft ist. Die Ausprägungsgrade der Facette werden bewertet mit: *trifft zu*, *trifft grösstenteils zu*, *trifft teilweise zu* und *trifft nicht zu*.

Der allgemeine Teil des Ratingsystems umfasst die Subkategorien *Funktion* und *Offenheit*. Sie werden öfters dazu verwendet, Experimentiereinheiten zu klassifizieren (Kircher et al., 2001, Tesch, 2005, Millar, 2009). In die vorhandene Systematik der Perspektiven experimentellen Handelns lassen sie sich jedoch nicht einfügen und werden deshalb als Kontrollvariablen im allgemeinen Teil zusammengefasst. So bleibt die Anschlussfähigkeit an bisherige Forschungsarbeiten zur unterrichtlichen Praxis experimentellen Handelns gewährleistet. Die Ausgestaltung beider Subkategorien basiert auf dem ‘Practical Activity Analysis Inventory’ von Millar (2009).

In Tabelle 5.4 werden die Perspektiven und der allgemeine Teil mit den assoziierten Facetten bzw. Subkategorien zusammengefasst und mit Abkürzungen versehen.

Verbunden mit dem theoretischen Hintergrund des Ratinginstruments sind die folgenden Forschungsfragen (vgl. Abschnitt 2.3):

- T1 Findet sich die Qualität experimentellen Handelns, aufgespannt durch die drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung*, in der untersuchten Teilstichprobe?
- T2 Lassen sich bezüglich der Perspektiven oder Facetten Unterschiede zwischen den Ländern nachweisen?
- T3 Wie verhalten sich die Ratings einzelner Experimentiereinheiten innerhalb einer Klasse zu ihrer summativen Beurteilung (Gesamtrating)?
- T4 Welche Merkmale der Kategorien *Funktion* und *Offenheit* lassen sich in den Experimentiereinheiten identifizieren?

Anhand des Instruments beurteilten die Beobachter einzelne Experimentiereinheiten sowie alle Experimentiereinheiten einer Doppelstunde als Ganzes (Gesamtrating) (vgl. Abschnitt 5.2.2). Es handelt sich somit um ein Event-basiertes Rating, wobei jeweils eine Experimentiereinheit (bzw. alle Experimentiereinheiten einer Doppelstunde beim Gesamtrating) ein Event darstellen. Üblicherweise umfassen Experimentiereinheiten die Phasen *Vorbereitung*, *Durchführung* und *Nachbereitung*. Im

Kategorie/Perspektive	Facette	Abk.	Skalenniv.
Allgemeiner Teil	Funktion	FU	Nominal
	Offenheit	OF	Binär
Kontextorientierung	Fachimmanente Einbettung	EB1	Ordinal
	Lebensweltliche Einbettung	EB2	Ordinal
	Zielklarheit	ZK	Ordinal
Reflexivität	Prozessreflexion	PR	Ordinal
	Ergebnisreflexion	ER	Ordinal
Theorieleitung	Kognitiver Anspruch	KA	Ordinal
	Theoretische Fundierung	TF	Ordinal
	Begriffsbildung	BB	Ordinal

Tabelle 5.4: Zusammenfassung der verwendeten Kategorien/Perspektiven, Facetten mit Abkürzungen und dem entsprechenden Skalenniveau. Nominal: FU1 ... FU8, Binär: 0000 ... 1111 (0 Nein, 1 Ja), Ordinal: 1 (trifft zu), 2 (trifft grösstenteils zu), 3 (trifft teilweise zu), 4 (trifft nicht zu).

Rahmen der Sichtstrukturanalyse wurden die Phasen bereits identifiziert. Sie basieren auf dem Verständnis von Tesch (2005). Wie für hoch-inferente Analysen üblich, wurden alle Experimentiereinheiten unabhängig durch zwei Beobachter beurteilt. Um den dafür notwendigen Aufwand zu reduzieren, wählte man 45 der insgesamt 99 Doppelstunden für die Analyse aus. Im Folgenden wird auf die Ziehung der dafür benötigten Teilstichprobe eingegangen.

5.2.1 Wahl der Teilstichprobe

Der grosse zeitliche Aufwand, der mit dem gewählten Ratingverfahren verbunden war, führte dazu, die Analyse lediglich an einer Teilstichprobe durchzuführen. Diese sollte einerseits die verschiedenen Schultypen und andererseits die Länder angemessen repräsentieren. Die Studienanlage bedingte zudem, dass das experimentelle Handeln in den gewählten Fällen (Doppelstunden) angemessen vertreten war. Letzteres Kriterium fand Berücksichtigung, indem ausschliesslich Fälle bei der Ziehung der Teilstichprobe zugelassen wurden, die mindestens eine längere Experimentiereinheit enthielten. Unter einer längeren Experimentiereinheit wird ein Experiment verstanden, welches inklusive Vor- und Nachbereitung länger als 15 Minuten dauert. Wie aufgrund der Ergebnisse der Sichtstruktur zu erwarten war, genügten fast alle deutschen Fälle dem erwähnten Kriterium (40 von 43). Auch die schweizerische Stichprobe wurde dadurch nur wenig eingeschränkt (27 von 31). Hingegen dauerte in 10 der 25 finnischen Doppelstunden keine Experimentiereinheit länger als 15 Minu-

ten. Somit entspricht die eingeschränkte finnische Stichprobe bezüglich des zeitlichen Umfangs der Experimentiereinheiten nicht mehr der ursprünglichen. Dem Umstand muss bei der Auswertung Rechnung getragen werden.

Aus den Fällen, die längere Experimentiereinheiten enthielten, wurde eine Teilstichprobe zufällig gezogen. Die angemessene Repräsentation der Länder und Schultypen erreichte man durch die Schichtung der Teilstichprobe. In Tabelle 5.5 sind die absoluten Häufigkeiten der Fälle der Teilstichprobe nach Land und Schultyp zusammengefasst. Insgesamt 45 Fälle wurden für die Tiefenstrukturanalyse vorgesehen. Eine ausführliche Auflistung der so ausgewählten Fälle findet sich im Anhang, Tabelle B.1.

Land	Schultypen				Σ
	0	1	2	3	
D	3	4	4	4	15
FIN	15	0	0	0	15
CH	3	4	4	4	15
Σ	21	8	8	8	45

Tabelle 5.5: Die nach Land und Schultyp ausgewiesenen Fallzahlen (Anzahl Doppelstunden) der Teilstichprobe.

5.2.2 Ablauf des Ratings

Sämtliche Experimentiereinheiten wurden durch den Entwickler des Instruments sowie durch jeweils einen Hilfsassistierenden beurteilt. Letztere bereitete man sorgfältig auf ihre Arbeit vor: Sie umfasste die Auseinandersetzung mit der relevanten fachdidaktischen Literatur, eine Einführung in die Kategorien, die Arbeit an Trainingsvideos der Vorgängerstudie (Knierim, 2008, Gerber, 2007) sowie das Studium der nicht für die Tiefenstrukturanalyse vorgesehenen Fälle der Gesamtstichprobe. Aufgrund der erwähnten intensiven Auseinandersetzung mit dem Instrument und dem fachdidaktischen Kontext im Team rechtfertigt sich der Terminus des Expertenratings.

Eine Sitzung, bei der jeweils alle Experimentiereinheiten zu einem Fall geratet wurden, gliederte sich in folgende Schritte:

Identifizierung der Einheiten. Die Frage, was eine Experimentiereinheit alles umfasst, lässt sich im realen Unterricht manchmal nicht eindeutig beantworten. Pro-

bleme bei der Zuordnung der Experimentiereinheiten ergaben sich vorwiegend bei zwei Formen experimentellen Handelns: bei den Versuchsreihen sowie bei der Abfolge mehrerer Variationen eines bestimmten Experiments. Beide Formen zeichnen sich durch eine enge inhaltliche Beziehung zwischen den verschiedenen Teilen des Experiments aus. Die Beziehungen konnten im Rahmen der Sichtstrukturanalyse nur in ungenügender Masse identifiziert werden. Gleichzeitig bewährte sich bei der Pilotierung im Fall von Versuchsreihen oder bei der Abfolge mehrerer Variationen eines bestimmten Experiments eine Gesamtbeurteilung. Das machte eine vorgängige Identifikation der erwähnten Formen experimentellen Handelns unerlässlich. Dabei wurde wie folgt vorgegangen: Die auf der Basis der Sichtstruktur identifizierten Experimentiereinheiten sichtetete man auszugsweise. Im Falle enger inhaltlicher Verflechtungen zwischen den Einheiten wurde über eine Zusammenlegung der betreffenden Einheiten befunden.

Sichtung der Einheiten. Die Experten sichteteten die Experimentiereinheiten in chronologischer Reihenfolge. Für das spätere Rating relevante Aussagen bzw. Vorkommnisse wurden von beiden Experten unabhängig festgehalten. Zusätzlich verfasste der Entwickler des Instruments eine Kurzbeschreibung des Experiments. Den beschriebenen Schritt führten die Experten für sich, ohne den Austausch von Bewertungen oder die Kommentierung von Vorkommnissen, durch.

Zuordnen der Ratings. Die Experten bewerteten jede Facette für sich auf der Grundlage des Ratinginstruments. Sie verfassten zu jeder Bewertung eine kurze Begründung. Dieser Schritt begann bereits im letzten Drittel der Sichtung einer Experimentiereinheit.

Vergleich und Konsensfindung. Anschliessend wurden Bewertungen und Begründungen beider Experten miteinander verglichen. Bei Unstimmigkeiten in der Bewertung wogen die Experten beide Begründungen gegeneinander ab. Der gefundene Konsens wurde in Form einer gemeinsamen Begründung festgehalten.

Gesamtrating. Beinhaltete ein Fall eine einzige Experimentiereinheit, so endete der Ratingprozess mit dem zuletzt genannten Schritt. Die Bewertung der Facetten jeder Perspektive übernahm man für das Gesamtrating der Doppelstunde. Oft beinhaltete ein Fall jedoch mehrere Experimentiereinheiten. In diesem Fall wurde ein

Gesamtrating erstellt, in dem alle Experimentiereinheiten eines Falls global bezüglich der drei Perspektiven bewertet wurden. Die Subkategorien des allgemeinen Teils (*Funktion* und *Offenheit*) eigneten sich aufgrund ihres Skalenniveaus nicht für ein Gesamtrating. Sie wurden deshalb für das Gesamtrating weggelassen.

Folgende Aspekte berücksichtigen die Experten bei der Vergabe des Gesamtratings: Die Bedeutung der einzelnen Experimentiereinheiten innerhalb eines Falls, Unterschiede in der Dauer der Experimentiereinheiten sowie den Gesamteindruck.

5.2.3 Interraterreliabilität

Die Experten ordneten für alle deutschen und schweizerischen Fälle unabhängig voneinander Ratings zu (vgl. Abschnitt 5.2.2). Für die entsprechenden 48 Experimentiereinheiten (30 Doppelstunden) konnte damit die Übereinstimmung der Ratings zwischen den Experten in Form eines Intraklassenkoeffizienten ermittelt werden. Aus den Ergebnissen, die in Tabelle 5.6 zusammengefasst sind, geht hervor, dass gute Übereinstimmung ($ICC3 > .6$) bei den Facetten *Fachimmanente Einbettung* und *Lebensweltliche Einbettung*, *Ergebnisreflexion*, *Kognitiver Anspruch* und *Theoretische Fundierung* erzielt werden konnten. Eher ungenügende Übereinstimmung werden bei den Facetten *Zielklarheit*, *Prozessreflexion* und *Begriffsbildung* festgestellt. Cohens κ für die no-

Facette	ICC3	F	Konf.
Fachimmanente Einbettung	.66	4.9	[.46, .79]
Lebensweltliche Einbettung	.66	4.8	[.46, .79]
Zielklarheit	.55	3.5	[.32, .72]
Prozessreflexion	.46	2.7	[.20, .66]
Ergebnisreflexion	.71	6.0	[.54, .83]
Kognitiver Anspruch	.69	5.4	[.50, .81]
Theoretische Fundierung	.72	6.2	[.55, .83]
Begriffsbildung	.57	3.6	[.34, .73]
Kategorie	κ	Konf.	
Funktion	.51	[.31, .71]	
Offenheit	-	[-, -]	

Tabelle 5.6: Intraklassenkoeffizienten (ICC3: two-way mixed model), F Statistik und das 95%-Konfidenzintervall, ausgewiesen für jede Facette. Für sämtliche Berechnungen des ICC3 gelten: $df1 = df2 = 47$ und $p < .001$. Cohens κ und das 95%-Konfidenzintervall für die Kategorie *Funktion*. Die Berechnungen basieren auf den Ratings der Fälle aus Deutschland und der Schweiz.

minale Variable *Funktion* beträgt: .51 (95%-Konfidenzintervall [.31, .71]). Für die Variable *Offenheit* wurde kein Übereinstimmungsmass berechnet.

Die Ergebnisse belegen, dass die Experten Experimentiereinheiten mit meist zufriedenstellender Übereinstimmung beurteilen. Allerdings überrascht die verhältnismässig tiefe Übereinstimmung der Kategorie *Funktion*. Bei der Entwicklung des Instruments wurde sie als eher niedrig-inferent eingestuft. Dank der durchgängig doppelten Kodierung werden die partiellen Unstimmigkeiten zwischen den Experten durch den Konsensfindungsprozess ausgeglichen, wodurch sich die Auswirkungen der Unstimmigkeit auf die Ergebnisse der vorliegenden Studie in Grenzen halten.

5.3 Das Instrument zur Analyse der Tiefenstruktur

5.3.1 Allgemeiner Teil

Funktion

In Bezug auf das Lernen in der Physik kann ein Experiment verschiedene Funktionen einnehmen. Oft werden im Unterricht Experimente eingesetzt, um Phänomene zu zeigen, die Schülerinnen und Schüler in ihrem Alltag nicht oder nicht unter kontrollierten Bedingungen beobachten können. Eine weitere typische Funktion, die Experimente im Unterricht einnehmen können, ist das Lösen eines technischen Problems (Bsp.: Wie muss eine Flurbeleuchtung verdrahtet werden, damit das Licht von zwei Schaltern aus unabhängig angesteuert werden kann?). Der Begriff der Funktion des Experiments wird bei Kircher et al. (2001, S.279) eingeführt. Kircher et al. unterscheiden dabei 14 verschiedene Funktionen des Experimentes. Für das vorliegende System wurde eine gröbere Einteilung vorgenommen, welche sich am *learning objective* bei Millar (2009) orientiert.

Die Experimentiereinheit übernimmt die folgende Funktion (*Zutreffendes Ankreuzen*):

- | | | |
|--|--------------------------|-----|
| Das Kennenlernen eines Laborgerätes oder Messinstruments. | <input type="checkbox"/> | FU1 |
| Das Lösen bzw. Erkennen eines technischen Problems. | <input type="checkbox"/> | FU2 |
| Durchführen einer oder mehrerer Messungen. | <input type="checkbox"/> | FU3 |
| Das Kennenlernen eines Phänomens. | <input type="checkbox"/> | FU4 |
| Eine physikalische Gesetzmässigkeit plausibel machen (quantitativ oder qualitativ). | <input type="checkbox"/> | FU5 |
| Ein physikalisches Konzept veranschaulichen. | <input type="checkbox"/> | FU6 |
| Das Erlangen eines besseren allgemeinen Verständnisses von den Arbeits- und Funktionsweisen der Naturwissenschaft. | <input type="checkbox"/> | FU7 |
| Anderes | <input type="checkbox"/> | FU8 |

FU1 Das Kennenlernen eines Laborgerätes oder Messinstruments steht im Vordergrund. *Beispiel:* Die Lehrperson führt in die Bedienung eines Multimeters ein. Dazu werden verschiedene Messungen durchgeführt.

- FU2** Diese Ausprägung bezieht sich auf technische oder konstruktive Problemstellungen. Als Lösung resultiert ein Produkt, welches den in der Problemstellung formulierten Anforderungen möglichst weitgehend genügt. Es handelt sich somit nicht um ein Experiment im naturwissenschaftlichen Sinn. *Beispiel:* Verdrahten einer Flurbeleuchtung mit zwei Schaltern.
- FU3** Es werden Messungen durchgeführt. Die Messungen werden nicht weiter mit einer Gesetzmässigkeit in Verbindung gebracht. Oft ist das Ziel einer solchen Einheit das Kennenlernen einer physikalischen Grösse sowie, ein Gefühl für auftretende Grössenordnungen zu bekommen. *Beispiel:* Mit dem Leistungsmessgerät wird die Leistung verschiedener Verbraucher bestimmt. Danach wird eine Rangliste erstellt.
- FU4** Dieser Ausprägung gehören Experimentiereinheiten an, bei denen das Kennenlernen eines Phänomens im Vordergrund steht. Zu dieser Ausprägung gehören oft qualitative Experimente, die in vielen Fällen als Lehrer-Demonstration organisiert sind. *Beispiel:* Durch einen gespannten Draht lässt die Lehrperson einen elektrischen Strom fließen. Die Schülerinnen und Schüler beobachten, dass der Draht zu glühen beginnt und durchhängt.
- FU5** Mit der Durchführung des Experiments wird eine physikalische Gesetzmässigkeit plausibel gemacht. Einerseits kann dies durch die Aufnahme einer Messreihe mit anschliessender Analyse geschehen. Andererseits ist auch ein qualitativer Abgleich zwischen dem Experiment und dem Gesetz denkbar. *Beispiel:* Schülerinnen messen an einem Widerstand Stromstärke und Spannung. Die Schülerinnen und Schüler erkennen den linearen Zusammenhang zwischen den beiden Grössen.
- FU6** Zu dieser Ausprägung gehören Experimente, bei denen das Verständnis eines physikalischen Konzeptes zentral ist. *Beispiel:* Ein elektrischer Stromkreis und ein Wasserstromkreis sind nebeneinander aufgebaut. Daran erläutert die Lehrperson die Wasseranalogie zum elektrischen Stromkreis.
- FU7** Zentral für diese Ausprägung sind epistemologische Aspekte experimentellen Handelns: Wie gelangt die Naturwissenschaft mit Experimenten zu Erkenntnis? Was für eine Bedeutung haben Messfehler bei der Interpretation eines Experiments? Was sind die wichtigen Phasen, die bei der Durchführung eines Experiments üblicherweise durchlaufen werden?

OF Offenheit

Die Dimension der Offenheit erfasst, inwiefern die Umsetzung einer Experimentiereinheit Möglichkeiten einer Einflussnahme durch die Schülerinnen und Schüler bietet ($J=Ja$, $N=Nein$).

Schülerinnen und Schüler können die Fragestellung eines Experiments mitbestimmen oder können aus mehreren Fragestellungen wählen.

J	N
---	---

OF1

Aus einer grösseren Auswahl von Materialien wählen Schülerinnen und Schüler jene aus, die ihnen für die Beantwortung einer Fragestellung oder Problemstellung relevant erscheinen.

J	N
---	---

OF2

Schülerinnen und Schüler werden aufgefordert, die Durchführung in einzelnen Aspekten im Hinblick auf die Beantwortung der Problemstellung selbst zu gestalten.

J	N
---	---

OF3

Schülerinnen und Schüler haben bei der Nachbereitung des Experiments Wahlmöglichkeiten.

J	N
---	---

OF4

Die Beurteilung nach den vier Kriterien bedarf keiner weiteren Ausführung.

5.3.2 Experimentelles Handeln als kontextorientierte Aktivität

Fachimmanente Einbettung. Die Lehrperson bereitet Schülerinnen und Schüler inhaltlich auf das Experiment vor.

Beschreibung: Die Lehrperson leitet zum Experiment über. Sie hebt vorher behandelte Fachinhalte hervor und setzt sie in Bezug zum Experiment. Andererseits werden die Schülerinnen und Schüler durch einen Ausblick auf das bevorstehende Experiment vorbereitet. Darüber hinaus gehört zu einer guten Einbettung auch das Wiederaufnehmen dieser Fachinhalte am Ende der Experimentiereinheit.

Trifft zu: Die Lehrperson bettet das Experiment klar erkennbar und gemäss obiger Beschreibung ein.

Trifft grösstenteils zu: Die Lehrperson bettet das Experiment grösstenteils gemäss obiger Beschreibung ein. Als Beobachter/ Beobachterin habe ich den Eindruck, die fachinhaltlichen Verknüpfungen zwischen Experiment und Vorwissen könnten noch ausgebaut werden.

Trifft teilweise zu: Die Lehrperson bettet das Experiment teilweise gemäss obiger Beschreibung ein. Als Beobachter/ Beobachterin erscheinen mir die fachinhaltlichen Verknüpfungen eher in einem losen Zusammenhang mit dem Experiment zu stehen.

Trifft nicht zu: Die Lehrperson bettet das Experiment nicht gemäss obiger Beschreibung ein. Dies kann bedeuten, dass ich als Beobachter/ Beobachterin vom Experiment überrascht werde oder dass auf die Ankündigung des Experiments direkt eine handlungsbezogene Vorbereitung folgt.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

Lebensweltliche Einbettung. Die Lehrperson bettet das Experiment in einen lebensweltlichen Kontext ein.

Beschreibung: Die Lehrperson leitet zum Experiment über. Dabei stellt sie das Experiment in einen lebensweltlichen Zusammenhang und aktiviert damit die Aufmerksamkeit und das Interesse der Schülerinnen und Schüler. Ein lebensweltlicher Zusammenhang ergibt sich durch die Verankerung des Experiments an einen gesellschaftlichen oder historischen Bezugspunkt. Darüber hinaus gehört zu einer guten Einbettung auch das Wiederaufnehmen der lebensweltlichen Zusammenhänge am Ende der Experimentiereinheit.

Trifft zu: Die Lehrperson bettet das Experiment klar erkennbar und gemäss obiger Beschreibung ein.

Trifft grösstenteils zu: Die Lehrperson bettet das Experiment grösstenteils gemäss obiger Beschreibung ein. Lebensweltliche Problemstellungen und Fragen werden aufgeworfen, aber nicht weiter in Bezug auf das Experiment vertieft.

Trifft teilweise zu: Die Lehrperson bettet das Experiment teilweise gemäss obiger Beschreibung ein. Lebensweltliche Problemstellungen und Fragen werden angedeutet. Die Verbindung zum Experiment bleibt unklar.

Trifft nicht zu: Die Lehrperson bettet das Experiment nicht gemäss obiger Beschreibung ein. Dies kann bedeuten, dass die Einbettung ausschliesslich an Fachinhalten erfolgt oder ganz ausbleibt.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

EB2

Zielklarheit. Mit dieser Dimension soll erfasst werden, inwiefern es der Lehrperson gelingt, die Problemstellung des Experimentes für Schülerinnen und Schüler verständlich und einleuchtend darzustellen.

Beschreibung: Ausgeprägte Zielklarheit zeichnet sich durch Klarheit der Problemstellung und Integration des gesamten Experiments aus.

Die Problemstellung des Experiments wird klar dargestellt und in einer für Schülerinnen und Schüler adäquaten Sprache formuliert. Die Problemstellung wird entweder zu Beginn der Einheit durch die Lehrperson vorgegeben oder erscheint als Ergebnis eines Aushandlungsprozesses zwischen der Klasse und der Lehrperson.

Die Problemstellung ist nicht kleinschrittig, d.h. sie umfasst das ganze Experiment. Bei parallel stattfindenden Schülerexperimenten können mehrere Problemstellungen vorkommen. Werden Problemstellungen schriftlich an Schülerinnen und Schüler abgegeben, so werden diese ebenfalls für die Bewertung berücksichtigt.

Trifft zu: Aus Sicht des Beobachters/ der Beobachterin erscheint die Experimentiereinheit ausgeprägt zielklar. Dies bedeutet: Die Problemstellung wird klar dargestellt. Zudem umfasst die Problemstellung das ganze Experiment.

Trifft grösstenteils zu: Aus Sicht des Beobachters/ der Beobachterin erscheint die Experimentiereinheit grösstenteils zielklar. Die Problemstellung wird weitgehend klar dargestellt. Zudem umfasst die Problemstellung das ganze Experiment.

Trifft teilweise zu: Aus Sicht des Beobachters/ der Beobachterin erscheint die Experimentiereinheit teilweise zielklar. Eine Problemstellung ist vorhanden, sie ist aber für Schülerinnen und Schüler schwer verständlich. Die Problemstellung umfasst das ganze Experiment.

Trifft nicht zu: Dem Beobachters/ der Beobachterin bleibt die Problemstellung bis zur Durchführung des Experiments unklar bzw. vorenthalten. Anstelle der Problemstellung tritt typischerweise ein kleinschrittiges Vorgehen.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

ZK

5.3.3 Experimentelles Handeln als reflexive Aktivität

Prozessreflexion. Die Lehrperson ermutigt Schülerinnen und Schüler in ausgeprägtem Mass, ihr Handeln bzw. ihre Ideen zu reflektieren: Wo stehe ich? Sind Korrekturen am Vorgehen notwendig? Ist meine Vermutung plausibel? Ist mein Erklärungsansatz genau genug formuliert? usw.

Beschreibung: Durch das Frage-/ Antwortverhalten der Lehrperson werden Schülerinnen und Schüler dazu ermutigt, längere Erklärungen abzugeben. Den stichwortartigen und ungenauen Antworten der Schülerinnen und Schüler begegnet die Lehrperson mit Nachhaken. Die Lehrperson zeigt grosses Interesse an den Erklärungsversuchen der Schülerinnen und Schüler und schafft damit eine wichtige Vertrauensbasis.

Die Lehrperson antizipiert übliche kognitive Prozesse der Schülerinnen und Schüler: Die Lehrperson entwickelt und reflektiert einen Gedankengang. Wichtig hierbei ist das modellhafte Vorgehen der Lehrperson sowie die Bemühung um die Konsistenz des Gedankengangs. Diese implizite Art der Ermutigung zur Reflexion tritt üblicherweise im Rahmen eines Lehrervortrags auf.

Generell soll die Qualität der Vorkommnisse höher gewichtet werden als die Häufigkeit.

Trifft zu: Es werden mindestens zwei Vorkommnisse beobachtet, bei denen die Lehrperson Schülerinnen und Schüler in ausgeprägtem Mass zur Reflexion im obigen Sinne anregt.

Trifft grösstenteils zu: Es wird ein Vorkommnis beobachtet, bei dem die Lehrperson Schülerinnen und Schüler in ausgeprägtem Mass zur Reflexion im obigen Sinne anregt.

Trifft teilweise zu: Es wird ein Vorkommnis beobachtet, bei dem die Lehrperson Schülerinnen und Schüler zur Reflexion im obigen Sinne anregt. Die Reflexion wirkt aber auf den Beobachter/ die Beobachterin oberflächlich.

Trifft nicht zu: Es wird kein Vorkommnis beobachtet, bei dem die Lehrperson Schülerinnen und Schüler zur Reflexion im obigen Sinne anregt. Einerseits kann dies bedeuten, dass stichwortartige Antworten der Schülerinnen und Schüler in ihrer Ungenauigkeit akzeptiert werden. Andererseits, dass die Lehrperson auf eine allgemein gehaltene Frage die Lösung präsentiert. Weiter antizipiert die Lehrperson nicht die üblichen kognitiven Prozesse der Lernenden.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

Ergebnisreflexion. Ergebnisse des Experiments werden zu einer übergeordneten griffigen Aussage zusammengefasst. Sofern vorhanden, wird die Problemstellung mit den Ergebnissen verglichen.

Beschreibung: Die entsprechende Situation tritt in der Regel im Rahmen der Nachbereitungsphase einer Experimentiereinheit auf. Dabei geht die Lehrperson auf die gefundenen Ergebnisse des Experiments ein und würdigt sie im Hinblick auf die Problemstellung, sofern von einer solchen ausgegangen wurde.

Meist werden Ergebnisse zusammengefasst und Vereinfachungen vorgenommen. Gelegentlich werden mit dem Experiment verbundene Anschlussfragen aufgeworfen und diskutiert.

Die notwendigen Bedingungen für eine Ergebnisreflexion sind: Fast alle Schülerinnen und Schüler können das Angebot nutzen. Die Lehrperson ist für die Qualität der Reflexion besorgt.

Trifft zu: Eine ausgeprägte Reflexion der Ergebnisse findet statt. Sofern eine Problemstellung den Ausgangspunkt der Experimentiereinheit bildet, werden die Ergebnisse mit diesen verglichen.

Trifft grösstenteils zu: Eine Reflexion der Ergebnisse findet statt. Sofern eine Problemstellung den Ausgangspunkt der Experimentiereinheit bildet, werden die Ergebnisse mit dieser verglichen. In einzelnen Teilaspekten weist die Reflexion Mängel auf, die aber für den Beobachter/ die Beobachterin nicht gravierend sind.

Trifft teilweise zu: In Ansätzen findet eine Reflexion der Ergebnisse statt. Als Beobachter/ Beobachterin erkenne ich jedoch offensichtliche Mängel, welche die Reflexion stören. Als offensichtliche Mängel gelten: Fehlender oder ungenügender Vergleich mit der vorhandenen Fragestellung. Unklare oder fehlerhafte Zusammenfassung der Ergebnisse. Als Folge davon scheint ein Zurückgreifen auf die Erkenntnisse aus dem Experiment nur schwer möglich.

Trifft nicht zu: Eine Reflexion der Ergebnisse findet nicht statt. Die Experimentierphase bricht somit kurz nach der Durchführung ab. Messergebnisse oder Beobachtungen werden mitgeteilt aber nicht reflektiert.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

ER

5.3.4 Experimentelles Handeln als theoriegeleitete Aktivität

Kognitiver Anspruch. Die Lehrperson gestaltet die Experimentiereinheit den Möglichkeiten der Schülerinnen und Schülern entsprechend kognitiv anspruchsvoll.

Beschreibung: Hierbei geht es um das Globalurteil der ganzen Experimentiereinheit. Schülerinnen und Schüler werden durch die Experimentiereinheit weder unter- noch überfordert. Die Auseinandersetzung mit dem Experiment ist für Schülerinnen und Schüler interessant und herausfordernd zugleich. Sie werden dazu animiert, dem Experiment bzw. Aspekten des Experiments auf den Grund zu gehen.

Trifft zu: Die Lehrperson hat die Experimentiereinheit gemäss obiger Beschreibung kognitiv anspruchsvoll gestaltet.

Trifft grösstenteils zu: Die Lehrperson hat die Experimentiereinheit gemäss obiger Beschreibung grösstenteils kognitiv anspruchsvoll gestaltet.

Trifft teilweise zu: Die Lehrperson hat die Experimentiereinheit gemäss obiger Beschreibung teilweise kognitiv anspruchsvoll gestaltet.

Trifft nicht zu: Die Experimentiereinheit ist für Schülerinnen und Schüler nicht kognitiv anspruchsvoll gestaltet. Die Schülerinnen und Schüler sind unter- oder überfordert.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

KA

Theoretische Fundierung. Die Lehrperson verbindet das Experiment mit einem Erklärungsmodell und regt Schülerinnen und Schüler zur Argumentation in diesem Modell an.

Beschreibung: Erklärungsmodelle¹ sind Vorstellungen, welche nicht unmittelbar an Observablen anknüpfen. Sie bilden den Rahmen, innerhalb dessen Ergebnisse bzw. Beobachtungen verstanden werden können. Erklärungsmodelle gehen nicht direkt aus Experimenten hervor. Sie verkörpern den kreativen Teil naturwissenschaftlichen Arbeitens.

Im Gegensatz dazu werden mit Gesetzen kausale Zusammenhänge, meist in Form einer mathematischen Relation, beschrieben. Sie können durch das Experiment vorläufig bestätigt werden. Allerdings bieten Sie keine Erklärungen, weshalb ein Gesetz die gefundene Form einnimmt.

Der Begriff des

Die Unterscheidung zwischen den Begriffen Erklärungsmodell und Gesetz soll am folgenden Beispiel erläutert werden: Wie verändert sich die Stromstärke im Draht, wenn die elektrische Spannung erhöht wird? Durch eine Messreihe finden Schülerinnen und Schüler folgendes: Die Stromstärke, die durch einen Draht fließt, ist proportional zur angelegten Spannung. Dies ist eine Gesetzmässigkeit (Ohmsches Gesetz). Weshalb das Gesetz die gefundene Form einnimmt, lässt sich durch verschiedene Erklärungsmodelle begründen: A) Anhand der Wasseranalogie des elektrischen Stromkreises. Hierbei werden die Spannung mit der Druckdifferenz über einen Rohrabschnitt und die elektrische Stromstärke mit der Wasserstromstärke verglichen. B) Anhand der Teilchenvorstellung der elektrischen Ladung: Die Teilchen werden durch die angelegte Spannung im Leiter beschleunigt und immer wieder durch Zusammenstöße mit Rumpfatomen abgebremst². Beide Erklärungsmodelle können die Proportionalität zwischen Stromstärke und Spannung begründen.

Je nach Experiment können Erklärungsmodelle nur bei einer oder bei mehreren Phasen experimentellen Handelns bedeutsam sein.

Trifft zu: Die Lehrperson bezieht sich während des experimentellen Handelns *wiederholt* auf ein Erklärungsmodell. Sie regt Schülerinnen und Schüler zur Argumentation in diesem Modell an.

¹Es wird nicht zwischen den verwandten Begriffen Modell und Theorie unterschieden.

²Das sog. Drude-Modell lässt sich gemäss neueren empirischen Ergebnissen nicht aufrecht erhalten.

Trifft grösstenteils zu: Die Lehrperson bezieht sich während des experimentellen Handelns auf ein Erklärungsmodell. Sie regt Schülerinnen und Schüler zur Argumentationen unter Verwendung dieses Modells an.

Trifft teilweise zu: Die Lehrperson bezieht sich während des experimentellen Handelns in Ansätzen auf ein Erklärungsmodell. Es kann sein, dass eine Gesetzmässigkeit für Schülerinnen und Schüler plausibel gemacht wird und dabei implizit ein Erklärungsmodell angewendet wird, ohne jedoch explizit auf dieses hinzuweisen.

Trifft nicht zu: Die Lehrperson leitet Schülerinnen und Schüler nicht dazu an, experimentelles Handeln aus der Perspektive eines Erklärungsmodells zu beurteilen. Die Ergebnisse aus dem Experiment werden nicht weiter klassifiziert und werden nicht im Rahmen eines theoretischen Konzeptes zusammengefasst.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

TF

Begriffsbildung. Die Lehrperson setzt Alltagssprache und Fachsprache bewusst ein und leitet Schülerinnen und Schüler zu einer adäquaten Ausdrucksweise an.

Beschreibung: Zu einem bewussten Einsatz der Sprache gehören: Begrifflichkeiten und ihre unterschiedliche Bedeutung in Alltag und Fach hervorheben und reflektieren, eine sorgfältige und fachlich korrekte und konsequente Wortwahl, für Schülerinnen und Schüler adäquate Formulierungen wählen, die nicht anbiedern oder schwer verständlich sind..

Zur Anleitung einer adäquaten Ausdrucksweise gehören: Auf die Wortwahl der Schülerinnen und Schüler achten, bei Fehlern die Möglichkeit zur Korrektur geben, sich selbst um eine adäquate Ausdrucksweise bemühen usw.

Trifft zu: Die Lehrperson drückt sich, wie oben beschrieben, sprachlich differenziert aus und leitet Schülerinnen und Schüler zu einer adäquaten Verwendung der Fachsprache an.

Trifft grösstenteils zu: Die Lehrperson drückt sich, grösstenteils wie oben beschrieben, sprachlich differenziert aus und leitet Schülerinnen und Schüler zu einer adäquaten Verwendung der Fachsprache an.

Trifft teilweise zu: Die Lehrperson drückt sich, teilweise wie oben beschrieben, sprachlich differenziert aus und leitet Schülerinnen und Schüler zu einer adäquaten Verwendung der Fachsprache an.

Trifft nicht zu: Die Lehrperson setzt Sprache nicht wie oben beschrieben ein. Die Lehrperson setzt die Sprache weitgehend unreflektiert ein. Merkmal dafür sind entweder saloppe, ungenaue oder falsche Aussagen sowie die Vermengung von Alltags- und Fachsprache oder im Gegenteil unnötig komplizierte nicht adressatengerechte Formulierungen.

Gesamturteil: (1=Trifft zu, 2=Trifft grösstenteils zu, 3=Trifft teilweise zu, 4=Trifft nicht zu)

1	2	3	4
---	---	---	---

5.4 Fallanalysen

Das in Abschnitt 5.2 beschriebene Analyseverfahren der Tiefenstruktur führt zu Befunden über die Ausprägung der angelegten Qualitätsmerkmale in der untersuchten Teilstichprobe. Damit ergibt sich eine Klassifizierung der Fälle bzw. der Experimentiereinheiten. Es ist somit naheliegend, Fälle ähnlicher Klassifizierung im Rahmen einer weiterführenden Analyse zusammen zu betrachten. Was unter Ähnlichkeit verstanden werden kann, wird im Folgenden erläutert. Mit der weiterführenden Analyse, also der qualitativen Fallanalyse, werden folgende Forschungsfragen untersucht (vgl. Abschnitt 2.3):

T5 Was zeichnet Experimentiereinheiten hoher Qualität aus? Welche Eigenschaften der Lehrperson sind damit verbunden?

T6 Inwiefern decken sich die Befunde der Fallanalysen mit den Ergebnissen aus dem Ratingverfahren?

T7 Welche neuen oder verfeinerten Kategorien ergeben sich aus den Fallanalysen?

Wie soll die Ähnlichkeit von Experimentiereinheiten bezüglich der Qualitätsmerkmale definiert werden? Es muss beachtet werden, dass die Bildung eines Summenwerts von verschiedenen Ratings aufgrund des ordinalen Charakters der Facetten nicht zulässig ist (vgl. Tabelle 5.4). Deshalb wird die Ähnlichkeit zweier Ratings über eine in Gower (1971, Fall (c) S.859) vorgeschlagene Metrik für ordinal skalierte Variablen definiert. Jedem Eintrag der Metrik entspricht dabei eine der acht Facetten EB1, EB2, ZK, PR, ER, KA, TF und BB.

Die systematische Identifikation ähnlicher Fälle erfolgt mittels einer hierarchischen Clusteranalyse nach dem *Ward*-Verfahren (Bortz, 1993). Dafür wurde das Programm *pvclust* eingesetzt (Suzuki & Shimodaira, 2009). Die im Programm implementierte bootstrap-resampling Methode erlaubte zudem die Berechnung eines p-Wertes (approximately unbiased probability value), welcher angibt, wie gut ein Cluster, d.h. eine Gruppe ähnlicher Experimentiereinheiten, durch die Daten unterstützt wird (Suzuki & Shimodaira, 2009). Insgesamt liessen sich 9 Cluster mit einem $p > .95$ identifizieren (vgl. Tabelle C.1). Sämtliche im Rahmen der Fallanalyse ausgewerteten Einheiten entstammen dem Cluster IX, welcher Fälle mit besonders hohen Ausprägungen (tiefen Ratingwerten) versammelt.

Das Vorgehen bei der Analyse der Experimentiereinheiten ist an Mayring (2002) orientiert. Dabei wurde wie folgt vorgegangen:

Transkription. Aus den Video- und Audiodaten der entsprechenden Experimentiereinheiten erstellte man ein Transkript der Äusserungen der Lehrperson.

Transkriptdurchgang und Kategorienbildung. Das Transkript wurde zeilenweise analysiert. Dabei stellte man sich die Leitfrage: Inwiefern ist die in der Zeile gemachte Aussage ein Beleg für das Zutreffen bzw. nicht Zutreffen der drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung*? Aus den Exzerpten, d.h. Belegen für das Zutreffen bzw. nicht Zutreffen der Perspektiven wurden Kategorien gebildet. Diese versah man mit aussagekräftigen, am Transkript orientierten Paraphrasen. Inhaltlich ähnliche Exzerpte wurden der bereits bestehenden Kategorie zugeordnet (Subsumption).

Revision. Die bei der Analyse gebildeten Kategorien überarbeitete man nach etwa der Hälfte des Transkripts eines Falles. Ähnliche Kategorien wurden zusammengefasst und übergeordnete Kategorien formuliert.

Abschliessen des Materialdurchgangs. In der zweiten Hälfte des Materialdurchgangs ordnete man den Exzerpten vorwiegend bestehende Kategorien zu. In der Phase wurden üblicherweise nur noch wenig neue Kategorien gebildet.

Diskussion und Vergleich mit den Ratings. Die Forschungsfragen T5 und T6 wurden auf der Basis der emergenten Kategorien und der zugeordneten Exzerpten für jede Experimentiereinheit beantwortet. Die Frage T7 beantwortete man für alle analysierten Experimentiereinheiten zusammen.

Nachfolgende Fälle. Die Behandlung der weiteren Fälle erfolgte nach dem eben beschriebenen Ablauf, wobei bereits vorhandene Kategorien aus den vorhergehenden Fallanalysen als Ausgangspunkt übernommen wurden.

Vergleich der Fälle. Um die Fälle miteinander vergleichen zu können, wendete man bei allen Fällen ein durch den Analyseprozess revidiertes Kategoriensystem an (vgl. Tabellen 7.10, 7.11 und 7.12). Dabei wurde dieses nicht mehr verändert.

Teil III

Ergebnisse

Kapitel 6

Analyse der Sichtstruktur experimentellen Handelns

Im vorliegenden Kapitel werden die Ergebnisse der Analyse der Sichtstruktur vorgestellt. Als Sichtstruktur bezeichnet man jene Merkmale unterrichtlichen Handelns, die ohne grösseren Interpretationsaufwand feststellbar sind. Die Schwerpunkte der Analyse werden durch folgende Forschungsfragen vorgegeben (vgl. Abschnitt 2.2):

- S1 Welche experimentellen Handlungsmuster lassen sich im videographierten Unterricht identifizieren?
- S2 Inwiefern bestehen zwischen den Ländern Unterschiede bei den experimentellen Handlungsmustern?
- S3 Inwiefern sind zwischen den verschiedenen Schultypen Unterschiede bei den experimentellen Handlungsmustern auszumachen?

6.1 Stichprobe

Die untersuchte Stichprobe umfasst die Daten der Kodierungen von 99 Klassen (Fällen). In Tabelle 6.1 ist die Aufteilung der Klassen auf die drei Schultypen ersichtlich (vgl. Abschnitt 4.2). Die Schultypen 1 bis 3 beschreiben die verschiedenen Leistungsniveaus der Sekundarstufe I, wobei 1 dem tiefsten, 2 dem mittleren und 3 dem höchsten Leistungsniveau entsprechen. Als Bsp. dafür seien für den Kanton Bern die Schultypen in aufsteigender Reihenfolge angeführt: Realschule (1), Sekundarschule (2) und Progymnasium (3). Dem Typ 0 werden Klassen zugeordnet, welche sich aus

Schülerinnen und Schülern verschiedener Schultypen (1-3) zusammensetzen oder bei denen nicht nach Schultypen unterschieden wird, wie das z.B. in Finnland der Fall ist.

Land	Schultypen				Σ
	0	1	2	3	
D	13	7	9	14	43
FIN	25	0	0	0	25
CH	4	9	11	7	31
Σ	42	16	20	21	99

Tabelle 6.1: Verteilung der Fälle (Klassen) auf die verschiedenen Schultypen: Typ 1 entspricht dem tiefsten, 2 dem mittleren und 3 dem höchsten Leistungsniveau. Klassen des Typs 0 umfassen Schülerinnen und Schüler verschiedener Schultypen (1-3) oder es wird nicht nach Schultypen unterschieden.

6.2 Analysemethodik

Die Daten zur Sichtstruktur experimentellen Handelns werden deskriptiv ausgewertet. Dabei hängen die gewählten Auswertungsverfahren von wichtigen Eigenschaften, wie Normalität oder Homogenität der zugrundeliegenden Verteilungen ab. Es zeigt sich, dass die entsprechenden Verteilungen weder dem Kriterium der Normalität noch dem Kriterium der Homogenität durchgängig genügen. Der Befund ergibt sich aus der Anwendung des Shapiro-Wilk und des Bartlett-Tests. Einerseits lässt sich beobachten, dass Verteilungen bei einem Median nahe bei 90 Minuten linksschief, bei einem Median nahe bei null rechtsschief sind. Auf der anderen Seite unterscheiden sich auch die Varianzen der Verteilungen deutlich voneinander.

Aus den genannten Gründen kommen für die Auswertung der Sichtstruktur nicht-parametrische Verfahren zur Anwendung. Für die Lage des Mittelwertes wird jeweils der Median angegeben. Das 1. und 3. Quartil sowie das Maximum und das Minimum geben Aufschluss über die Streuung der Verteilung. Beim Vergleich zweier Mittelwerte wird der Wilcoxon Rangsummentest für zwei unabhängige Stichproben verwendet (Sachs & Hedderich, 2006). Die Signifikanzniveaus bei multiplen Mittelwertvergleichen, wie sie beim Ländervergleich vorkommen, werden nach Bonferroni korrigiert. Somit sind die Signifikanzniveaus für die drei Länderpaare bzw. die drei Länder: $\alpha = .05/3 \approx .017$ (*), $.01/3 \approx .003$ (**) und $.001/3 \approx .0003$ (***)

Die Ergebnisse zu den Kategorien werden jeweils in Tabellen zusammengefasst (vgl. Tabellen 6.2-6.11). In einer ersten Tabelle (Tabellen 6.2 und 6.4) sind für die Länderstichproben sowie für das gesamte Sample Lage- und Streuungsparameter der Verteilungen aufgeführt. Die Spalten umfassen: Minimum (Min), 1. Quartil (1Q), Median (Med), 3. Quartil (3Q) und das Maximum der Stichprobe (Max) sowie das 95 %-Konfidenzintervall des Medians (Konf.). Sofern nicht anders vermerkt, sind die Angaben in Minuten pro Doppelstunde zu verstehen. Eine zweite Tabelle (Tabellen 6.3, 6.5 und 6.11) umfasst paarweise Vergleiche von Verteilungen verschiedener Variablen. Die Spalten umfassen: Rangstatistik (W), Irrtumswahrscheinlichkeit (p), das 95%-Konfidenzintervall der Verschiebung der Verteilungen (Konf.) sowie die mittlere Verschiebung (Diff.). Die Spalten W und p haben die Einheit 1, die Einheiten der Spalten Konf. und Diff. sind, sofern nicht anders vermerkt, in Minuten pro Doppelstunde angegeben.

6.3 Ländervergleich

6.3.1 Unterricht im Kontext experimentellen Handelns

Der Unterricht im Kontext experimentellen Handelns hat in der analysierten Stichprobe eine herausragende Bedeutung. Das äussert sich insbesondere in der Unterrichtszeit, die im Kontext mit experimentellem Handeln steht. Dazu werden alle zeitlichen Intervalle gezählt, die der *Vorbereitung*, *Durchführung* oder der *Nachbereitung* eines Experiments zugeordnet werden können (vgl. Anhang, Abschnitt A.1). So definiert, stehen in der Hälfte der untersuchten Fällen 30 bis 65 Minuten der gesamten Unterrichtszeit (90min) im Zusammenhang mit experimentellem Handeln (Tabelle 6.2, Exp. Handeln gesamt).

Interessanterweise lassen sich bezogen auf die Länderstichproben deutliche Unterschiede festmachen, wie Abbildung 6.1 zeigt. Im deutschen Physikunterricht stehen im Mittel 65 Minuten, also zwei Drittel der gesamten Unterrichtszeit im Zusammenhang mit Experimenten. Im Gegensatz dazu wird in Finnland im Mittel nur knapp ein Drittel (28min) der Unterrichtszeit dafür verwendet. Der Schweizer Physikunterricht liegt hier mit der Hälfte der Unterrichtszeit (42min), die für experimentelles Handeln eingesetzt wird, zwischen Deutschland und Finnland. Die Länderstichproben unterscheiden sich bezüglich ihrer zentralen Tendenz bzw. Mediane alle signi-

fikant (vgl. Tabelle 6.3). Zwischen Deutschland und Finnland ergibt sich sogar ein höchst signifikanter, zwischen Deutschland und der Schweiz ein hoch signifikanter Unterschied. Die Länderstichproben weisen ähnlich grosse Streuungsparameter auf: Interquartilsabstand (IQA) (31-38)min, Spannweite (SPA) (66-79)min, wobei die deutsche Stichprobe linksschief, die finnische und schweizerische eher rechtsschief sind.

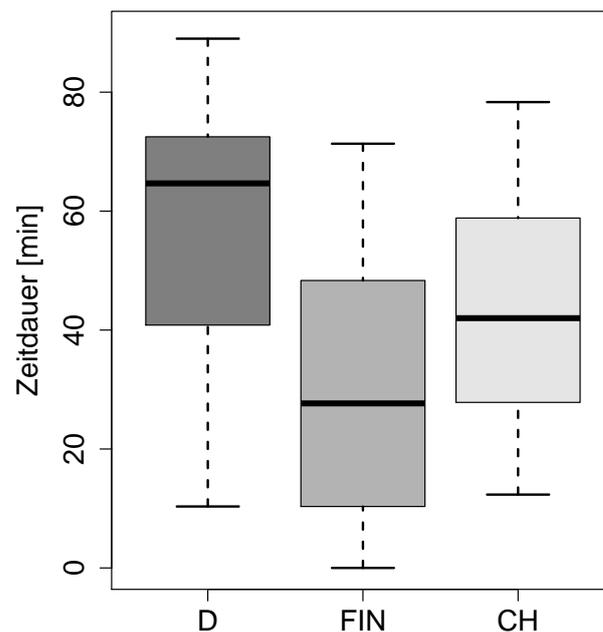


Abbildung 6.1: Die Zeitdauer pro Doppelstunde, die im deutschen, finnischen und Schweizer Physikunterricht im Zusammenhang mit experimentellem Handeln steht, entspricht den Angaben unter *Experimentelles Handeln insgesamt* von Tabelle 6.2.

6.3.2 Die Phasen experimentellen Handelns

Typischerweise setzt sich eine Experimentiereinheit aus den Phasen *Vorbereitung*, *Durchführung* und *Nachbereitung* zusammen (vgl. Anhang, Abschnitt A.1). Die Phasen *Vorbereitung* und *Nachbereitung* umfassen neben der inhaltlichen Auseinandersetzung auch alle organisatorischen Tätigkeiten im Zusammenhang mit dem Experiment: Auf- bzw. Abbau des Experiments, Einteilung in Gruppen, Ablaufplanung usw. Schliesslich beinhaltet die *Durchführung* jenen Teil der Experimentiereinheit, bei dem Messdaten erhoben bzw. Beobachtungen angestellt werden.

Land	Min	1Q	Med	3Q	Max	Konf.
<i>Keine</i>						
D	0	7	18	35	71	[10, 25]
FIN	16	31	53	71	89	[40, 66]
CH	6	19	39	51	64	[29, 48]
Ges.	0	16	33	51	89	[27, 39]
<i>Vorbereitung</i>						
D	3	10	15	23	34	[12, 18]
FIN	0	3	8	12	24	[5, 10]
CH	2	5	9	12	19	[8, 11]
Ges.	0	6	10	18	34	[8, 12]
<i>Durchführung</i>						
D	5	12	19	25	40	[16, 22]
FIN	0	4	11	22	34	[5, 17]
CH	4	10	19	30	46	[13, 25]
Ges.	0	10	18	25	46	[16, 20]
<i>Nachbereitung</i>						
D	1	12	18	23	39	[15, 20]
FIN	0	3	6	15	21	[2, 9]
CH	3	8	14	18	29	[12, 17]
Ges.	0	7	15	20	39	[13, 17]
<i>Exp. Handeln insgesamt</i>						
D	10	41	65	73	89	[57, 72]
FIN	0	10	28	48	71	[16, 40]
CH	12	28	42	59	78	[33, 51]
Ges.	0	30	49	65	89	[43, 55]

Tabelle 6.2: Zur Kategorie E1 sind die Lage- und Streuungsparameter der Subkategorien *Keine*, *Vorbereitung*, *Durchführung* und *Nachbereitung* aufgeführt. Unter *Experimentelles Handeln insgesamt* wird die Summe aus Vorbereitungs-, Durchführungs- und Nachbereitungszeit ausgewiesen. Die Lage- und Streuungsparameter sind in Minuten pro Doppelstunde zu verstehen.

Abbildung 6.2 zeigt, dass die Phasen *Vorbereitung*, *Durchführung* und *Nachbereitung* im deutschen Physikunterricht zu etwa gleichen Anteilen vorkommen. Während sich die Phasen in Deutschland bezüglich ihren Anteilen nicht unterscheiden, bestehen signifikante Unterschiede zwischen allen Phasen im Schweizer Physikunterricht: Die *Durchführung* macht mit 46% den grössten Anteil der Phasen aus, die *Vorbereitung* mit 20% den kleinsten. In Finnland unterscheidet sich der Anteil der *Durchführung* (39%) von den Anteilen der *Vorbereitung* und der *Nachbereitung* (30%, 30%) ebenfalls signifikant.

Variable	W	p	Konf. [min]	Diff. [min]
<i>D-FIN</i>				
Keine	113	.0000***	[-48, -22]	-35
Vorbereitung	811	.0005**	[4, 12]	8
Durchführung	744	.009*	[2, 14]	8
Nachbereitung	858	.0001***	[5, 15]	10
Gesamt	894	.0000***	[18, 41]	29
<i>D-CH</i>				
Keine	263	.03	[-24, -2]	-13
Vorbereitung	976	.0007**	[3, 11]	7
Durchführung	688	.8	[-6, 6]	1
Nachbereitung	842	.06	[0, 8]	4
Gesamt	953	.002**	[6, 25]	15
<i>FIN-CH</i>				
Keine	303	.006*	[8, 37]	22
Vorbereitung	326	.3	[-5, 2]	-2
Durchführung	242	.02*	[-15, -2]	-7
Nachbereitung	216	.005*	[-10, -2]	-6
Gesamt	240	.02*	[-26, -3]	-15

Tabelle 6.3: Mittelwertsvergleiche zur Kategorie *Experimentierphasen*. H_0 -Hypothesen: Die Länderpaare (D-FIN, D-CH oder FIN-CH) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. sind in Minuten pro Doppelstunde zu verstehen.

Absolut gesehen ergeben sich bei den Subkategorien *Vorbereitung*, *Durchführung* und *Nachbereitung* signifikante Unterschiede (Tabelle 6.3): Im finnischen Unterricht wird im Vergleich zum Unterricht der anderen Länder signifikant weniger Zeit für die *Durchführung* und die *Nachbereitung* der Experimente aufgewendet. Der deutsche und Schweizer Unterricht lässt sich diesbezüglich nicht unterscheiden. Am signifikant längsten vorbereitet wird in Deutschland (Median 15min, Tabelle 6.2). Die entsprechende Phase dauert im finnischen und im schweizerischen Unterricht ähnlich lang (Median 8 bzw. 9min).

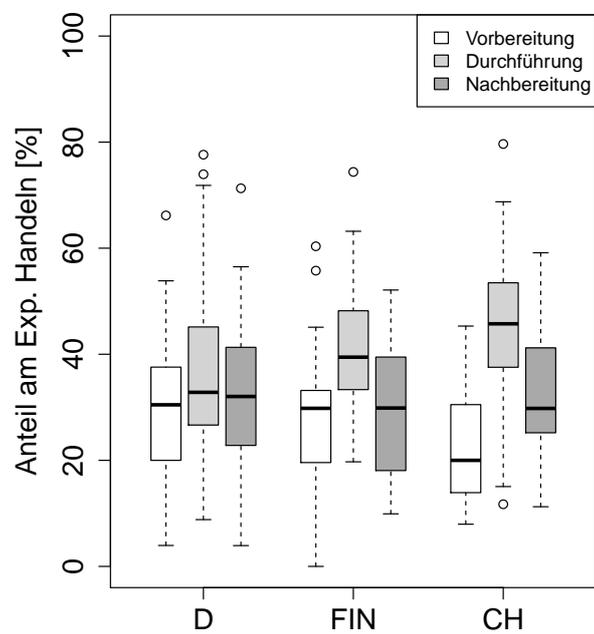


Abbildung 6.2: Der Anteil in Prozenten der *Vorbereitung*, *Durchführung* und *Nachbereitung* am experimentellen Handeln insgesamt, ausgewiesen für den deutschen, finnischen und den Schweizer Physikunterricht.

6.3.3 Art der Beschäftigung

Die Sequenzen der *Vor-* und *Nachbereitung* eines Experiments werden durch die Kategorie *Art der Beschäftigung* entweder einer inhaltlichen oder einer organisatorischen Beschäftigung zugeordnet (vgl. Anhang, Abschnitt A.1). In einigen Intervallen lässt sich eine inhaltliche Beschäftigung nicht klar von einer organisatorischen unterscheiden. Gemäss Kategoriensystem werden solche Intervalle der inhaltlichen Beschäftigung zugeordnet.

Land	Min	1Q	Med	3Q	Max	Konf.
<i>Inhaltliche Beschäftigung</i>						
D	56	78	86	94	100	[82, 90]
FIN	39	53	80	86	100	[69, 91]
CH	42	72	84	96	100	[77, 91]
Ges.	40	72	85	93	100	[81, 88]
<i>Organisatorische Beschäftigung</i>						
D	0	5	14	22	44	[10, 18]
FIN	0	14	20	47	61	[9, 31]
CH	0	4	16	28	58	[9, 23]
Ges.	0	7	16	31	66	[12, 19]

Tabelle 6.4: Zur Kategorie G3 sind die Lage- und Streuungsparameter der Subkategorien *Inhaltliche Beschäftigung* und *Organisatorische Beschäftigung* aufgeführt. Die Lage- und Streuungsparameter sind in Prozenten der Vor- und Nachbereitungszeit zu verstehen.

In Tabelle 6.4 sind die Streu- und Lageparameter der Subkategorien *Inhaltliche* und *Organisatorische Beschäftigung* in Prozenten der Vor- und Nachbereitungszeit aufgeführt. Insgesamt 85% der Vor- und Nachbereitungszeit wird inhaltlich gearbeitet. Ausschliesslich 16% macht der Anteil der rein organisatorischen Vor- und Nachbereitungszeit aus. Zwischen den Ländern lassen sich keine signifikanten Unterschiede feststellen.

6.3.4 Datenerfassung beim Experimentieren

So wie bei den Phasen *Vor-* und *Nachbereitung* wird auch die Phase der *Durchführung* anhand weiterer Kategorien spezifiziert. Diese sind: *Datenerfassung* K1, *Organisationsform* K2a, *Sachbegegnung* K5 und bei Schülerexperimenten zusätzlich *Organisationsform* K2b und *Arbeitsverteilung* K3 (vgl. Anhang, Abschnitt A.1).

Anhand der Kategorie *Datenerfassung* K1 wurde beurteilt, ob qualitative oder quantitative Experimente durchgeführt wurden. Als qualitativ bezeichnet man Experimente, deren Observablen direkt über die Sinne beobachtet werden (Vergleich von Helligkeiten, Beobachten eines Zeigerausschlags, Demonstration eines Wasserkreislaufs als Analogon zum elektrischen Stromkreis o.ä.). Dagegen sind quantitative Experimente mit einer meist schriftlichen Erfassung von Werten eines Messinstrumentes verbunden (Punkte einer Strom-Spannungskennlinie erfassen, Leistungen von Verbrauchern messen o.ä.). Bei Schülerexperimenten, die sowohl qualitative als auch quantitative Experimente beinhalteten, konnte zudem die Subkategorie *Unklar* vergeben werden. Allerdings geschah das ausschliesslich in vier Fällen (3 Finnland, 1 Deutschland).

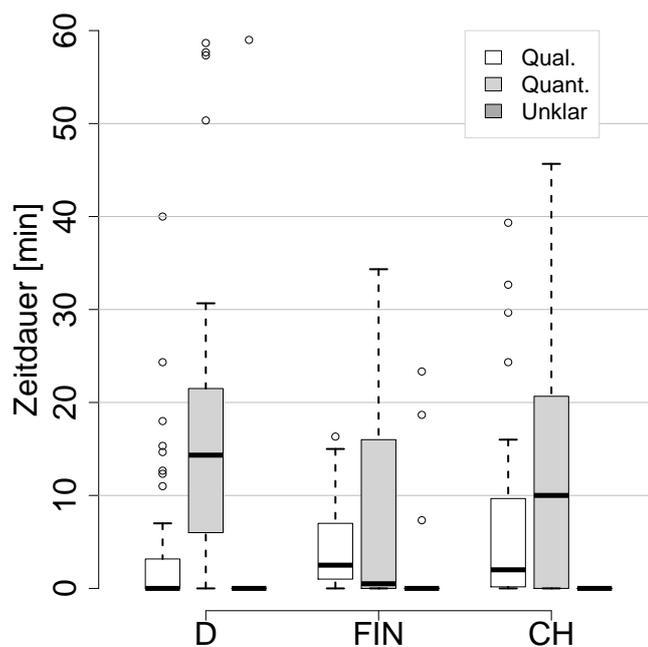


Abbildung 6.3: Zeitdauer pro Doppelstunde, während der qualitative (weiss) bzw. quantitative (hellgrau) Experimente durchgeführt werden. Die Kategorie *Unklar* (im Diagramm durch die Markierung des Medians bei null wiedergegeben; der Interquartilsabstand verschwindet hier) umfasst jenen Teil der *Durchführung*, bei der qualitative und quantitative Experimente gleichzeitig bearbeitet werden.

In Abbildung 6.3 sind die Zeiten, während denen qualitative bzw. quantitative Experimente durchgeführt werden, als Boxplots aufgetragen. Aus der dargestellten Subkategorie *Unklar* mit verschwindendem Interquartilsabstand sind die erwähnten vier Vorkommnisse (Finnland 3, Deutschland 1) ersichtlich (Ausreisser). Die Abbil-

dung zeigt, dass in Deutschland und der Schweiz die *Durchführung* quantitativer Experimente gegenüber den qualitativen zeitlich überwiegt. Signifikant ist der Unterschied jedoch nur in Deutschland, wie aus Tabelle 6.5 hervorgeht. Insbesondere das Resultat in Deutschland bestätigt die Erwartung, dass quantitative Experimente durch ihre zeitlich aufwändige *Datenerfassung* weitaus mehr Unterrichtszeit beanspruchen als qualitative. Ein eher umgekehrtes Verhalten zeigt Finnland: Der Median der Durchführungszeit qualitativer Experimente überwiegt jenen der quantitativen. Allerdings kann hier kein signifikanter Unterschied belegt werden. Verglichen mit dem deutschen Physikunterricht dauert die Durchführung quantitativer Experimente in Finnland signifikant weniger lang, die Durchführung qualitativer Experimente in der Schweiz signifikant länger. Allerdings ist letzterer Unterschied bei einer mittleren Verschiebung der Mediane von einer Minute (vgl. Diff., Tabelle 6.5) klein.

In Tabelle 6.6 sind absolute und relative Häufigkeiten von Doppelstunden mit qualitativen (QL), quantitativen (QT) und Experimenten der Kategorie *Unklar* (UN) aufgelistet. Auch in Bezug auf die Häufigkeiten bestätigt sich für Deutschland die grosse Bedeutung quantitativer Experimente gegenüber qualitativen: Quantitative Experimente können fast doppelt so häufig beobachtet werden wie qualitative.

Variable	W	p	Konf. [min]	Diff. [min]
<i>D-FIN</i>				
Qualitativ	407	.080	[-2, 0]	-1
Quantitativ	776	.002**	[2, 14]	8
<i>D-CH</i>				
Qualitativ	436	.008*	[-4, 0]	-1
Quantitativ	752	.400	[-2, 8]	2
<i>FIN-CH</i>				
Qualitativ	342	.400	[-4, 1]	0
Quantitativ	268	.040	[-12, 0]	-4
<i>Qualitativ-Quantitativ</i>				
D	365	.000**	[-16, -6]	-11
FIN	328	.800	[-4, 2]	0
CH	376	.100	[-11, 0]	-4

Tabelle 6.5: Mittelwertvergleiche zur Kategorie *Datenerfassung*. H_0 -Hypothesen: Die Stichprobenpaare (D-FIN, D-CH, FIN-CH bzw. Qualitativ-Quantitativ) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. werden in Minuten angegeben.

Land	QL	QT	UN
D	17 (0.40)	36 (0.84)	1 (0.02)
FIN	17 (0.77)	11 (0.50)	3 (0.14)
CH	23 (0.74)	22 (0.71)	0 (0.00)
Σ	57 (0.59)	69 (0.72)	4 (0.04)

Tabelle 6.6: Kategorie *Datenerfassung* K1: Absolute und relative (in Klammern) Häufigkeiten der Subkategorien *Qualitativ* (QL), *Quantitativ* (QT) und *Unklar* (UN). Bei den relativen Häufigkeiten entspricht 1 der Anzahl Doppelstunden eines Landes, während denen Experimente durchgeführt werden (D: 43, FIN: 22, CH: 31).

6.3.5 Organisationsform und Arbeitsverteilung

Die Kategorien *Organisationsform* K2a und K2b und *Arbeitsverteilung* K3 gehören inhaltlich zusammen. Deshalb werden sie im Folgenden zusammen ausgewertet. Die Kategorie *Organisationsform* besteht aus den Subkategorien *Demonstration Lehrperson* (DL), *Demonstration Schülerin/Schüler* (DS), *Schülerin/Schüler einzeln* (SE), *Partnerarbeit* (PA) und *Gruppenarbeit* (GR). Für Schülerexperimente wird noch weiter differenziert in die *Durchführung an Stationen* (ST) und einem *arbeitsteiligen Vorgehen* (AT) (Kategorien K2b und K3). Die Bezeichnungen für die Subkategorien entsprechen den gebräuchlichen Fachtermini. Für die Beschreibung der Kategorien bzw. Subkategorien sei auf Abschnitt A.1 verwiesen.

In Abbildung 6.4 sind die drei Subkategorien *Demonstration Lehrperson*, *Partnerarbeit* und *Gruppenarbeit* der Kategorie *Organisationsform* zusammengefasst. Anhand der Boxplots lässt sich ablesen, dass in mehr als dreiviertel der Fälle die *Durchführung* von Demonstrationsexperimenten weniger als zehn Minuten dauert. Dagegen dauert die *Durchführung* von Gruppenarbeiten im deutschsprachigen Unterricht in der Hälfte der Fällen mehr als 10 Minuten. Bei einzelnen Fällen in Deutschland beträgt die *Durchführung* in Gruppen gegen 60 Minuten (ohne *Vor-* und *Nachbereitung!*), also Zweidrittel der Unterrichtszeit. Trotz der Deutlichkeit der erwähnten Unterschiede im Diagramm lassen sich bezogen auf die Länderstichproben keine signifikanten Unterschiede zwischen den beiden Subkategorien belegen. Weiter zeigt sich, dass nur in 9 von insgesamt 96 Fällen Experimente als Partnerarbeit durchgeführt werden (vgl. auch Tabelle 6.7). Zur Erinnerung: Bei Partnerarbeiten teilen sich zwei Lernende ein Experiment. Im Gegensatz dazu setzt sich eine Gruppe bei einer Gruppenarbeit aus drei oder mehr Schülerinnen und Schülern zusammen.

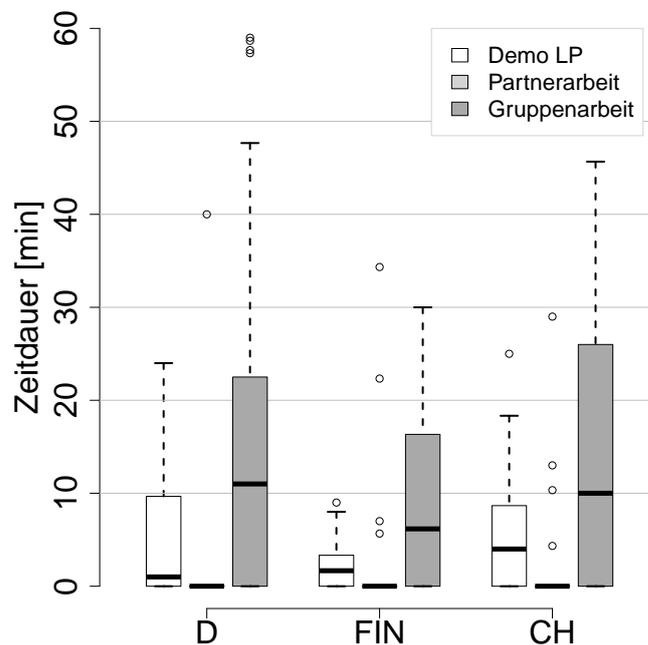


Abbildung 6.4: Zeitdauer, die pro Doppelstunde im Mittel für Demonstrationsexperimente der Lehrperson (weiss), für Schülerexperimente als Partnerarbeit (hellgrau), als Gruppenarbeit (dunkelgrau) aufgewendet wird.

Ergänzend zu Abbildung 6.4 sind in Tabelle 6.7 die absoluten, in Klammern die relativen Häufigkeiten der eben besprochenen sowie der übrigen Subkategorien von *Organisationsform* und *Arbeitsverteilung* aufgeführt. Hierbei fällt auf, dass die Fälle bezüglich Organisationsform und Arbeitsverteilung wenig Vielfalt zeigen: Experimentelles Handeln findet hauptsächlich in der Form einer Lehrer-Demonstration oder einer Gruppenarbeit statt. Eher selten sind Experimente, bei denen Schülerinnen und Schüler ein Experiment vorführen (5 Vorkommnisse), alleine (3) bzw. zu zweit durchführen (9). Bei Schülerexperimentierphasen werden den Gruppen meist feste Experimente zugewiesen. Das Experimentieren an Stationen, welches den Lernenden Wahlmöglichkeiten bezüglich der Reihenfolge der *Durchführung* von Experimenten oder die Möglichkeit, Experimente wegzulassen, bietet, kommt praktisch nicht vor (3 Vorkommnisse).

Des Weiteren sind Schülerexperimente meist arbeitsgleich organisiert; d.h. jede Gruppe bzw. jede Schülerin, jeder Schüler bearbeitet die gleiche Aufgabe. In ausschliesslich 12 der 70 Klassen, bei denen mindestens ein Schülerexperiment (GR, PA und SE) vorkommt, wählt die Lehrperson ein arbeitsteiliges Vorgehen. Der Umstand lässt erahnen, dass das gegenseitige Präsentieren von Ergebnissen und Erfahrungen

beim Experimentieren eher eine untergeordnete Bedeutung hat. Die Subkategorie DL ist in allen Ländern mit relativen Häufigkeiten von 0.53 bis 0.63 ähnlich stark vertreten. Vergleichbar oft kommt auch die Subkategorie GR vor (0.58-0.63). Das bedeutet, dass sich die Lehrenden bei der Gestaltung der Doppelstunde in mehr als der Hälfte aller Fälle für mindestens ein Demonstrationsexperiment bzw. für eine Gruppenarbeit entschieden haben.

Land	DL	DS	SE	PA	GR	ST	AT
D	23 (0.53)	4 (0.09)	2 (0.05)	1 (0.02)	26 (0.60)	2 (0.05)	7 (0.16)
FIN	14 (0.63)	0 (0.00)	0 (0.00)	4 (0.16)	14 (0.63)	0 (0.00)	2 (0.09)
CH	19 (0.61)	1 (0.03)	1 (0.03)	4 (0.18)	18 (0.58)	1 (0.03)	3 (0.10)
Σ	56 (0.58)	5 (0.05)	3 (0.03)	9 (0.09)	58 (0.60)	3 (0.03)	12 (0.13)

Tabelle 6.7: Absolute und relative Häufigkeiten (in Klammern) der Subkategorien: *Demonstration Lehrperson* (DL), *Demonstration Schülerin/Schüler* (DS), *Schülerin/Schüler einzeln* (SE), *Partnerarbeit* (PA), *Gruppenarbeit* (GR), *Stationen* (ST), *Arbeitsteilig* (AT). Relative Häufigkeiten: 1 entspricht der Anzahl Doppelstunden eines Landes, bei denen Experimente stattfinden (D: 43, FIN: 22, CH: 31).

6.3.6 Sachbegegnung

Mit der Kategorie *Sachbegegnung* wird erfasst, welche Materialien bei der Durchführung von Experimenten vorwiegend zum Einsatz kommen. Die Subkategorie *Physikalische Geräte* wird dann einem Zeitintervall zugeordnet, wenn Materialien und Geräte zum Einsatz kommen, die speziell für den Physikunterricht oder für die Laborverwendung hergestellt wurden. Zur Subkategorie *Alltagsgeräte* gehören Intervalle, bei denen Materialien und Geräte zum Einsatz kommen, die auch im eher alltäglichen Leben ihre Verwendung finden. Schliesslich werden mit der Subkategorie *Bildschirmexperiment* Computersimulationen von Experimenten zusammengefasst.

Aus Abbildung 6.5 geht hervor, dass in allen untersuchten Ländern vorwiegend Experimente mit physikalischen Geräten durchgeführt werden. In mehr als der Hälfte der Fälle findet überhaupt keine *Durchführung* mit Alltagsgeräten statt. Die mittlere Durchführungszeit (Median) von Experimenten mit Physikalischen Geräten macht in allen Ländern etwa gleich viel aus und liegt zwischen 10 und 15 Minuten pro Doppelstunde. Die mittlere Durchführungszeit mit Alltagsgeräten unterscheidet sich in allen Ländern hoch bis höchst signifikant von jener mit physikalischen Geräten,

wie Tabelle 6.8 zeigt. Der Vergleich einzelner Subkategorien zwischen den Ländern ergibt jedoch keine signifikanten Unterschiede.

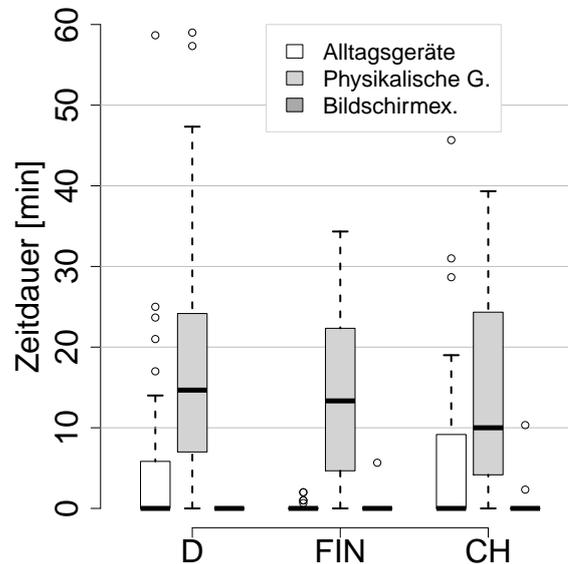


Abbildung 6.5: Zeitdauer, die pro Doppelstunde im Mittel für die *Durchführung* von Experimenten mit Alltagsgeräten (weiss), mit physikalischen Geräten (hellgrau, verschwindender Interquartilsabstand) und am Bildschirm (dunkelgrau) aufgewendet wird.

Die bisherigen Befunde zur Nutzung von Alltagsgeräten bzw. physikalischen Geräten bestätigen sich weiter mit Blick auf Tabelle 6.9. In allen Ländern ist die Verwendung physikalischer Geräte beim Experimentieren deutlich häufiger als der Einsatz von Alltagsgeräten. Der Unterschied erscheint jedoch etwas weniger deutlich als in Abbildung 6.5. Das hängt damit zusammen, dass Alltagsgeräte häufiger im Zusammenhang mit, gegenüber den Schülerexperimenten generell kürzer dauernden, Demonstrationsexperimenten eingesetzt werden. Wie sich auch schon aus Abbildung 6.5 ablesen lässt, kommen insgesamt nur drei Fälle mit Bildschirmexperimenten vor (1 FIN, 2 CH).

Variable	W	p	Konf. [min]	Diff. [min]
<i>Alltagsgeräte-Physikalische Geräte</i>				
D	387	.0000***	[-15, -7]	-12
FIN	63	.0000***	[-17, -5]	-8
CH	259	.0012**	[-14, -4]	-7

Tabelle 6.8: Mittelwertsvergleiche zur Kategorie *Sachbegegnung*. H_0 -Hypothese: Die Stichprobenpaare (Alltagsgeräte-Physikalische Geräte) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. werden in Minuten angegeben.

Land	PH	AL	BI
D	35 (0.81)	15 (0.35)	0 (0.00)
FIN	21 (0.95)	5 (0.23)	1 (0.05)
CH	25 (0.81)	11 (0.35)	2 (0.06)
Σ	81 (0.84)	31 (0.32)	3 (0.03)

Tabelle 6.9: Kategorie *Sachbegegnung* K5: Absolute und relative (in Klammern) Häufigkeiten der Subkategorien Physikalische Geräte (PH), Alltagsgeräte (AL) und Bildschirmexperiment (BI). Relative Häufigkeiten: 1 entspricht der Anzahl Doppelstunden eines Landes, bei denen Experimente stattfinden (D: 43, FIN: 22, CH: 31).

6.3.7 Die Experimentiereinheit

Bisher wurden die im Kontext experimentellen Handelns stehenden Zeiten summarisch ausgewiesen (vgl. Tabelle 6.2). Dabei fehlen Angaben über die Anzahl der Experimente pro Doppelstunde sowie die pro Experiment aufgewendete Zeit. Aufgrund der bisher vorgestellten Ergebnisse stellen sich folgende Fragen: Lässt sich der hohe Anteil experimentellen Handelns im deutschen Physikunterricht durch eine gegenüber Finnland und der Schweiz höhere Zahl von Experimentiereinheiten pro Doppelstunde erklären? Oder sind vielmehr die Unterschiede in der mittleren Dauer der Experimentiereinheiten für die Länderunterschiede verantwortlich?

Land	Anz. Experimentiereinheiten							
	0	1	2	3	4	5	6	7
D	0	21	9	8	1	1	1	2
FIN	3	7	6	6	2	1	0	0
CH	0	7	8	11	5	0	0	0
Σ	3	35	23	25	8	2	1	2

Tabelle 6.10: Anzahl der Fälle, bei denen pro Doppelstunde insgesamt n ($n = 0, 1, 2, \dots, 7$) Experimentiereinheiten durchgeführt wurden.

Für die Beantwortung der Fragen werden die Kategorien *Experimentierphasen* E1 und *Nummerierung Experimentiereinheiten* E2 herangezogen. Aus Tabelle 6.10 ist ersichtlich, wie viele Experimente pro Doppelstunde durchgeführt wurden. Mit dem exakten Test nach Fisher kann gezeigt werden, dass das Merkmal *Anzahl der Experimentiereinheiten* vom Merkmal *Land* abhängt ($p = .04$, $\alpha = .05$). Somit lassen sich Unterschiede in der Häufigkeit der Experimente zwischen den Ländern belegen. Der paarweise Ländervergleich ergibt jedoch nur signifikante Unterschiede zwischen der Schweiz und Deutschland ($p = .03$). Median und Modus der Häufigkeiten betragen: D und FIN: Median 2, Modus 1; CH Median 3, Modus 3.

In Abbildung 6.6 sind die Länderverteilungen der Dauer einer Experimentiereinheit dargestellt. Hier zeigen sich höchst signifikante Unterschiede zwischen Deutschland und Finnland (vgl. Tabelle 6.11): In Deutschland dauert eine Experimentiereinheit höchst signifikant länger (35min) als in Finnland (15min). Die Signifikanz bleibt auch reduziert auf die Organisationsformen Gruppenarbeit (D 61min, FIN 23min) und Lehrer-Demonstration (D 12min, FIN 5min) bestehen. Zwischen Deutschland und der Schweiz sind mit Ausnahme der Lehrer-Demonstrationen ebenfalls signifikante Unterschiede auszumachen. Eine Gruppenarbeit (inkl. Vor- und Nachbe-

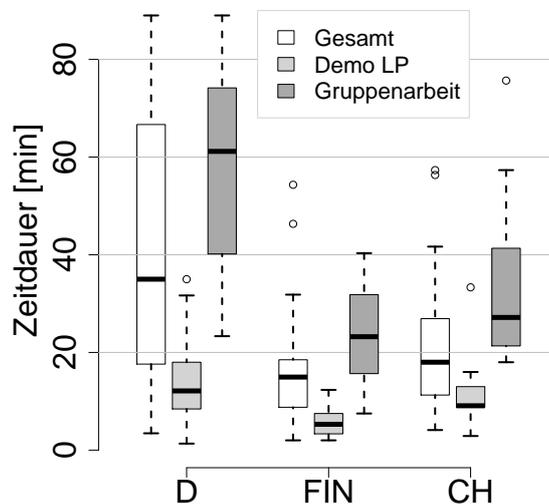


Abbildung 6.6: Die Zeitdauer, die im deutschen, finnischen und Schweizer Physikunterricht pro Experimentiereinheit gesamthaft (weiss), als Lehrer-Demonstration (hellgrau) und als Experimentiereinheit in Gruppen (dunkelgrau) aufgewendet wird.

reitungszeit) dauert zwischen drei (CH) bis fünf mal (D) länger als eine Lehrer-Demonstration. Bezogen auf die ausgangs gestellten Fragen lässt sich festhalten, dass sich die Unterschiede bezüglich der im Kontext experimentellen Handelns stehende Unterrichtszeit sowohl aus den abweichenden Häufigkeiten als auch aus der unterschiedlichen mittleren Dauer der Experimentiereinheiten zusammensetzen: Das gilt insbesondere für die Länderpaare Deutschland-Schweiz und Finnland-Schweiz. Zwischen Deutschland und Finnland scheint hauptsächlich die verschieden grosse Dauer der Experimentiereinheiten für das Ergebnis in Abbildung 6.1 verantwortlich zu sein.

Vergleich	W	p	Konf. [min]	Diff. [min]
<i>D-FIN</i>				
Lehrer Demonstration	268	0.0002***	[4, 12]	7
Gruppenarbeit	315	0.0000***	[23, 48]	36
Gesamt	739	0.0002***	[8, 35]	19
<i>D-CH</i>				
Lehrer Demonstration	247	0.2	[-1, 7]	2
Gruppenarbeit	326	0.0002***	[13, 42]	28
Gesamt	930	0.004*	[4, 28]	14
<i>FIN-CH</i>				
Lehrer Demonstration	36	0.0007**	[-7, -2]	-5
Gruppenarbeit	81	0.2	[-17, 3]	-6
Gesamt	244	0.08	[-10, 1]	-5
<i>Lehrer Demonstration-Gruppenarbeit</i>				
D	8	0.0000***	[-57, -34]	-47
FIN	6	0.0000***	[-26, -11]	-19
CH	11	0.0000***	[-26, -12]	-17

Tabelle 6.11: Mittelwertsvergleiche zur Kategorie *Organisationsform*. H_0 -Hypothesen: Die Stichprobenpaare (D-FIN, D-CH, FIN-CH oder Lehrer Demonstration-Gruppenarbeit) unterscheiden sich nicht bezüglich ihrer zentralen Tendenz. Die Spalten Konf. und Diff. sind in Minuten pro Experimentiereinheit zu verstehen.

6.4 Schultypenvergleich

Im Folgenden wird die Stichprobe nach Unterschieden bezüglich der Schultypen untersucht. Da das finnische Schulsystem nur eine Gesamtschule kennt, wird diese Landerstichprobe bei der Auswertung nicht berucktigt. Die Klassen aus Deutschland und der Schweiz werden fur die Analyse zusammen als eine Stichprobe betrachtet.

In Abbildung 6.7 ist die im Kontext experimentellen Handelns stehende Unterrichtszeit (Gesamt) sowie die pro Doppelstunde eingesetzte Vorbereitungs-, Durchfuhrungs- und Nachbereitungszeit fur die drei Schultypen aufgefuhrt. Sowohl gesamthaft als auch fur die einzelnen Phasen lassen sich keine signifikanten Landerunterschiede festmachen. Das gilt auch fur die ubrigen Variablen. Insbesondere lassen sich auch keine Unterschiede bei den Haufigkeiten oder den mittleren Dauern von Demonstrationsexperimenten bzw. Experimenten in Gruppen nachweisen.

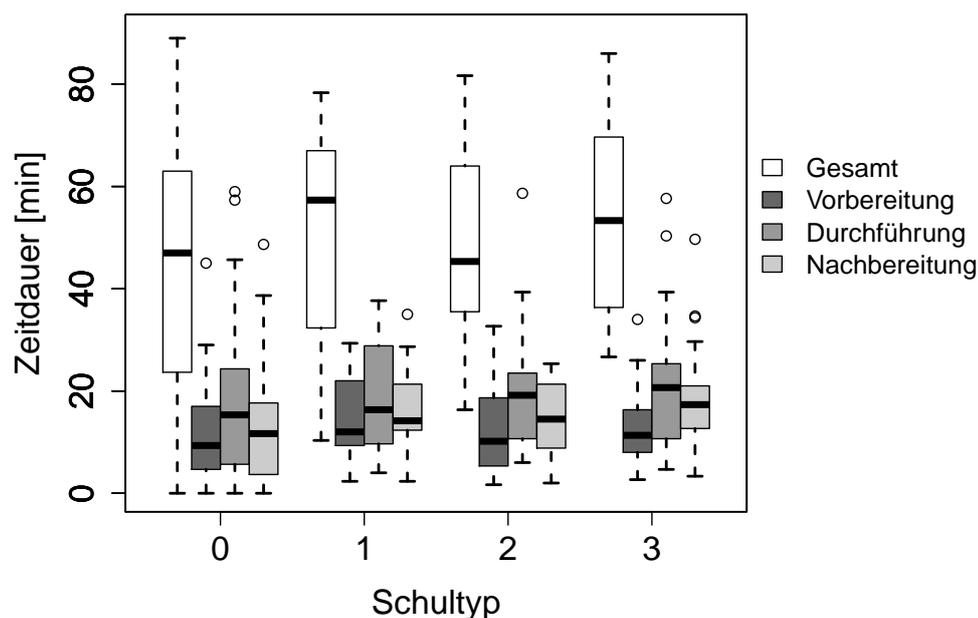


Abbildung 6.7: Die Zeitdauer pro Doppelstunde die in den Schultypen 0-3 fur experimentelles Handeln gesamthaft (weiss), fur die Vorbereitung (dunkelgrau), *Durchfuhrung* (grau) und *Nachbereitung* (hellgrau) eingesetzt wird (Stichproben D und CH zusammengenommen). Typ 1, tiefstes; 2, mittleres und 3 hochstes Leistungsniveau (Typ 0, typengemischte Klasse).

Kapitel 7

Analyse der Tiefenstruktur experimentellen Handelns

Im vorliegenden Kapitel wird die Analyse der Tiefenstruktur experimentellen Handelns vorgestellt. Anders als die Sichtstruktur bezieht sich die Tiefenstruktur auf Aspekte, welche sich nicht alleine durch die Registrierung von Vorkommnissen bei der Beobachtung des Unterrichts erschliessen lassen. Die Entschlüsselung der Tiefenstruktur erfordert von den beurteilenden Experten somit eine verstärkte Interpretationsleistung. Während die Analyse der Sichtstruktur auf eher wertfreien Merkmalen beruht, ist mit der Analyse der Tiefenstruktur eine Bewertung verbunden. Zur Entschlüsselung der Tiefenstruktur von Unterrichtsvideos, insbesondere der Experimentiereinheiten, sind verschiedene methodische Zugänge denkbar. Hier wurde als Vorgehen die Tiefenstrukturanalyse einer Teilstichprobe auf Basis des in Abschnitt 5.2 beschriebenen hoch-inferenten Ratinginstruments gewählt.

Das Vorgehen ermöglicht die Beurteilung nach vorgegebenen Kategorien und damit die Identifikation von ‘good-practice’ Beispielen, sowie die Untersuchung von Länderspezifika. Theoretisch basiert das entwickelte Analyseinstrument auf drei Perspektiven (vgl. Abbildung 7.1): Experimentelles Handeln als kontextorientierte Aktivität, als reflexive Aktivität und als theoriegeleitete Aktivität. Im Folgenden werden die Perspektiven vornehmlich mit den Begriffen *Kontextorientierung*, *Reflexivität* und *Theorieleitung* betitelt.

Die eben beschriebene kategoriale, hoch-inferente Analyse der Unterrichtsvideos wird mit Analysen von Lehrertranskripts ausgewählter Fälle¹ ergänzt. Die Unter-

¹Zum Vorgehen der Fallauswahl siehe Abschnitt 5.4.

suchung der Transkripts basiert auf der qualitativen Inhaltsanalyse von Mayring (2002) (vgl. Abschnitt 5.4). Die qualitative Inhaltsanalyse der Lehrertranskripts hat einerseits zum Ziel, Beispiele experimentellen Handelns zu illustrieren und dient andererseits der Validierung des Ratinginstruments und der Generierung neuer Kategorien (vgl. Kapitel 2).

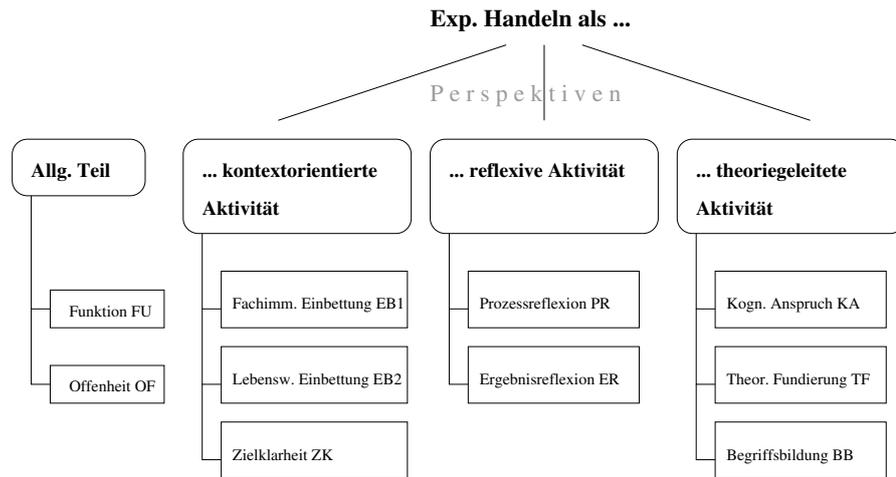


Abbildung 7.1: Die Struktur des Ratinginstruments besteht aus den drei Perspektiven experimentellen Handelns und einem allgemeinen Teil (abgerundete Rechtecke). Den Subkategorien bzw. Facetten (Rechtecke) entsprechen Items, die auf einer 4-stufigen Likert-Skala beurteilt werden.

Klassen bzw. Fälle erhielten im Rahmen der Untersuchung eine Identifikationsnummer (ID). Sie setzt sich aus dem Länderkürzel (D: Deutschland, F: Finnland und S: Schweiz), dem Schultyp (0: typengemischte Klasse, 1: unteres Niveau, 2: mittleres Niveau und 3: oberes Niveau), der Schulnummer (Laufnummer in jedem Land bei 1 beginnend) sowie der Klassennummer (Laufnummer) zusammen. Ein Unterrichts-video mit der ID S-1-05-01 verweist somit auf eine Schweizer Klasse des unteren Niveaus.

7.1 Analyse auf Klassenebene

Die Analyse auf Klassenebene ergibt sich aus der summarischen Beurteilung aller Experimentiereinheiten einer Klasse. Die verwendete Methodik wurde bereits in Abschnitt 5.2 ausgeführt. Dieser zufolge werden die Facetten der drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung* für die ganze Klasse eingeschätzt.

Die Perspektive *Kontextorientierung* zeigt auf, inwiefern Experimentiereinheiten im Unterricht mit Fachinhalten oder mit lebensweltlichen Aspekten verbunden werden. Ergänzend dazu eröffnet die Perspektive *Reflexivität* Einblicke in das kommunikative Verhalten des Lehrenden und seine Fähigkeit, während der Experimentierphasen Freiräume für Metakognitionen zu schaffen. Als dritte Perspektive zeigt die *Theorieleitung*, wie stark qualitative Erklärungsansätze mit dem experimentellen Handeln verbunden werden und von welcher sprachlichen Qualität die dabei verwendeten Formulierungen sind. Die in Abbildung 7.1 ebenfalls angeführte Kategorie *Allg. Teil* geht nicht in die Analyse auf der Klassenebene ein. Die Auswertung der damit assoziierten Subkategorien *Funktion* und *Offenheit* wird im nächsten Abschnitt besprochen. Im Zusammenhang mit der Analyse auf der Klassenebene werden folgende Forschungsfragen untersucht (vgl. Abschnitt 2.3):

T1 Findet sich die Qualität experimentellen Handelns, aufgespannt durch die drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung*, in der untersuchten Teilstichprobe wieder?

T2 Lassen sich bezüglich der Perspektiven oder Facetten Unterschiede zwischen den Ländern nachweisen?

In Abbildung 7.2 sind die Daten zu jeder Facette auf Klassenebene zusammengefasst, wobei *weiss=trifft zu*, *hellgrau=trifft grösstenteils zu*, *dunkelgrau=trifft teilweise zu* und *schwarz=trifft nicht zu* entsprechen. Die Fälle sind so geordnet, dass Klassen mit hohen Qualitätsmerkmalen experimentellen Handelns ganz rechts zu finden sind.

Das eher dunkle Bild lässt bereits einige grundsätzliche Schlüsse zu: Die in den Facetten definierten Merkmale für Qualität sind generell wenig ausgeprägt in der untersuchten Stichprobe zu finden. Besonders gilt das für die Facetten *Lebensweltliche Einbettung* (EB2) und *Theoretische Fundierung* (TF), weniger für die Facette *Begriffsbildung* (BB). Vom allgemeinen Eindruck unabhängig, lassen sich jedoch einzelne Fälle identifizieren, bei denen mehrere Facetten ausgeprägt vorkommen. So z.B. in S-1-05-01, F-0-16-01, S-3-31-01 und D-3-22-01. Somit findet sich die durch das Instrument definierte Qualität experimentellen Handelns in einzelnen Fällen der untersuchten Stichprobe wieder. Vier ausgewählte Experimentiereinheiten der Doppelstunden am rechten Rand werden in einer nachfolgenden Fallanalyse genauer untersucht (vgl. Abschnitt 7.3).

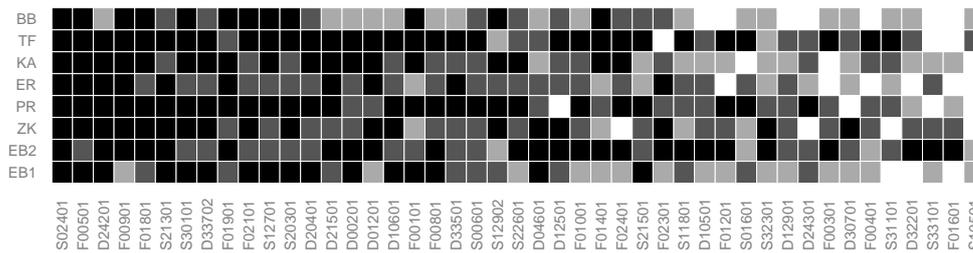


Abbildung 7.2: Überblick über die Ergebnisse der Tiefenstruktur: Zu den acht Facetten der drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung* sind die Ergebnisse für jede Klasse zusammengefasst. *Weiss=trifft zu*, *hellgrau=trifft grösstenteils zu*, *dunkelgrau=trifft teilweise zu* und *schwarz=trifft nicht zu*. Die Klassen sind entsprechend ihres Rating-Summenwerts geordnet.

Ausgehend von dem ersten Überblick folgt anhand der Tabelle 7.1 und den Abbildungen 7.3-7.5 eine detailliertere Analyse der Ergebnisse. Wie Tabelle 7.1 zeigt, liegen die Mediane bei fast allen Facetten zwischen 3 (trifft teilweise zu) und 4 (trifft nicht zu). Eine Ausnahme bildet einzig die Facette *Begriffsbildung*. Hier beträgt der Median 2 (trifft grösstenteils zu). Die Breite des 95%-Konfidenzintervalls des Medians liegt durchwegs zwischen 0.4 und 1. Im analysierten Physikunterricht haben damit die drei Perspektiven experimentellen Handelns, beschrieben durch ihre Facetten, eine untergeordnete Bedeutung. Allerdings weist eine namhafte Minderheit der Fälle hohe Ausprägungen bezüglich einzelner Facetten auf. Den Abbildungen 7.3-7.5 ist zu entnehmen, dass Fälle mit Ausprägungen von 1 (trifft zu) oder 2 (trifft grösstenteils zu) mit folgenden Häufigkeiten auftreten: *Fachimmanente Einbettung* 19 (42%), *Lebensweltliche Einbettung* 4 (9%), *Zielklarheit* 8 (18%), *Prozessreflexion* 6 (13%), *Ergebnisreflexion* 13 (29%), *Kognitiver Anspruch* 14 (31%), *Theoretische Fundierung* 5 (11%) und *Begriffsbildung* 24 (53%). Es gibt somit durchaus Fälle, die den vorausgesetzten Qualitätsmerkmalen partiell genügen.

Aus den Abbildungen 7.3-7.5 werden zudem länderspezifische Unterschiede bei den Facetten *Fachimmanente Einbettung*, *Lebensweltliche Einbettung*, *Prozess Reflexion* und *Begriffsbildung* ersichtlich. Bei der *Fachimmanenten Einbettung* liegt der Median bei Finnland bei 2 (trifft grösstenteils zu), während er bei Deutschland und der Schweiz bei 3 (trifft teilweise zu) liegt. Dagegen scheinen Experimente in der Schweiz etwas besser lebensweltlich eingebettet zu sein als jene im deutschen und im finnischen Unterricht (Median (Med.) CH 3, FIN & D 4). Die Facetten *Prozess*

Facette	Med	Mod	IQR	Konf.
Fachimm. Einbett.	3	-	2	[2.5, 3.5]
Lebenswelt. Einbett.	4	4	1	[3.8, 4.2]
Zielklarheit	3	3	1	[2.8, 3.2]
Prozess Reflexion	4	4	1	[3.8, 4.2]
Ergebnis Reflexion	3	3	1	[2.8, 3.2]
Kogn. Anspruch	3	4	2	[2.5, 3.5]
Theor. Fundierung	4	4	1	[3.8, 4.2]
Begriffsbildung	2	2	2	[1.5, 2.5]

Tabelle 7.1: Ergebnisse der Tiefenstrukturanalyse auf Klassenebene. Zu jeder Facette sind angegeben: Median (Med), Modalwert (Mod), Interquartilsabstand (IQR) sowie das 95%-Konfidenzintervall des Medians (Konf.).

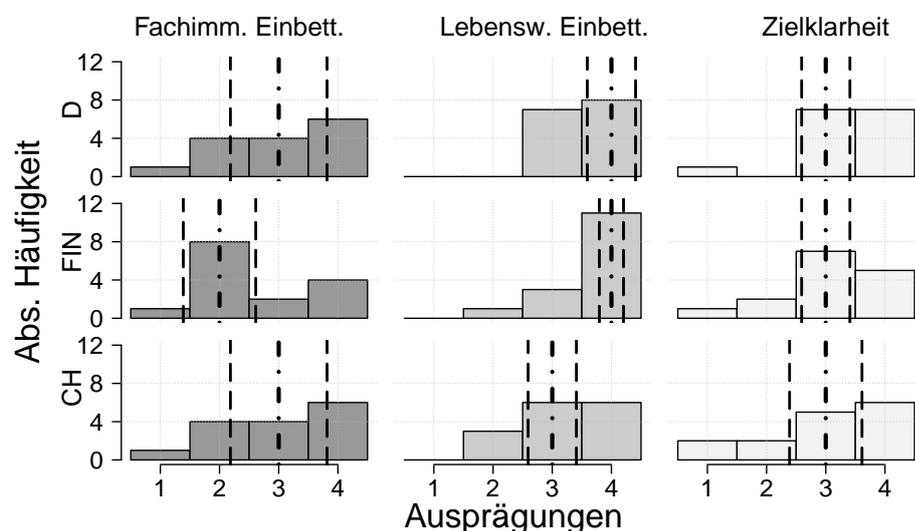


Abbildung 7.3: Kontextorientierung und assoziierte Facetten wie *Fachimmanente Einbettung*, *Lebensweltliche Einbettung* und *Zielklarheit*: Häufigkeitsverteilungen der Gesamtratings pro Land und Facette. Ausprägungen: 1=trifft zu, 2=trifft grösstenteils zu, 3=trifft teilweise zu und 4=trifft nicht zu. Median, strichpunktierte Linie; Grenzen des 95%-Konfidenzintervalls des Medians, gestrichelte Linien.

Reflexion und *Begriffsbildung* sind im Vergleich zu den übrigen Ländern ausgeprägter in Deutschland vorzufinden (Med. D 3/2, FIN & CH 4/3). Mit Ausnahme der Facette *Begriffsbildung* sind sämtliche Länderunterschiede nicht signifikant (Begriffsbildung: D-FIN $W = 66$ $p = .05$; D-CH $W = 57$, $p = .02$; FIN-CH $W = 116$, $p = .9$).

Was bedeuten die gefundenen Mediane der Ratings für den beobachteten Unterricht? Im Folgenden wird der Frage für jede Facette einzeln nachgegangen:

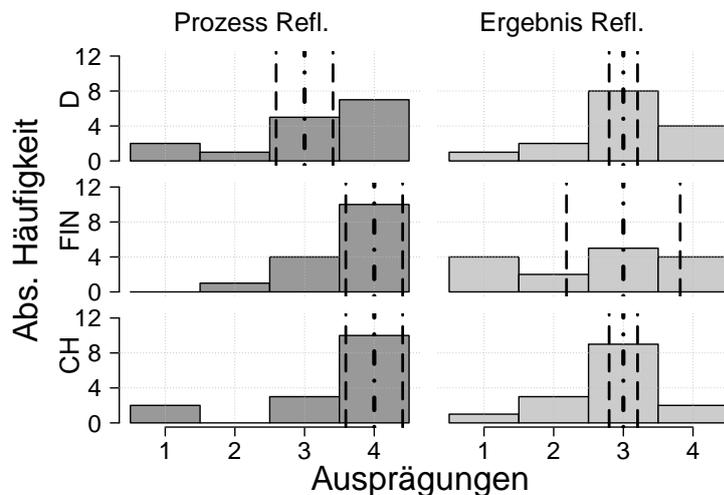


Abbildung 7.4: Reflexivität und assoziierte Facetten wie *Prozess Reflexion* und *Ergebnis Reflexion*: Häufigkeitsverteilungen der Gesamtratings pro Land und Facette. Ausprägungen: 1=trifft zu, 2=trifft grösstenteils zu, 3=trifft teilweise zu und 4=trifft nicht zu. Median, strichpunktierte Linie; Grenzen des 95%-Konfidenzintervalls des Medians, gestrichelte Linien.

Fachimmanente Einbettung, Median (Med) 3, Interquartilsabstand (IQR)

2. Die Lehrperson bettet das Experiment fachinhaltlich teilweise ein. Typischerweise werden dabei relevante Fachinhalte aufgegriffen, diese jedoch nicht oder in ungenügender Masse mit dem bevorstehenden Experiment verbunden. Meist beschränkt sich die Phase, in der Ansätze einer Einbettung zu beobachten sind, auf den Beginn der Experimentiereinheit. Der Aspekt der inhaltlichen Vorbereitung kommt generell zu kurz.

Lebensweltliche Einbettung, Med 4, IQR 1. Eine lebensweltliche Einbettung findet nicht statt. Das bedeutet in den meisten Fällen, dass sich die Lehrperson an einer Fachsystematik orientiert und dabei auf eine Darlegung der lebensweltlichen Relevanz des Themas verzichtet. Dadurch fehlen wichtige Anknüpfungspunkte, die dazu beitragen, dass Schülerinnen und Schüler Experimente motivierend erleben.

Zielklarheit, Med 3, IQR 1. Die Experimentiereinheit wird von der Lehrperson teilweise zielklar angeleitet. Demzufolge wird die mehrheitlich vorhandene Problem- bzw. Fragestellung nur ansatzweise klar dargestellt. Der Mangel an Klarheit kann durch die Formulierung oder die mangelnde Begründung der Problemstellung be-

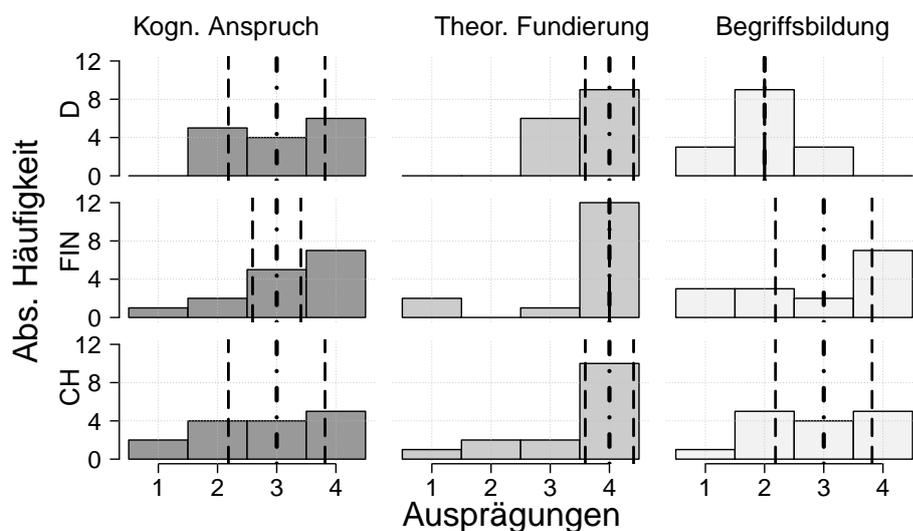


Abbildung 7.5: Theorieleitung und assoziierte Facetten wie *Kognitiver Anspruch*, *Theoretische Fundierung* und *Begriffsbildung*: Häufigkeitsverteilungen der Gesamtratings pro Land und Facette. Ausprägungen: 1=trifft zu, 2=trifft grösstenteils zu, 3=trifft teilweise zu und 4=trifft nicht zu. Median, strichpunktierte Linie; Grenzen des 95%-Konfidenzintervalls des Medians, gestrichelte Linien.

dingt sein. Sehr häufig führt auch eine kleinschrittige Vorgehensweise beim experimentellen Handeln zu Defiziten bei der Zielklarheit.

Prozess Reflexion, Med 4, IQR 1. Die Facette *Prozess Reflexion* findet sich in den analysierten Fällen kaum. D.h., eine Mehrheit der Lehrpersonen ermutigt Schülerinnen und Schüler nicht dazu, ihr Handeln und ihre Ideen zu reflektieren. Weiter führt das Frageverhalten der Lehrpersonen meist zu stichwortartigen Antworten der Lernenden. Anstelle von inhaltlichen Rückfragen bei Kurzantworten, stehen oft Ausführungen der Lehrperson. Zudem antizipiert die Lehrperson kaum die üblichen kognitiven Prozesse der Lernenden.

Ergebnis Reflexion, Med 3, IQR 1. Eine Reflexion der Ergebnisse findet teilweise statt. Zwar werden Ergebnisse zusammengefasst, dabei treten jedoch Schwierigkeiten auf. Zu nennen sind in dem Zusammenhang: die übereilt abgeschlossene Zusammenfassung der Ergebnisse infolge von Zeitknappheit, das Offenlassen der ursprünglichen Problemstellung sowie die ungenügende Ergebnissicherung.

Kognitiver Anspruch, Med 3, IQR 2. Die untersuchten Experimentiereinheiten sind teilweise kognitiv anspruchsvoll. Unter einem adäquaten kognitiven Anspruch wird eine stimmige Passung zwischen Anspruchsniveau des Experiments und den Schülerfähigkeiten und Interessen verstanden. Die teilweise Erfüllung der Ausprägung deutet darauf hin, dass das Experiment die Schülerinnen und Schüler zu wenig fordert, oder diese überfordert. Häufig geht eine Überforderung mit Geräten und Messinstrumenten zusammen mit einer Unterforderung bei der theoretischen Anlage des Experiments einher.

Theoretische Fundierung, Med 4, IQR 1. Die meisten Experimentiereinheiten sind nicht theoretisch fundiert. Damit ist gemeint, dass Experimente nur selten mit einer Idee bzw. einem Erklärungsmodell verbunden werden. Physikalische Vorgänge, die sich bei einem Experiment abspielen, werden somit nicht qualitativ beschrieben. Dagegen dominiert die quantitative Beschreibung durch Gesetzmässigkeiten.

Begriffsbildung, Med 2, IQR 2. Die Lehrperson setzt die Sprache im Zusammenhang mit experimentellem Handeln grösstenteils bewusst und reflektiert ein. Sie leitet Schülerinnen und Schüler in den meisten Fällen zu einer adäquaten Ausdrucksweise an und reflektiert die unterschiedliche Bedeutung von Termini in Fach- und Alltagssprache.

7.2 Analyse auf der Ebene einzelner Experimentiereinheiten

Bisher wurden die Ergebnisse auf der Ebene der Klassen ausgewiesen (Gesamtratings). Diese gehen aus der summarischen Beurteilung aller Experimentiereinheiten einer Klasse hervor (vgl. Abschnitt 5.2). Der aktuelle Abschnitt ist der Analyse der 68 Experimentiereinheiten aller 45 Klassen gewidmet. Hierbei wird der Zusammenhang zwischen dem Gesamtrating und den Ratings der Experimentiereinheiten untersucht. Zudem soll der allgemeine Teil mit den Kategorien *Funktion* und *Offenheit* deskriptiv analysiert werden. Die erwähnten Kategorien finden aufgrund des Skalenniveaus (nominal, binär, vgl. Tabelle 5.4) keinen Eingang im Gesamtrating. Folgende Fragen sind Gegenstand der Untersuchung (vgl. Abschnitt 2.3):

T3 Wie verhalten sich die Ratings einzelner Experimentiereinheiten innerhalb einer Klasse zu ihrer summativen Beurteilung (Gesamtrating)?

T4 Welche Merkmale der Kategorien *Funktion* und *Offenheit* lassen sich in den Experimentiereinheiten identifizieren?

Aus den Ergebnissen wird ersichtlich, dass die Ratings einzelner Experimentiereinheiten einer Klasse meist wenig variieren. In Tabelle 7.2 sind die Häufigkeiten zu den verschiedenen Spannweiten der Einzelratings zusammengefasst. So zeigt sich, dass sich in 67% bis 87% der Klassen die Ratings einer Facette über die verschiedenen Experimentiereinheiten nicht verändern ($\Delta = 0$). Besonders stabil verhalten sich die Ratings bei der *Begriffsbildung* (87%). Die häufigsten Veränderungen innerhalb einer Klasse lassen sich bei der Facette *Kognitiver Anspruch* (67%) identifizieren. Hier wird die unterschiedlich starke Anbindung der Facetten an Persönlichkeitsmerkmale der Lehrperson erkennbar: Während der *Kognitive Anspruch* deutlich von der einzelnen Experimentiereinheit und weniger von der Lehrerpersönlichkeit abhängt, kann bei der Facette *Begriffsbildung* von einem umgekehrten Verhältnis ausgegangen werden. Aufgrund der insgesamt eher geringen Unterschiede bei den Ratings verschiedener Experimentiereinheiten einer Klasse wird im Folgenden auf eine Auswertung bezüglich der drei Perspektiven auf der Ebene einzelner Experimentiereinheiten verzichtet.

Es ist auf der Ebene der Experimentiereinheiten hingegen wichtig zu erfahren, welche fachdidaktischen Funktionen die jeweiligen Experimente im Unterricht übernehmen. Die Ergebnisse hierzu sind in Tabelle 7.3 zusammengefasst. Daraus wird

Facette	Spannweite Δ			
	0	1	2	3
Fachimmanente Einbettung	33 (73)	7 (16)	4 (9)	1 (2)
Lebensweltliche Einbettung	33 (73)	6 (13)	6 (13)	0 (0)
Zielklarheit	32 (71)	11 (24)	2 (4)	0 (0)
Prozess Reflexion	33 (73)	9 (20)	3 (7)	0 (0)
Ergebnis Reflexion	36 (80)	6 (13)	1 (2)	2 (4)
Kognitiver Anspruch	30 (67)	8 (18)	7 (16)	0 (0)
Theoretische Fundierung	37 (82)	3 (7)	4 (9)	1 (2)
Begriffsbildung	39 (87)	5 (11)	1 (2)	0 (0)

Tabelle 7.2: Spannweite Δ der Ratings aller Experimentiereinheiten einer Klasse. Aufgetragen werden absolute und relative Häufigkeiten () (in % bezogen auf die 45 Klassen der Teilstichprobe).

ersichtlich, dass bei drei von vier Experimenten Merkmal FU3 (Durchführen einer oder mehrerer Messungen) oder FU5 (Eine physikalische Gesetzmässigkeit plausibel machen) zugeordnet wurde. Eine Experimentiereinheit des Merkmals FU3 umfasst eine oder mehrere Messungen, die nicht mit einer Gesetzmässigkeit, einem Modell oder einer Theorie in Verbindung gebracht werden.

Typisch für das Merkmal FU3 ist die Experimentiereinheit 1 von D-1-25-01: Im Rahmen eines Demonstrationsexperiments misst der Lehrer die Leistung beim Einsatz verschiedener Geräte, wie z.B eines Fernsehers, einer Herdplatte oder einer Bohrmaschine. Anhand der Messungen erstellt die Lehrperson zusammen mit der Klasse eine Rangordnung der Geräte.

Bei einer Experimentiereinheit mit dem Merkmal FU5 steht das Plausibel machen einer Gesetzmässigkeit im Zentrum. Mit der Einheit 2 von S-2-03-01 lässt sich das gut veranschaulichen: Schülerinnen und Schüler erwärmen Wasser mit einem Tauchsieder und messen dabei die zugeführte elektrische Leistung sowie den Temperaturanstieg. Es wird die Frage untersucht, wie viel Energie benötigt wird, um einen Liter Wasser um ein Grad Celsius zu erwärmen.

Während drei Viertel der Experimente zu den Merkmalen FU3 und FU5 gehören, lässt sich ein Viertel der Experimente den Merkmalen FU2, FU4 und FU6 zu etwa gleichen Teilen zuordnen (FU2 Das Lösen bzw. Erkennen eines technischen Problems, FU4 Das Kennenlernen eines Phänomens, FU6 Ein physikalisches Konzept veranschaulichen). Die Merkmale FU1, FU7 und FU8 kommen nicht bzw. praktisch nicht vor (FU1 Das Kennenlernen eines Laborgerätes oder Messinstruments,

FU7 Das Erlangen eines besseren allgemeinen Verständnisses von den Arbeits- und Funktionsweisen der Naturwissenschaft, FU8 Anderes).

Merkmale der Kategorie <i>Funktion</i>		n
FU1	Das Kennenlernen eines Laborgerätes oder Messinstruments	1
FU2	Das Lösen bzw. Erkennen eines technischen Problems	5
FU3	Durchführen einer oder mehrerer Messungen	25
FU4	Das Kennenlernen eines Phänomens	4
FU5	Eine physikalische Gesetzmässigkeit plausibel machen	27
FU6	Ein physikalisches Konzept veranschaulichen	6
FU7	Das Erlangen eines besseren allgemeinen Verständnisses von den Arbeits- und Funktionsweisen der Naturwissenschaft	0
FU8	Anderes	0
Anzahl Experimentierheiten insgesamt		68

Tabelle 7.3: Anzahl der Experimentiereinheiten n, die den Merkmalen FU1 bis FU8 zugeordnet sind.

Die Kategorie *Funktion* und die Facetten der drei Perspektiven sind nicht unabhängig voneinander: Experimente des Merkmals FU6 weisen signifikant häufiger hohe Ausprägungen bei den Facetten *Prozess Reflexion*, *Kognitiver Anspruch* und *Theoretische Fundierung* auf als jene, die den Merkmalen FU3 bis FU5 zugeordnet wurden. Das wird durch den exakten Fisher-Test der drei 4-Feldtafeln in Tabelle 7.4 belegt. Dafür wurden die Ausprägungen der Facetten dichotomisiert und die Merkmale der Kategorie *Funktion* umkodiert. Für die übrigen Facetten lässt sich kein Zusammenhang zu den beiden Merkmalen FU3-5 (Zusammenfassung der Merkmale FU3, FU4 und FU5) und FU6 nachweisen.

	PR		KA		TF	
	FU3-5	FU6	FU3-5	FU6	FU3-5	FU6
trifft zu	4	4	14	5	4	4
trifft nicht zu	52	2	42	1	52	2
odds ratio	23.1		14.3		23.1	
<i>p</i>	.002		.008		.002	

Tabelle 7.4: Zusammenhang zwischen der *Funktion* der Experimentiereinheit und den Facetten *Prozess Reflexion*, *Kognitiver Anspruch* und *Theoretische Fundierung*. Die Daten wurden auf die Form einer 4-Feldertafel mit den Faktoren Facette und Funktion umkodiert. Alternativ-Hypothese H_1 : *odds ratio* ist grösser als 1. *p* gibt die Irrtumswahrscheinlichkeit für das Eintreten von H_1 an. $\alpha = .05$.

Die Offenheit experimentellen Handelns drückt aus, inwiefern Schülerinnen und Schüler auf Entscheidungen und Handlungen im Zusammenhang mit dem Experiment Einfluss nehmen können (vgl. Ratingsystem, Abschnitt 5.3). Identifiziert wurden die Merkmale: Mitbestimmung bei der Wahl der Fragestellung (OF1), bei der Wahl der Materialien (OF2), bei der Durchführung (OF3) sowie bei der Nachbereitung (OF4).

Die Ergebnisse zeigen, dass Offenheit bei den untersuchten Experimentiereinheiten praktisch keine Bedeutung hat (vgl. Tabelle 7.5): In nur 4 der 68 Experimentiereinheiten zeigen sich Ansätze von Offenheit. Während die Merkmale OF1 und OF4 gar nie auftreten, registriert man 3 Vorkommnisse bei OF2 und ein Vorkommnis bei OF3.

	Merkmale der Kategorie <i>Offenheit</i>	n
OF1	Schülerinnen und Schüler können die Fragestellung eines Experiments mitbestimmen oder können aus mehreren Fragestellungen wählen.	0
OF2	Aus einer grösseren Auswahl von Materialien wählen Schülerinnen und Schüler jene aus, die ihnen für die Beantwortung einer Fragestellung oder Problemstellung relevant erscheinen.	3
OF3	Schülerinnen und Schüler werden aufgefordert, die Durchführung des Experiments in einzelnen Aspekten im Hinblick auf die Beantwortung der Problemstellung selbst zu gestalten.	1
OF4	Schülerinnen und Schüler haben bei der Nachbereitung des Experiments Wahlmöglichkeiten.	0

Tabelle 7.5: Anzahl der Experimentiereinheiten n, die den Merkmalen OF1 bis OF4 zugeordnet sind.

7.3 Qualitative Analyse ausgewählter Experimentiereinheiten

Im Folgenden werden auf der Basis von Lehrertranskripts einzelne Experimentiereinheiten qualitativ analysiert. Die Grundlage dafür bilden Experimentiereinheiten, die anhand des Ratingsystems besonders ausgeprägt bewertet wurden. Das genaue Vorgehen bei der Auswahl und der Analyse der Einheiten wurde in Abschnitt 5.4 beschrieben. Im Gegensatz zur Beurteilung der Einheiten auf Basis von Bild und Ton, wie das bei den vorhergehenden Analysen der Fall war, ist die Beurteilungsgrundlage bei Transkripten auf die Sprache reduziert. Die Untersuchung wird somit eine neue Sicht auf die Einheiten ermöglichen und ergibt aus mehreren Gründen einen Mehrwert: Beispiele gelingenden experimentellen Handelns werden illustriert. So können Aspekte der Qualität direkt mit Äusserungen der Lehrenden verbunden werden. Gleichzeitig eröffnet die Fallanalyse die Möglichkeit, die Beurteilung anhand des Ratingsystems an einzelnen Beispielen zu bestätigen bzw. kritisch zu hinterfragen. Ausserdem können aus den Fällen neue Kategorien der Qualität experimentellen Handelns hervorgehen.

Anknüpfend an die Ausführungen werden die Ziele der Analyse durch folgende Fragen vorgegeben (vgl. Abschnitt 2.3):

- T5 Was zeichnet Experimentiereinheiten hoher Qualität aus? Welche Eigenschaften der Lehrperson sind damit verbunden?
- T6 Inwiefern decken sich die Befunde der Fallanalysen mit den Ergebnissen aus dem Ratingverfahren?
- T7 Welche neuen oder verfeinerten Kategorien ergeben sich aus den Fallanalysen?

7.3.1 Experimentiereinheit 1, S-1-05-01

Vierstufiger Schalter

Der Lehrer A präsentiert der Klasse folgende Problemstellung: Für eine Leuchte im Wohnzimmer mit vier Glühlampen muss ein Stufenschalter entwickelt werden. Wahlweise sollen damit eine, zwei, drei oder alle vier Glühlampen gleichzeitig betrieben werden können (vgl. Tabelle 5.4). Er knüpft mit der Aufgabe an die vorhergehende Problemstellung an, bei der bereits ein einfacher Schalter von den Lernenden gebaut wurde. Schülerinnen und Schüler haben den Auftrag, den Stufenschalter in Gruppen zu entwickeln und danach einen Prototyp zu erstellen. Den Gruppen gehören jeweils zwei bis drei Lernende an. Lehrer A geht von Gruppe zu Gruppe, lässt sich dabei über den Arbeitsstand informieren und bietet bei Problemen seine Unterstützung an. Für die Entwicklung und Umsetzung des Schalters stehen folgende Materialien zur Verfügung: Papier für Skizzen, Draht, Nägel, Schrauben, Blechstreifen, Kork und Holzbretter. Das passende Werkzeug liegt ebenfalls bereit. Im Rahmen der Schlussbesprechung teilt jede Gruppe ihren Arbeitsstand mit. Hierbei werden auch Verbesserungsmöglichkeiten besprochen.

Transkriptanalyse

Die Analyse des Transkripts des Lehrers A wurde nach den in Abschnitt 5.4 beschriebenen Schritten durchgeführt.

Diese Fragen. Die Kategorie *Diese Fragen* bezieht sich auf Äusserungen des Lehrers A während der Vorbereitungsphase des Experiments. Darin streicht er drei Grundfragen, er nennt sie “diese Fragen”, bei der Lösung eines technischen Problems heraus: 1) *Um was geht es?* 2) *Was wird gebaut?* 3) *Warum ist man sicher, dass es geht?* Die Grundfragen bilden die Subkategorien zur Kategorie *Diese Fragen*. Sie sind jedoch nicht nur während der Vorbereitung präsent. Vielmehr beruft A sich wiederholt während der Durchführung und der Nachbereitung auf sie. Zur Verdeutlichung der Subkategorien werden jeweils exemplarisch Exzerpte angegeben.

Der Subkategorie *Um was geht es?* gehören Äusserungen von A an, die auf das Bewusstmachen und die Verdeutlichung der gesetzten Fragestellung hinzielen. Ex-

zerpte zur Subkategorie treten einerseits während der Vorbereitungsphase auf, innerhalb der die Fragestellung vorgestellt wird. Das geschieht im Rahmen eines fragend-entwickelnden Unterrichtsgesprächs. Andererseits ruft A die Fragestellung immer wieder im Kontakt mit den Gruppen in Erinnerung.

Also, noch einmal bitte, Michael² >um Aufmerksamkeit bittend< es geht um die Frage "wie viele": wie viele Stufen oder wie viele Positionen der Schalter braucht. (00:19:17)

Aber jetzt gehts ja darum, dass du es erklären kannst und dass du eine Lösung findest oder: also da kannst du schon so zeichnen. Und du kannst es auch so bauen. (00:46:19) [7-1]

Die Subkategorie *Warum ist man sicher, dass es geht?* verweist auf Exzerpte des Transkripts, in welcher A die Schülerinnen und Schüler auffordert, die skizzierte Schaltung zu durchdenken. Dabei sollen sich die Lernenden vorstellen, wo im Schalt-schema der Strom durchfließt bzw. wo der Strom durchfließen sollte. Weiter gehören Exzerpte dazu, bei denen A laut denkt und dabei den Schülerinnen und Schüler modellhaft seine Überlegungen schildert. So wird für die Lernenden sichtbar, dass experimentelles Handeln ganz wesentlich mit kognitiven Prozessen verbunden ist.

Ja, ja, du musst dir mal ganz genau überlegen, ganz genau: wo fließt der Strom durch, wo soll der Strom durchfließen. (00:43:08)

Aber, wart schnell >unterbricht den hämmernden S< der, der Strom geht ja nachher hier zurück, wenn du hier auf diese Lampe ansteuerst mit einem Kabel, dann geht, sind ja die verbunden, dann geht der Strom hier zurück >fragend<. (00:59:03) [7-2]

Feedback. Die Kategorie bezieht sich auf verschiedene Äusserungen von A, die zum Feedback geben gehören. Darin eingeschlossen ist die Bewertung und Reflexion der in Gruppen durchgeführten Arbeiten und das Aufzeigen von Entwicklungsmöglichkeiten. Demzufolge unterteilt sich die Kategorie *Feedback* weiter in die Subkategorien *Bewerten* und *Entwickeln*. Der Subkategorie *Bewerten* werden Exzerpte zugeordnet, bei der A Lösungen der Schülerinnen und Schüler bewertet. Meist paraphrasiert A in einem ersten Schritt den Vorschlag des Lernenden:

Kann- Kannst du dann ausschalten, wenn mit einem Draht auf alle vier... Du sagst, du gehst mit diesem Draht auf den Kreis hier. Die Füße bilden wie ein

²Auftretende Namen wurden geändert

-, oder ein Viereck. Und wenn du hier drauf gehst, ist ja immer bei allen vieren Strom. Kannst du dann nur eine, eine Lampe brennen lassen? (00:49:22) [7-3]

Diese und ähnliche Äusserungen unterstreichen die Ernsthaftigkeit, mit der sich A jeder einzelnen Lösung der Gruppen annimmt. Meist entspricht die von den Schülerinnen und Schülern vorgeschlagene Lösung noch nicht der eigentlichen Lösung des Problems. Hierbei kommt der Aspekt des Bewertens hinzu:

Also hier haben wir zwei oder, wenn wir die einzeln zählen, fünf Schalter. Eins, zwei, drei, vier, fünf und es wäre schon gut, wenn man alles in einem Schalter hätte, wäre schon gut. (00:48:02) [7-4]

Es ist auffallend wie vorsichtig A die Lösungen der Gruppen bewertet. Die vorgestellte Lösung, eine Leuchte mit fünf Schaltern zu betreiben, entspricht in keiner Weise den gestellten Anforderungen. Trotzdem honoriert A die Arbeit und spricht davon, dass "es gut wäre" die fünf Schalter in einem integriert zu haben. Der Subkategorie *Entwickeln* sind Exzerpte zugeordnet, in welchen A Entwicklungsmöglichkeiten von Lösungen anspricht oder von den Schülerinnen und Schülern aufnimmt. Wie bei *Bewerten*, zeigt sich auch hier die Vorsicht von A, wenn es darum geht, eine Lösung zu korrigieren oder zu verbessern. Oft verwendet A dafür die Möglichkeitsform. A im Gespräch mit einer Gruppe:

Gibt es eine Möglichkeit, dass man das ändern könnte? Man muss ja irgendwie von hier die Lampen einzeln ansteuern können. Also, statt dass du diese Verbindung hier machst, könnte man doch diese Verbindung auch hier machen, so, dass man diese ansteuern kann, dann diese, dann diese und dann alle vier. Das wäre eine Möglichkeit, oder? (01:05:00) [7-5]

Im Rahmen der Besprechung des Arbeitsstandes der Gruppen gegen Ende der Experimentiereinheit kommt der Aspekt der Entwicklung einer möglichen Lösung öfters vor. Hierbei werden auch Ideen, die nicht weitergeführt wurden, angesprochen. A in der Rolle des Moderators:

Ja, man müsste schauen, wie man, wie kann man das mit der einzelnen Lampe machen, ja. Und jetzt bist du weggekommen von dem und nimmst sie auseinander und probierst jetzt von hier jede einzelne anzusteuern. Okay? (01:12:12) [7-6]

Gruppenmoderation. Zu der Kategorie gehören Aussagen von A, welche die Zusammenarbeit in Gruppen betreffen. Sie unterteilen sich in solche, die A vor Beginn

der Gruppenarbeitsphase gewissermassen als Anleitung zur Zusammenarbeit gibt (*Anleiten*) und jenen, die A im direkten Kontakt mit den Gruppen macht (*Integrieren, Austausch fördern*). Zur Kategorie *Anleiten* gehört wesentlich ein Exzerpt aus der Vorbereitungsphase des Experiments.

Es gibt ein, äh, Hinweis, der mir wichtig ist. Wir haben, Michi >um Aufmerksamkeit bittend< wir haben letztes Mal gesehen, dass es schwierig ist in der Gruppe, sich zu einigen, was man bauen will und dann hat es so Gespräche gegeben: nein wir machens so, nein auf diese Art, mach das selber und so weiter. Also diese Momente, wo man sich nicht ganz einig ist. Es gibt dazu eine Möglichkeit, die ich euch, die ich euch empfehle, nämlich: Nehmt ein Blatt Papier, ein Blatt Papier und probiert es aufzuzeichnen. Das ist eine gute Möglichkeit, um miteinander zu diskutieren, welche Lösung man baut. Ich habe hier Notiz- ein Notizblock. Man kann Papier nehmen und darauf zeichnen. Und beim Zeichnen soll man sich überlegen: wie muss der Strom fliessen, also wo soll er durchgehen, ja. (00:35:09)

[7-7]

Danach beginnen die Schülerinnen und Schüler mit der Arbeit in den Gruppen. Im Gespräch mit diesen versucht A wiederholt, Schülerinnen und Schüler, die sich eher passiv verhalten, in die Auseinandersetzung mit einzubeziehen.

Sandra, hast du auch eine Idee oder hast du einfach so mit-? (00:48:24)

Und jetzt, Monika, schau mal, was er erklärt. (00:51:13)

[7-8]

Solche und ähnliche Äusserungen wurden der Subkategorie *Integrieren* zugeordnet. Bei der Besprechung des Arbeitsstandes gegen Ende der Experimentiereinheit tauschen sich die Gruppen gegenseitig aus (*Austausch fördern*). Hierbei nimmt A die Rolle des Moderators ein.

Äh, Kevin, kannst du schnell sagen, was ihr bis jetzt gemacht habt, an was ihr gebaut habt, was ihr gedacht habt, welche Lösung ihr prob- probiert habt? (01:10:07)

[7-9]

Im Gespräch zwischen A und den Schülergruppen wird schnell klar, dass die Aufforderung, über Erfahrungen zu berichten, durchaus so umfassend verstanden werden soll. A ist es offensichtlich wichtig, dass auch Versuche, die nicht zu der eigentlichen Lösung geführt haben, mitgeteilt werden. Trotzdem ufern hierbei die Rückmeldungen, dank der Führung des Gesprächs durch A, nicht aus.

Mitteilen. Die Kategorie beinhaltet wichtige Aspekte der Kommunikation zwischen A und einer Schülerin oder einem Schüler. Oft initiiert A das Gespräch durch die Aufforderung, man möge ihm die Lösung bzw. eine Lösungsidee erklären. Die dazu passenden Exzerpte sind in der Subkategorie *Erklären* zusammen gefasst. Nachdem die Erklärung oder ein Teil davon erfolgt ist, äussert A seine Einsicht bzw. seinen Widerspruch (*Einsehen*) oder paraphrasiert die Erklärung (*Paraphrasieren*). Bei der Subkategorie *Erklären* finden sich solche und ähnliche Äusserungen:

Kannst du mir das erklären? Ich hab vor- ich hab mich vorhin gefragt, was du da zeichnest. (00:58:40) [7–10]

Die Aufforderung, etwas zu erklären, wird hier mit einer Bekundung des Interesses am Beitrag der Schülerin verbunden. A integriert damit gleichzeitig die Schülerin in die Diskussion mit der Gruppe. Somit gehört das Exzerpt auch der Subkategorie *Integrieren* der Kategorie *Gruppenmoderation* an.

In einem anderen Beispiel bestätigt A seine Einsicht wie folgt:

Das sehe ich ein, so wie du sagst, ja. Wäre eine Möglichkeit. (00:45:55) [7–11]

Öfters paraphrasiert A auch Erklärungen der Schülerinnen und Schüler. Wahrscheinlich geht es hierbei darum, Lösungswege nachzuvollziehen und offene Details zu klären.

Du hast hier einen Schalter, wo du sagst es gibt Strom auf dieses Blättchen und wenn du eine Lampe willst, machst du hier eine Verbindung, wenn du zwei m-willst, machst du hier und hier, hier, hier, hier gibt drei und alle zusammen, alle zusammen gibt vier. (00:47:21) [7–12]

Falldiskussion

Zusammenfassend weist das Verhalten von A drei besondere Merkmale auf: Eine ausgeprägte kommunikative Kompetenz, eine besondere Gewichtung des prozesshaften experimentellen Handelns und, etwas weniger deutlich als letztere beiden, ein theoriegeleitetes Vorgehen.

Die besonderen Merkmale im Verhalten von A lassen sich aus der Transkriptanalyse und der Zusammenfassung der Kategorien in Abbildung 7.6 rekonstruieren. Das kommunikative Verhalten von A ist geprägt von einem Interesse an den Lösungsvorschlägen und Ideen der Schülerinnen und Schüler. Das Interesse zeigt sich über

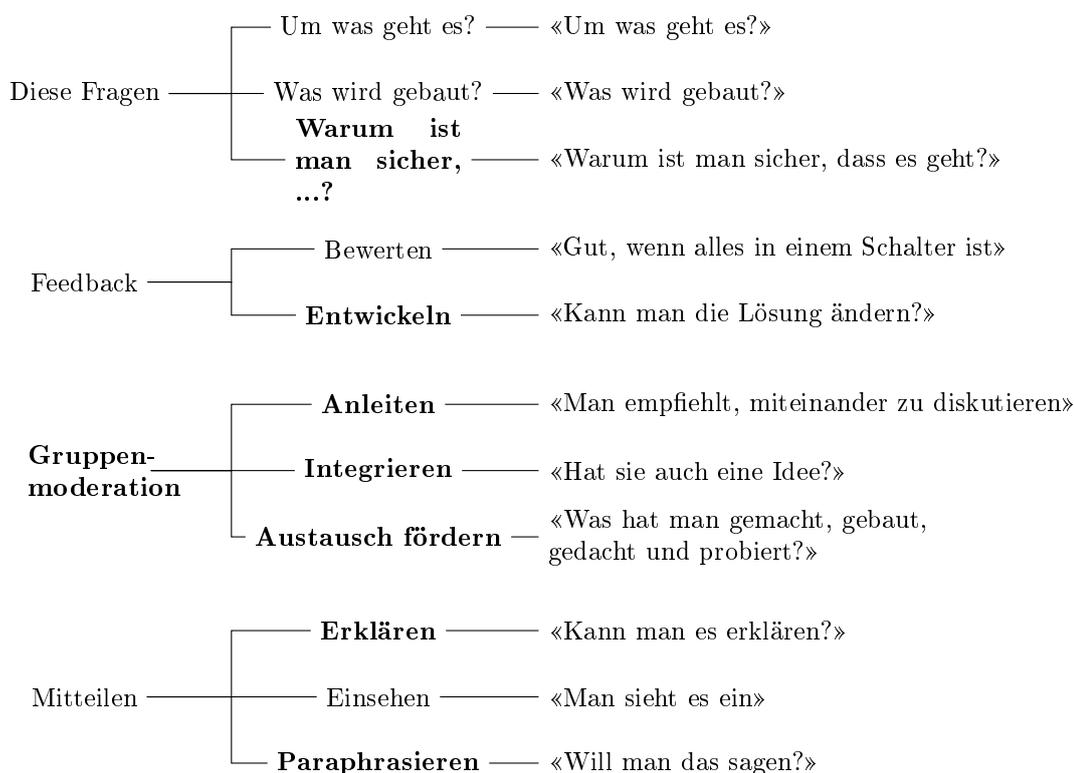


Abbildung 7.6: Strukturdiagramm der Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, S-1-05-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.

alle Kategorien hinweg. Besonders deutlich wird es bei den Kategorien *Feedback* und *Mitteilen*. Redewendungen wie “du sagst, du gehst [...]”, “hab mich vorhin gefragt was du [...]” oder “das sehe ich ein, so wie du sagst [...]” sind Beispiele von Indikatoren für dieses Interesse. A zeigt in seinem Verhalten ein Bewusstsein für das Potential, aber auch für die Schwierigkeiten von Gruppenarbeiten. Das wird dann deutlich, wenn A die Gruppenarbeit anleitet und dabei eine “Möglichkeit” präsentiert, wie man sich in der Gruppe einigen kann oder, wenn er Schülerinnen und Schüler, die sich aus der Gruppe ausklinken, ins Gespräch zu integrieren versucht ([7–7] und [7–8]). Die Besprechung des Arbeitsstandes legt A zudem auch so an, dass die Gruppen gegenseitig voneinander lernen können ([7–9]).

Neben der ausgeprägt kommunikativen Kompetenz von A, die aus der Transkriptanalyse sehr deutlich hervortritt, lässt sich ein weiteres Merkmal der Lehrperson erkennen: die besondere Gewichtung des Prozesshaften beim experimentellen Handeln. Ersichtlich wird das zum einen aus der Kategorie *Diese Fragen*, welche

fast prototypisch die wiederkehrenden Fragen eines Problemlöseprozesses beinhaltet. Zum anderen sind für A bei der Besprechung des Arbeitsstandes insbesondere auch Versuche, die nicht zum gewünschten Ziel geführt haben, von grosser Bedeutung. Formulierungen wie “[...] was ihr bis jetzt gemacht habt, an was ihr gebaut habt, was ihr gedacht habt [...]” oder “[...] und jetzt bist du weggekommen von dem [...]” unterstreichen dies. Mit der Gewichtung des Prozesshaften einher gehen die Gespräche darüber, wohin der Strom fliesst bzw. fließen soll ([7-2]). Hier wird ein theoriegeleitetes Vorgehen erkennbar. Eingeleitet wird das Vorgehen durch die Frage: “Warum ist man sicher, dass es geht?” Es ist anzunehmen, dass die Schülerinnen und Schüler ein Modell des elektrischen Stromes kennengelernt haben. Allerdings wird es während der Experimentiereinheit von A nicht mehr aufgegriffen.

Vergleich mit den Ratings

Die qualitative Analyse des Transkripts ist in folgenden Punkten mit den Ratings besonders stimmig (vgl. Tabelle 7.6): Die Fähigkeit von A, auf die Lernenden eingehen zu können, findet sich in den Ratings der *Prozessreflexion* ($1=trifft\ zu$) und der *Ergebnisreflexion* ($1=trifft\ zu$). Zum Ergebnis ersterer Facette trägt zudem die auffallende Betonung des Prozesshaften beim experimentellen Handeln bei. Wie auch aus der Transkriptanalyse hervorgeht, ist die Perspektive experimentellen Handelns als theoriegeleitete Aktivität weniger deutlich erkennbar. Zu den Ratings der Facetten *Fachimmanente* und *Lebensweltliche Einbettung* ($2=trifft\ grösstenteils\ zu$) fehlt hingegen eine Entsprechung in der qualitativen Analyse. Möglicherweise ist eine Einbettung bei der zeilenweisen Analyse des Transkripts schwierig zu erkennen.

Sichtstruktur									
Vorb.		Durchf.		Nachb.		Σ			
22		32 _{Gr}		22		76			
Tiefenstruktur									
FU	OF	EB1	EB2	ZK	PR	ER	KA	TF	BB
2	0110	2	2	1	1	1	1	3	2

Tabelle 7.6: Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall S-1-05-01: Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Die Durchführung findet in Gruppen statt (Abkürzungen der Facetten, vgl. Abbildung 7.1).

7.3.2 Experimentiereinheit 1, F-0-16-01

Die vorliegende Experimentiereinheit stammt aus der finnischen Teilstichprobe. Die entsprechende Sequenz wurde von einer finnisch- und deutschsprachigen Hilfsassistentin transkribiert und anschliessend ins Deutsche übersetzt.

Reihenschaltung von Batterien

Der Lehrer B führt in das Experiment ein, wobei er an dem bereits behandelten Arbeits- und Leistungsbegriff der Mechanik anknüpft und diesen mit der gegenwärtigen Fragestellung, was Leistung und Arbeit für einen elektrischen Stromkreis bedeuten, verbindet. Die Schülerinnen und Schüler sollen folgende zwei Stromkreise miteinander vergleichen, indem sie Stromstärken und Spannungen messen: einen einfachen Stromkreis, bestehend aus einer Glühlampe und einer Batterie, einen Stromkreis, bestehend aus einer Glühlampe und zwei in Reihe geschalteten Batterien. Dazu fordert B die Lernenden auf, sich zu den nachfolgenden Fragen Gedanken zu machen: 1) Bei welcher der beiden Schaltungen leuchtet die Glühlampe leistungsfähiger? 2) Bei welcher Schaltung wird mehr Energie "verbraucht"? Die beiden Fragen sowie die Schaltschemen der Stromkreise werden an der Wandtafel festgehalten.

Die nachfolgende Schülerexperimentierphase ist als Partnerarbeit organisiert (vgl. Tabelle 7.7). Während der Phase geht B von Tisch zu Tisch und hilft bei Problemen. Schülerinnen und Schüler, die mit der experimentellen Umsetzung schneller fertig sind, dürfen die Glühlampe durch einen Motor ersetzen und die analogen Fragen für diesen Fall untersuchen. Nach dem Aufräumen der Materialien werden die Messwerte im Plenum vorgetragen und verglichen. Dazu beantworten Schülerinnen und Schüler die beiden Fragen intuitiv. Daran anknüpfend präsentiert B eine Computeranimation. Hierbei wird ein Wasserstromkreis einem elektrischen Stromkreis gegenübergestellt. B diskutiert die Parallelen der beiden Kreisläufe und illustriert Veränderungen der Spannung bzw. der Druckdifferenz und der elektrischen Stromstärke bzw. des Massenstroms. Die Erkenntnisse aus dem Schülerexperiment und aus der Animation werden durch B zusammengefasst. Darauf aufbauend formuliert er den Zusammenhang zwischen elektrischer Leistung, Spannung und Stromstärke.

Transkriptanalyse

Die Kategorien, welche sich aus dem vorherigen Fall ergaben, werden in die Beurteilung des neuen Falls aufgenommen (vgl. Abbildung 7.6). Aufgrund der durchgeführten Revision, ergab sich eine Anpassung der Struktur des Kategoriensystems: Zur Verbesserung der Übersichtlichkeit werden alle Kategorien einer der folgenden drei Ebenen zugeordnet: I) Ideenebene II) Objektebene III) Sozial-kommunikative Ebene. Weiter wurden die bestehenden Subkategorien aus *Diese Fragen* in ihrer Bedeutung erweitert. Deshalb tragen sie auch neue Bezeichnungen. Diese sind: Subkategorie *Um was geht es?* → *Problemstellen*, *Wie setzt man es um?* → *Gestalten* und *Warum ist man sicher, dass es geht?* → *Interpretieren*. Im Folgenden werden ausschliesslich die für den vorliegenden Fall bedeutsamen Kategorien besprochen und mit Exzerpten illustriert.

Ideenebene. Diese Fragen. Die Kategorie *Diese Fragen* setzt sich neu aus den Subkategorien *Problemstellen*, *Gestalten* und *Interpretieren* zusammen. Zur Subkategorie *Problemstellen* werden sämtliche Exzerpte zugeordnet, welche der Klärung der Fragestellung bzw. Problemstellung dienen. Hierbei fällt auf, dass B sich sehr wage ausdrückt: Mit folgenden Exzerpten wird die Experimentiereinheit eingeleitet.

Unser Programm heute sieht so aus: Wir sprechen über elektrische Leistung und auch ein bisschen über Arbeit und Energie. (00:01:16)

Wir machen auch praktische Aufgaben und Messungen, ihr dürft am Anfang der Stunde selbst messen und dann üben wir auch ein bisschen, wie man elektrische Leistung rechnen kann. (00:01:25) [7–13]

Offenbar rückt B die Information über den Ablauf der Stunde bzw. der Experimentiereinheit gegenüber einer Problemstellung in den Vordergrund. Die eigentliche Problemstellung wird dann im Verlauf der Vorbereitung klarer formuliert.

In welcher (Schaltung A oder B) leuchtet die Lampe -eurer Meinung nach- effizienter bzw. funktioniert effizienter? Also...in welcher Schaltung funktioniert die Lampe effizienter? (00:10:02) [7–14]

Obige Fragestellung wird von B nicht weiter begründet oder in eine übergeordnete Fragestellung eingebettet.

Die Subkategorie *Gestalten* umfasst Äusserungen von B, die einen experimentellen Aufbau begründen bzw. hinterfragen sollen. Während die erwähnte Kategorie bei S-1-05-01 sehr bedeutsam ist (Was wird gebaut?), spielt sie im vorliegenden Fall keine Rolle. Der experimentelle Aufbau wird weder begründet noch hinterfragt.

Machen wir das mal mit einer Batterie von 1 Komma 5 Volt, so dass die Glühlampe leuchtet. Das alte bekannte Schaltbild. (00:05:56)

Aber jetzt sollten wir uns das von einer anderen Sichtweise überlegen und gucken, wie das funktioniert. (00:05:49) [7–15]

Offenbar haben die Schülerinnen und Schüler den einfachen Stromkreis mit einer Glühlampe und einer Batterie bereits bei einem anderen Experiment kennengelernt. Mit der “anderen Sichtweise” verweist B auf den Umstand, dass Energien und Leistungen des elektrischen Stromkreises thematisiert werden sollen.

Zur Subkategorie *Interpretieren* zählen Äusserungen der Lehrperson, welche entweder Interpretationen beinhalten oder zu solchen auffordern. Dabei lassen sich Interpretationen nicht klar von Beobachtungen oder von wiedergegebenem Wissen abgrenzen. Die Spannweite der Äusserungen reicht deshalb auch von Exzerpten, welche sich direkt auf das Experiment beziehen bis zu solchen, welche eher in einem losen Zusammenhang dazu stehen. Die folgenden Exzerpte verdeutlichen dies:

Das hier wirkt ein bisschen niedrig, sodass vielleicht die Batterie leer ist oder irgendwas... Naja, eine leere Batterie hin oder her, aber diese Ergebnisse gehen trotzdem in die richtige Richtung, sodass man hier merkt, dass die Spannung steigt, so wie sie in einer Reihenschaltung auch steigen soll, und dann verursacht diese Steigung der Spannung natürlich, dass der Strom auch wächst. (00:24:13)

Genau, also gibt es eine Anziehungskraft zwischen zwei verschiedene Ladungen bzw. zwischen Ladungen mit ungleichem Vorzeichen. Diese positiven Ladungen hier ziehen die negativen an und wir sagen dann, dass die elektrische Anziehungskraft durch den Leiter und das elektrische Feld übermittelt wird. So, dass die Elektronen durch das elektrische Feld diese Kraft erleben und fangen an, sich zu bewegen. Und wenn die durch die Lampe fliessen, entsteht Reibung und durch diese Reibung entsteht Wärme und die Lampe wird heiss und fängt an zu glühen. (00:32:50) [7–16]

Letztere Äusserung erfolgt auf die Antwort eines Schülers zur Frage, was die Elektronen in Bewegung setzt. Aus zwei Gründen ist sie typisch für B: Sie bringt die grosse Genauigkeit und Sorgfalt zum Ausdruck, die B bei seinen Ausführungen zeigt.

Sie ist zudem typisch, weil B auf die stichwortartige Schülerantwort (“Ladungsunterschied”) nicht mit einer Nachfrage reagiert, sondern das Stichwort gleich selbst ausführt.

Ideenebene. Lebensweltlicher Kontext. Die Kategorie *Lebensweltlicher Kontext*, welche ebenfalls der Ideenebene zugeordnet ist, beinhaltet Exzerpte, die Bezüge des Experiments mit der Lebenswelt der Schülerinnen und Schüler herstellen. Bei B spielt die Kategorie eine untergeordnete Rolle. Kontextbezüge stellt B vorwiegend innerhalb der Fachinhalte her (siehe nachfolgender Paragraph zum fachimmanenten Kontext). Eine Ausnahme bildet die Reflexion des Begriffs “Energieverbrauch”, die der Subkategorie *Sprachgebrauch* zugeordnet wird.

Was ist damit, ich habe ja das Wort verbraucht in Anführungszeichen gesetzt, also wie könnten wir den Energieverbrauch hier in der Lampe verstehen? Verschwindet die Energie komplett? Die Lampe bekommt ja durch den Strom chemische Energie von der Batterie. Aber verschwindet die Energie komplett? (00:25:28) [7–17]

Die Äusserung unterstreicht, dass B alltagssprachliche Formulierungen reflektiert und dabei das Bewusstsein für die Unterschiede zwischen Fach- und Alltagssprache bei den Lernenden fördert. Wie beim Exzerpt zu *Interpretieren* neigt der Lehrende auch hier dazu, die gestellte Frage gleich selbst zu beantworten.

Ideenebene. Fachimmanenter Kontext. Zur Kategorie gehören Äusserungen von B, welche Bezüge zwischen verschiedenen Themen der Physik beinhalten. B verbindet das Experiment ausgeprägt mit anderen, bereits behandelten Themen. Die entsprechenden Exzerpte sind der Subkategorie (*Erinnern*) zugeordnet. Dies deshalb, weil den Bezügen oft die Formel, “Erinnert ihr euch?”, vorangeht.

Jetzt, wo wir über elektrische Leistung und Energie sprechen, könnten wir uns ein wenig daran erinnern, woran man sich so noch erinnern kann, ich meine wir hatten ja im Herbst den Mechanik-Kurs und die Begriffe Arbeit und Leistung und Energie. (00:01:54)

Und was ist denn jetzt mit der Leistung? Wie könnte man die elektrische Leistung genauer definieren? In der Mechanik hat sie ja die Geschwindigkeit der Arbeit beschrieben. Hier kann man eigentlich auch so denken... (00:39:26) [7–18]

Beim ersten Exzerpt fällt wiederum die Unverbindlichkeit auf, die bei B im Zusammenhang mit Aufforderungen zu beobachten ist: “[...] könnten wir uns ein wenig

daran erinnern, [...]”. Auf die Äusserung folgt eine ausführliche Rekapitulation der Begriffe Arbeit, Energie und Leistung in der Mechanik durch die Lehrperson. Besonders Gewicht legt B darauf, die beiden Themenbereiche Mechanik und Elektrizitätslehre explizit miteinander in Verbindung zu bringen. Das lässt sich aus dem zweiten Exzerpt gut erkennen. Mit “Hier” wird dabei auf den elektrischen Stromkreis verwiesen. Im Zusammenhang mit dem *Erinnern* kommen wiederholt Fragekonstrukte vor, die nach folgendem Schema ablaufen: Zunächst fragt B etwas eher Allgemeines. Melden sich nicht unmittelbar darauf Schülerinnen und Schüler, stellt B die Frage nochmals, aber diesmal in einem Anwendungsbeispiel. Folgendes Exzerpt veranschaulicht das Verhalten von B:

Könnt ihr euch daran erinnern, wie eine Reihenschaltung funktioniert? >wartet kurz< Das ist ja auch länger her, dass wirs gemacht haben. Wie schaltet man die Batterien da? (00:08:33) [7–19]

Objektebene. Messen. Etliche Transkriptstellen stehen in einem Zusammenhang mit dem Messen. Die diesbezüglichen Äusserungen werden entsprechend ihres Inhaltes entweder der Subkategorie *Erklären* oder der Subkategorie *Vergleichen* zugeordnet. Letzterer gehören sämtliche Exzerpte an, bei welchen B zum Vergleich der Messwerte aufruft oder selbst Messwerte vergleicht. Äusserungen, welche die Messmethodik zum Inhalt haben, werden der Subkategorie *Erklären* zugeordnet.

Fangen wir damit an, dass wir als Erstes was messen und dann basiert darauf fangen wir an, über die Leistung elektrischer Geräte nachzudenken. (00:04:07) [7–20]

Die Äusserung charakterisiert treffend den Ablauf des durchgeführten Experiments. Sie zeigt auch das Selbstverständnis, mit der Erkenntnisgewinn beim experimentellen Handeln mit einem induktivistischen Vorgehen verbunden wird. Erklärungsversuche für ein Phänomen beginnen also erst nach bereits durchgeführten Messungen: messen und *nachher* erklären.

Währenddem die Schülerinnen und Schüler das Experiment durchführen, geht B zu den Gruppen und hilft beim Anschliessen der Messgeräte. Einerseits erteilt B dabei Handlungsanweisungen, z.B. zum Aufbau der Schaltung (siehe nachfolgenden Paragraph *Handlungen* und Abbildung 7.7), andererseits weist B wiederholt darauf hin, was aufgeschrieben werden soll. Hierbei werden die Schwierigkeiten der Schüle-

rinnen und Schüler deutlich, welche sie im Umgang mit den Messgeräten haben. So verwechseln mehrere Gruppen die Spannungs- mit der Stromstärkemessung.

Und jetzt schreibt ihr diese Ergebnisse auf! Das hier ist die Spannung und dieses da der Strom. Da, auf dem ersten. Das hier ist also die Spannung. Also das untere. (00:18:26) [7-21]

Nach der Durchführung des Experiments werden im Rahmen einer längeren Plenumsphase die Messergebnisse verglichen. Die folgenden Exzerpte stammen aus der Subkategorie *Vergleichen*:

So, gucken wir dann, wie es mit den Ergebnissen aussieht. Ich habe schon ein wenig rum geschaut, wenigstens die Gruppe von Mark hatte in etwa richtige Ergebnisse, also ... (00:22:19)

Ihr habt ja in den Gruppen ein wenig unterschiedliche Ergebnisse. Aber verändert sie nicht, radiert eure Eigenen nicht weg! (00:22:35) [7-22]

Bei obigen Äusserungen kommt die Ambivalenz zwischen dem Wissen von B über die "richtigen Ergebnisse" und dem Geltenlassen unterschiedlicher Messwerte zum Ausdruck.

Objektebene. Handlungen. Während der Durchführung des Experiments ist B intensiv damit beschäftigt, den Aufbau der Schaltung anzuleiten. Dies scheint trotz des Auffrischens der Reihen- und Parallelschaltung sowie des sorgfältigen Besprechens des Schaltschemas im Vorfeld der Durchführung notwendig zu sein. Im Vordergrund steht hierbei, «wie es gemacht wird», und nicht, «wieso es gemacht wird».

Nee, ein anderes Kabel hierhin, dazwischen. Also auf dem Bild könnt ihr immer gucken wo das Kabel entlang ziehen soll. (00:15:27)

Ja. Und jetzt macht ihr einfach eine zweite Batterie dazu. Das hier braucht ihr gar nicht anzufassen, ihr macht da eine andere Batterie dahinter, in Reihe. (00:18:41) [7-23]

Äusserungen, die eher der *Sozial-Kommunikativen Ebene* zugeordnet werden könnten, kommen bei B praktisch nicht vor.

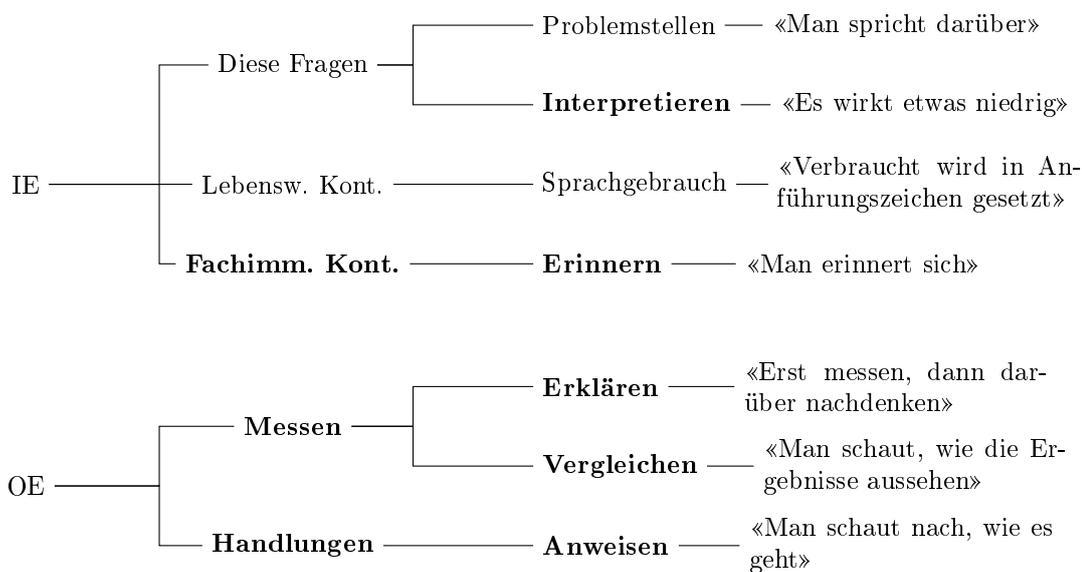


Abbildung 7.7: Strukturdiagramm der Ebenen (Ideenebene IE, Objektebene OE), Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, F-0-16-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.

Falldiskussion

Die Transkriptanalyse und die sich daraus ergebenden Kategorien bzw. Subkategorien (Abbildung 7.7) zeigen folgende Merkmale im Verhalten des Lehrers B: Er setzt das Experiment dazu ein, Fachinhalte auszuführen und Verbindungen zwischen verschiedenen verwandten Themen und Konzepten aufzuzeigen. Die bei B ausgeprägte vertikale Vernetzung zeigt sich ein erstes Mal während der Einführung. Hier stellt B Bezüge zwischen dem Arbeits-, Leistungs- und Energiebegriff in der Mechanik und der Elektrizitätslehre her (vgl. [7–18]). B nutzt im Verlauf des Experiments weitere Gelegenheiten, solche und ähnliche Bezüge herzustellen. Die Bedeutung der Fachinhalte für das Experiment wird dabei auch durch die Genauigkeit in den Formulierungen von B unterstrichen und zeigt sich beispielsweise bei [7–16].

Das Verständnis für die physikalischen Vorgänge und die in dem Zusammenhang etablierten Formulierungen gehören für B zusammen. Dies wird in [7–17] offenbar, wenn B die Schülerinnen und Schüler darauf hinweist, dass der Begriff des Energieverbrauchs eigentlich nicht zu einer physikalischen Vorstellung passt. Entgegen letzterem Exzerpt sind es hier nicht Fachtermini, die eingeführt werden, sondern Begrifflichkeiten aus dem Alltag, die im Kontext physikalischer Vorstellungen reflektiert

werden. Die Verbindung der Computeranimation und die dabei herausgearbeitete Analogie zwischen dem elektrischen und dem Wasserstromkreis mit den Messergebnissen aus den Schülerexperimenten zeigen, dass das Experiment hauptsächlich von Ideen und Modellen und weniger vom Messen bestimmt ist.

Die bisher erwähnten Verhaltensweisen von B sind alle der Ideenebene zuzuordnen (vgl. Abbildung 7.7). Neben der ausgeprägten Ideenebene zeigt das Verhalten von B auch wichtige Aspekte, die der Objektebene zugehörig sind. Widersprüchlich zur festgestellten Fokussierung auf Modelle und Ideen von B ist, dass er bei der Organisation des Experiments ganz im Sinne einer induktiven Logik, die Messung vor die Überlegung stellt. Indizien dafür finden sich z.B. in der wagen Ankündigungen des Experiments [7–13] oder im Exzerpt [7–20]: “[...] wir als Erstes was messen und dann basiert darauf fangen wir an, über die Leistung elektrischer Geräte nachzudenken [...]”. Tatsächlich ist die Schülerexperimentierphase für die Lernenden eher wenig kognitiv anspruchsvoll. Dazu trägt auch das Verhalten von B bei, wenn er den Lernenden Teile der Durchführung abnimmt oder sie rezeptartig anleitet (vgl. [7–21] und [7–23]).

Im Gegensatz zum Verhalten von A, kommen bei B die Kategorien *Feedback*, *Gruppenmoderation* und *Mitteilen*³ praktisch nicht vor; d.h., B gibt den Schülerinnen und Schülern keine Anleitungen zum Arbeiten in Gruppen und bietet keine Gelegenheit für einen Gruppen übergreifenden Austausch. Weiter fordert B Lernende nicht zu längeren Erklärungen auf. Ungenaue, stichwortartige Antworten entgegnet B, indem er die Antwort des Lernenden gleich selbst ausführt. Einzig während der Schülerexperimentierphase paraphrasiert B in Ansätzen Äusserungen der Schülerinnen und Schüler.

Vergleich mit den Ratings

Eine Übereinstimmung der Transkriptanalyse mit den Ratings zeigt sich für folgende Verhaltensweisen von B (vgl. Tabelle 7.7): die stark ausgeprägte vertikale Vernetzung, die sich in der fachimmanenten Einbettung niederschlägt (EB1, *1=trifft zu*), das eher wenig zielklare Vorgehen von B, welches sich in der Unsicherheit bei der Ankündigung des Experiments widerspiegelt (ZK, *3=trifft teilweise zu*), das Zusammenbringen von Erkenntnissen aus der Animation und den Messungen als Reflexion der Erkenntnisse (ER, *1=trifft zu*), die Betonung der Ideen und Modelle verbunden

³Diese Kategorien werden der Sozial-Kommunikativen Ebene zugeordnet.

mit einem angemessenen, aber hohen Anspruchsniveau (TF, *1=trifft zu* und KA, *2=trifft grösstenteils zu*) sowie die Genauigkeit in den Formulierungen von B (BB, *1=trifft zu*). Zunächst weniger stimmig erscheint das Frageverhalten von B, welches nicht zu einer eher ausgeprägten *Prozess Reflexion* (*2=trifft grösstenteils zu*) passen mag. Das lässt sich so erklären, dass B zwar nicht zu längeren Schülerantworten anregt, dafür aber umso ausgeprägter typische Gedankengänge antizipiert. Dieser Aspekt gehört ebenfalls zur *Prozess Reflexion* (vgl. Abschnitt 5.3).

Sichtstruktur									
Vorb.		Durchf.		Nachb.		Σ			
12		7 _{Pa} /6 _{DI}		20		45			
Tiefenstruktur									
FU	OF	EB1	EB2	ZK	PR	ER	KA	TF	BB
6	0000	1	4	3	2	1	2	1	1

Tabelle 7.7: Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall F-0-16-01: Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Die Organisation der Durchführung teilt sich hälftig in eine Partnerarbeit (Pa) und eine Lehrer-Demonstration (DI) (Abkürzungen der Facetten, vgl. Abbildung 7.1).

7.3.3 Experimentiereinheit 1, D-1-29-01

Mechanische und elektrische Energieübertragung im Vergleich

Das Experiment steht am Anfang einer Doppelstunde nach dem Urlaub. Anhand eines Demonstrationsexperiments nimmt Lehrer C das Thema von vor den Ferien wieder auf. Darin hebt C ein an einer Schnur befestigtes Massenstück mittels einer Vorrichtung, bestehend aus Spindel, Transmission und Kurbel, einen Meter hoch. C fordert die Schülerinnen und Schüler auf, den Vorgang zu beschreiben. Im Rahmen eines fragend-entwickelnden Unterrichtsgesprächs kommt man dazu, dass C an der Kurbel Arbeit verrichtet, welche über die Transmission auf die Spindel übertragen wird. Diese rollt die Schnur ein, an dessen Ende sich das Massenstück befindet: Hubarbeit wird am Massenstück verrichtet. Dabei wiederholt C wichtige Begriffe zur mechanischen Energie.

In einer nächsten Phase des Experiments wird dasselbe Massenstück mit einer neuen Vorrichtung, bestehend aus Motor, Leiter und Dynamo, gehoben. C nimmt das fragend-entwickelnde Unterrichtsgespräch wieder auf und unterstreicht dabei die Ähnlichkeiten der beiden Vorrichtungen. Es resultiert die Erkenntnis, dass elektrische Stromkreise, analog einer Transmission in der Mechanik, Energie übertragen. Eine Schülerin erfährt im Rahmen einer Demonstration, dass die Bedienung des Dynamos bei verschiedenen leistungsfähigen Glühlampen ganz unterschiedlich schwer geht. Somit macht C den Energietransport erfahrbar. Zum Abschluss stellt C die Verbindung zwischen der Dynamo-Motor Vorrichtung und der Energieversorgung zu Hause her.

Transkriptanalyse

Bereits definierte Kategorien bzw. Subkategorien werden für die qualitative Analyse des aktuellen Falls übernommen und an neue Aspekte entsprechend angepasst (vgl. Abbildungen 7.6, 7.7 und 7.8). Die bei Fall F-0-16-01 eingeführte Zuordnung von Kategorien zur *Ideen-, Objekt- und Sozial-kommunikativen Ebene* bleibt unverändert für den neuen Fall bestehen. Bei den Kategorien, also eine Hierarchiestufe weiter unten, ergeben sich folgende Anpassungen: Die Kategorien *Mitteilen* aus F-0-16-01 und *Feedback* aus S-1-05-01 werden in der Kategorie *Schülerbeiträge verarbeiten* zu-

sammengeführt. Hinzu kommt neu die Kategorie *Vorzeigen* mit ihren Subkategorien *Selbst vorzeigen* und *Vorzeigen lassen*. Schliesslich ergeben sich weitere Neuerungen bei den Subkategorien: Den Kategorien *Lebensweltlicher* und *Fachimmanenter Kontext* wird die Subkategorie *Bezüge herstellen* untergeordnet. Die Kategorie *Fachimmanenter Kontext* umfasst die bestehende Subkategorie *Erinnern*. Zusätzlich wird letztere Kategorie um die Subkategorie *Analogien aufzeigen* erweitert. Die für den aktuellen Fall wichtigen Kategorien und Subkategorien werden nachstehend aufgeführt und diskutiert.

Ideenebene. Diese Fragen. Aufgrund der Anlage des aktuellen Experiments – es geht um das Beschreiben eines Vorganges – drängt sich eine Problem- bzw. eine Fragestellung weniger auf. Dementsprechend lässt sich zur Subkategorie *Problemstellen* kein Exzerpt zuordnen. Anstelle der Fragestellung tritt die Aufforderung, den Vorgang, also das Hochheben des Massenstücks, zu beschreiben (siehe Subkategorie *Vorzeigen*). Die der Subkategorie *Interpretieren* zugeordneten Exzerpte können als Interpretationen von C verstanden werden.

Ja, elektrische Energie wird jetzt hier durch den Strom übertragen. Also, die Aufgabe des Stromkreises ist es also, elektrische Energie zu übertragen, Energie zu übertragen. (00:05:47) [7–24]

Dagegen werden Schülerinnen und Schüler kaum dazu angeregt, eigene Interpretationen abzugeben. Der experimentelle Aufbau ist nicht Gegenstand von Diskussionen. Demzufolge finden sich zur Subkategorie *Gestalten* keine Exzerpte.

Somit ist die Kategorie *Diese Fragen* im vorliegenden Fall nur partiell vertreten.

Ideenebene. Lebensweltlicher Kontext. In der Kategorie finden sich Exzerpte, bei denen sich das Experiment oder einzelne Aspekte davon auf die Lebenswelt der Schülerinnen und Schüler beziehen. C bettet das Experiment wiederholt in einen lebensweltlichen Kontext ein. Darüber hinaus scheint das Experiment auf diese Bezüge hin konzipiert zu sein.

Na, aus eurer Lebenserfahrung, wo habt ihr so ein Gerät schon mal selber betrieben? Pierre? [...] Und das heisst? [...] Ja, das ist der Dynamo. So können wir das hier auch erstmal nennen, es gibt noch einen weiteren Ausdruck, Generator, aber für uns ist das erstmal der Dynamo. (00:07:03) [7–25]

So werden den Schülerinnen und Schülern bekannte Geräte Elementen des Experiments zugeordnet. Auch sprachlich findet ein Bezug zum lebensweltlichen Kontext statt. C zum Begriff "Stromzähler":

Ja, so nennt ihr das. Das ist aber falsch. Das Ding muss wie heißen? Was wird, was wird dort eigentlich registriert? [S: Ich glaub die Elektronen, die da durchfließen]. Nene, bleiben wir mal bei den Begriffen, die wir hier haben. Wenn wir sie hier beschäftigen würden, als Kurbler, dann möchte sie was bezahlt haben? [S: Ihre Arbeit] Richtig, und das ist ein Arbeitsmessgerät, ein Energiemessgerät. (00:12:42) [7-26]

Einerseits fordert C die Schülerinnen und Schüler dazu auf, den Begriff des Stromzählers physikalisch zu hinterfragen. Auf der anderen Seite verwendet C den Begriff der Arbeit einmal im Alltagskontext und einmal im physikalischen Kontext. Dabei überträgt er den Alltagskontext unreflektiert auf den physikalischen Kontext: «Arbeit will bezahlt werden; daher heißt das Gerät Arbeitsmessgerät».

Ideenebene. Fachimmanenter Kontext. Analog zum lebensweltlichen Kontext werden Exzerpte, die verschiedene Fachinhalte miteinander verbinden, zur Kategorie *Fachimmanenter Kontext* gezählt. Solche Bezüge finden sich in der aktuellen Experimentiereinheit wiederholt. Wie schon bei den lebensweltlichen Bezügen beobachtet werden konnte, werden solche meist nicht weiter ausgeführt oder vertieft.

Und jetzt wollen wir das >Arbeit und Energie< wieder aufgreifen. Wir fangen auch mit einem Versuch, mit einem kleinen Versuch an, der für uns jetzt hier nicht besonders schwer ist. (00:01:33)

Fangen wir hier an. Markus [S1: Wurde Arbeit geleistet] wie vorhin,... hinten? [S1: Wird das umgewandelt..., S2: Hubarbeit] Ja, wieder Hubarbeit. Was ist denn jetzt anders, Jonas? Was sind das? [S: Drehfelder] >Gelächter< Die sehen anders aus, ja aber jetzt mal noch ein bisschen auf das Wesentliche. Denk nochmal hier dran. (00:04:25) [7-27]

C beginnt anhand des Experiments das Thema Arbeit und Energie wieder aufzugreifen. Die Durchführung des Experiments wird dabei in ein fragend-entwickelndes Unterrichtsgespräch integriert. C geht, wie obige Exzerpte verdeutlichen, sehr kleinschrittig vor und richtet durch Hinweise, wie «hier anfangen» oder «daran denken» das Augenmerk gezielt auf bestimmte Elemente bzw. Aspekte des Experiments.

Sozial-kommunikative Ebene. Schülerbeiträge verarbeiten. Grosse Teile des Demonstrationsexperiments sind in ein fragend-entwickelndes Unterrichtsgespräch integriert (vgl. Abschnitt 7.3.3). Charakteristisch dafür sind die kurzen Fragen von C und die ebenso kurzen Antworten der Schülerinnen und Schüler. Sie scheinen teilweise als Stichwortgeber für die weiteren Ausführungen von C zu fungieren.

Was ist passiert? >lange Stille, C schaut in Klasse< Samuel [...]. Was war also meine Aufgabe hier? [S: Bewegung in mechanische Konstruktion] Was hab ich gemacht? [S: Na Sie haben Arbeit gemacht] Ich hab gearbeitet, ja und hier hinten? [S: Auch Arbeit verrichtet]. (00:01:56) [7–28]

Interessant ist hierbei, dass die Fragen der Lehrperson zunächst sehr offen formuliert werden. Längere Antworten wären damit also durchaus angebracht. Die Entstehung solcher wird aber durch die schnelle Abfolge von Fragen nicht unterstützt: Nach einem ersten Stichwort eines Schülers oder einer Schülerin stellt die Lehrperson bereits eine neue Frage, welche dabei korrigierend eine veränderte Zielrichtung vorgibt. Damit intendiert C, die Aufmerksamkeit auf bestimmte Aspekte des Experiments zu lenken. Die einer offenen Ausgangsfrage nachfolgenden Fragen werden zunehmend geschlossen formuliert:

Reingesteckt haben wir auf dieser Seite? [S: Bewegungsenergie] Ja, mechanische Energie, Bewegungsenergie, also mechanische. Auf der rechten Seite kam was an? [S: elektrische Energie, also elektrischer Strom] Ganz am Ende jetzt, sind wir schon. (Gemurmel). Klar, kam wieder Arbeit an. Meinetwegen jetzt hier Hubbardarbeit. (00:07:54) [7–29]

Die im Zusammenhang mit der Kategorie *Schülerbeiträge verarbeiten* erwähnten Exzerpte wurden der Subkategorie *Einfordern* zugeordnet. Zu der Kategorie zählen Äusserungen der Lehrperson, welche zu einer Erklärung bzw. zu einer Antwort der Schülerinnen und Schüler aufrufen. Andere mögliche Aspekte, die im Zusammenhang mit der Verarbeitung von Schülerbeiträgen auftreten können, wie das Paraphrasieren, Bewerten, Entwickeln oder das Einsehen, kommen kaum vor.

Sozial-kommunikative Ebene. Vorzeigen. Exzerpte zur erstmals eingesetzten Kategorie *Vorzeigen* repräsentieren einen neuen Aspekt experimentellen Handelns, der insbesondere bei Demonstrationsexperimenten auftritt: Lehrende oder Schülerinnen und Schüler zeigen im Plenum ein Experiment vor. Dabei gibt die Lehrper-

son Hinweise zu möglichen Beobachtungsschwerpunkten. Unterschieden werden die beiden Subkategorien *Selbst vorzeigen* und *Vorzeigen lassen*. Während zur ersten Subkategorie Exzerpte zugeordnet werden, bei denen es um Aspekte des Präsentierens geht, beinhaltet letztere Äusserungen des Lehrenden, die zum Vorzeigen eines Experiments aufrufen. Im aktuellen Fall sind beide Subkategorien vertreten:

Ihr guckt mal einfach zu und sollt das dann nachher mit den Begriffen, die wir kennen, beschreiben. (00:01:43)

Jetzt tausch ich das aus. Und jetzt? >S kurbelt< Ja, reicht. Kannst du mal so erzählen, was Dir aufgefallen ist? [S: Am Anfang wurde anscheinend weniger Arbeit gebraucht, also ich musste nicht so stark drehen und bei der Grossen musste ich halt stärker drehen, damit die Lampe leuchtet]. (00:10:35) [7–30]

Mit der ersten Äusserung beginnt C die Demonstration der mechanischen Vorrichtung (vgl. Kurzbeschreibung). Das zweite Exzerpt entspricht jener Stelle, bei der C eine Schülerin am Dynamo kurbeln lässt. Dabei fordert er sie auf, zwischen verschiedenen Lasten zu vergleichen. Hierbei lässt er der Schülerin genügend Zeit, ihre Wahrnehmung und eine daraus abgeleitete Interpretation zu formulieren.

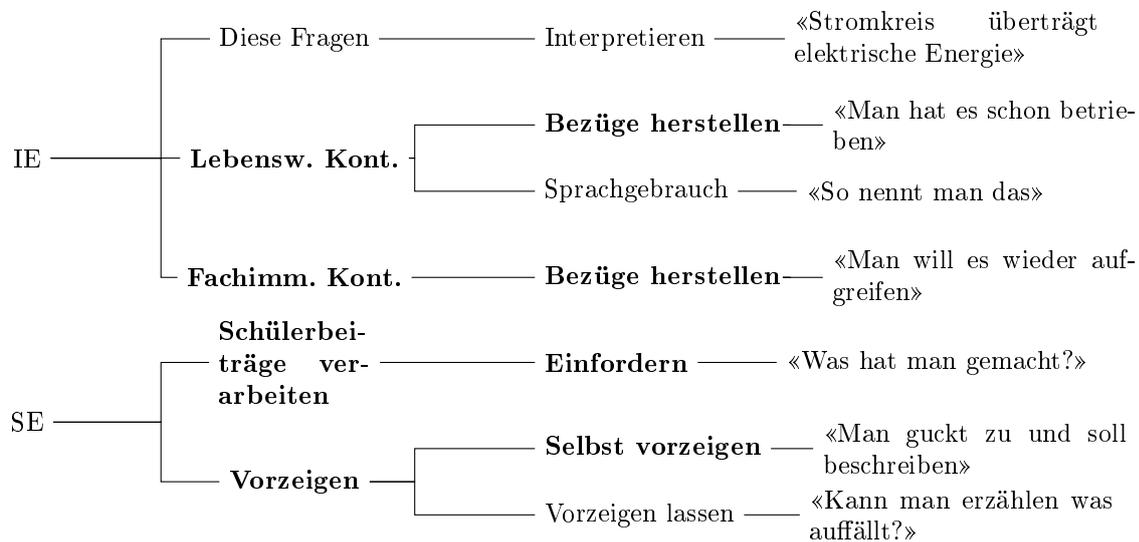


Abbildung 7.8: Strukturdiagramm der Ebenen (Ideenebene IE, Sozial-kommunikative Ebene SE), Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, D-1-29-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.

Falldiskussion

Im Vergleich zu den vorherigen Fällen dauert diese Experimentiereinheit deutlich kürzer (15min bzw. 76min und 45min). Innerhalb der kurzen Zeit gelingt es der Lehrperson C, verschiedene fachimmanente und lebensweltliche Bezüge aufzuzeigen. Sie finden auf mehreren Ebenen statt: Ein mechanisches System wird mit einem analogen elektrischen System verglichen, das elektrische System dient weiter als Prototyp für ein Energieversorgungssystem. Dabei werden die einzelnen Elemente dem realen System zugeordnet (vgl. [7–25] und [7–27]). Zusätzlich macht C die Schülerinnen und Schüler auf den physikalisch falschen Begriff “Stromzähler” aufmerksam. Interessanterweise geschieht das, indem er den Begriff der Arbeit unreflektiert einmal im Alltagskontext und dann wieder im fachlichen Kontext verwendet ([7–26]).

Es ist die eben erwähnte vertikale Vernetzung, welche die Fälle D-1-29-01 und F-0-16-01 ähnlich erscheinen lässt. Unterschiede zu F-0-16-01 lassen sich jedoch beim ausgeprägteren lebensweltlichen Bezug erkennen, ebenso bei der fehlenden *Objektebene* und der stattdessen hervortretenden *Sozial-kommunikativen Ebene*. Diese kann bei C wie folgt charakterisiert werden: Der Lehrer stellt im Rahmen des fragend-entwickelnden Unterrichtsgesprächs auffällig viele Fragen. Eine Sequenz von Fragen beginnt üblicherweise mit einer offenen Frage, auf die dann in zunehmendem Masse geschlossene Fragen folgen ([7–28] und [7–29]). Vertiefungsmöglichkeiten werden von C dabei kaum genutzt. Die Lernenden erscheinen als Stichwortgeber für die weiteren Ausführungen der Lehrperson. Hier wird ein Kontrast zwischen den Lehrpersonen A und C erkennbar. Im Gegensatz zu A gelingt es C nicht, Schülerinnen und Schüler zu längeren Ausführungen zu ermutigen (eine Ausnahme bildet [7–30]). Dadurch ergibt sich ein, sowohl auf der Objekt- wie auf der Ideenebene eng geführtes Demonstrationsexperiment, das durch eine starke Ablauforientierung⁴ geprägt ist.

Vergleich mit den Ratings

Die Transkriptanalyse ergibt in Bezug auf die Einschätzung der folgenden Facetten ein konsistentes Bild (vgl. Tabelle 7.8): die ausgeprägt vorhandene *Fachimmanente* und *Lebensweltliche Einbettung* (EB1, 1=trifft zu und EB2, 2=trifft grösstenteils zu), die durch das kleinschrittige Vorgehen bedingte fehlende *Zielklarheit* (ZK, 4=trifft nicht zu), der angemessene *Kognitive Anspruch* (KA, 1=trifft zu) sowie die mit dem

⁴Vgl. Fischler (1994).

modellhaften Vorgehen erklärbar weitgehend vorhandene *Theoretische Fundierung* (TF, 2=trifft grösstenteils zu).

Aus dem Transkript weniger deutlich treten die ausgeprägte *Begriffsbildung* (BB, 1=trifft zu) und die grösstenteils zutreffende *Ergebnisreflexion* (ER, 2=trifft grösstenteils zu) hervor. Eher wenig stimmig erscheint die zutreffende *Prozessreflexion* (PR, 1=trifft zu). Im Transkript finden sich nur vereinzelt Hinweise, die sich der Facette zuordnen liessen.

Sichtstruktur									
Vorb.		Durchf.		Nachb.		Σ			
4		2 _{DI}		9		15			
Tiefenstruktur									
FU	OF	EB1	EB2	ZK	PR	ER	KA	TF	BB
6	0000	1	2	4	1	2	1	2	1

Tabelle 7.8: Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall D-1-29-01: Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Das Experiment wird als Lehrerdemonstration durchgeführt (Abkürzungen der Facetten, vgl. Abbildung 7.1).

7.3.4 Experimentiereinheit 1, S-3-31-01

Energieabgabe des Generators

Die Lehrperson D beginnt die Experimentiereinheit mit der Repetition der wichtigen Grössen Stromstärke, Spannung und Energie. In einem fragend-entwickelnden Unterrichtsgespräch werden diese auf Grundgrössen zurückgeführt und ihre Einheiten angegeben. Anschliessend verknüpft D die erwähnten Grössen des elektrischen Stromkreises mit den analogen Grössen eines Wasserkreislaufs. Anhand einer Glühlampe, die an einem Generator mit Kurbel angeschlossen ist, demonstriert D das Zusammenspiel der besprochenen physikalischen Grössen und wirft die Frage auf, wie die vom Generator abgegebene Energie bei bekannter Spannung berechnet werden könnte. Die verschiedenen Schülervorschläge werden von D erörtert, weiter entwickelt oder verworfen. Dabei wechseln sich zwei unterrichtliche Arbeitsformen ab, das fragend-entwickelnde Unterrichtsgespräch und der Lehrervortrag. Die Zusammenhänge zwischen elektrischer Energie, Spannung, Stromstärke, Leistung und der Zeit werden so ermittelt. Das Ergebnis und wichtige Teilschritte hält D an der Wandtafel fest.

Transkriptanalyse

Für die Analyse der vorliegenden Exzerpte konnten die bereits bestehenden Ebenen, Kategorien und Subkategorien übernommen werden. Die einzige Ausnahme bildete die Kategorie *Gruppenmoderation* von Fall S-1-05-01 (vgl. Abbildung 7.6), welche in *Moderieren* umbenannt und damit erweitert wurde. Das ermöglichte die Zuordnung von Äusserungen von D, die eine Integration aller Schülerinnen und Schüler beim fragend-entwickelnden Unterrichtsgespräch zum Ziel hatten.

Ideenebene. Diese Fragen. Äusserungen zu der Kategorie verteilen sich hauptsächlich auf die Subkategorien *Problemstellen* und *Interpretieren*. Ersterer Subkategorie wurden jene Exzerpte zugeordnet, die die Ausgangsfrage des Experiments beinhalten. Auffallend dabei ist, dass D immer wieder zur Ausgangsfrage zurückkehrt. Damit verleiht D der Frage das angemessene Gewicht und ermuntert Schülerinnen und Schüler, sich vertieft mit ihr auseinanderzusetzen.

[...] wenn wir jetzt hier wissen wollen, wie viel Energie dort >zum Generator< raus kommt, wie können wir das jetzt ausrechnen? (00:13:15)

Ich möchte jetzt wissen, was für, für eine Stromrechnung wir bekommen, ja? Was für einen Energieverbrauch wir haben, nach einer gewissen Zeit, wenn wir eine Lampe laufen lassen, leuchten lassen, die ganze Zeit, jetzt zum Beispiel hier so ne Leuchtstoffröhre, oder ne Glühlampe, die ganze Zeit. (00:15:54) [7-31]

Wie aus den Exzerpten der Subkategorie *Problemstellen* hervorgeht, zeigt D ein großes Interesse daran, den Vorgang des Experiments sprachlich und begrifflich genau zu fassen. Das wird noch deutlicher bei den Exzerpten der Subkategorie *Interpretieren*. Hierbei verknüpft D Beobachtungen, Modellvorstellungen und Kontextbezüge miteinander. Dies geschieht auf verschiedenen kognitiven Anspruchsniveaus, wie die beiden folgenden Exzerpte zeigen:

So, und wenn wir jetzt ne ganze Weile drehen, dann kommt die ganze Zeit Energie raus und wenn wir ne Stunde drehen, dann wird dann eben ne Stunde lang hier Licht abgegeben und Wärme abgegeben, so wie bei uns jetzt hier die Lampen, die geben dauernd Energie ab, und dann warten wir ne Weile und dann nach nem Jahr kriegen wir dann die Stromrechnung. (00:14:12)

[...] Dann hätten wir ne Spannung von 100 Volt, genau [...] Wir haben ne Spannung vielleicht von 100 Volt oder von 230 Volt, das heisst 230 Joule pro Coulomb wird, wir brauchen 230 Joule um eine Ladungsmenge von einem Coulomb zu trennen, ja, hier, das wird hier getrennt und dann werden hier, wenn wir jetzt, angenommen wir hätten jetzt 230 Volt, ja, 230 Joule pro Coulomb hier wieder frei, pro Coulomb. (00:18:27) [7-32]

Ideenebene. Fachimmanenter Kontext. Fachimmanente Kontextbezüge treten bei D in vielfältigen Variationen auf. Dagegen kommen lebensweltliche Bezüge wenig vor. Wenn sie auftreten, werden sie stark an die Fachinhalte angeknüpft.

Im analogen Fall im, im äh Wasserkreislauf haben wir das ja auch. Was ist denn der Antrieb, wenn ein Wasser, wenn Wasser fließt, Mark? [...] Ja, die Erdanziehungskraft und wann fließt das Wasser? Die Erdanziehungskraft ist ja auch in einer Pfütze da. Da fließt aber jetzt noch kein Wasser, Lukas? [...] Ja, genau, die Höhendifferenz. (00:05:21) [7-33]

Sozial-kommunikative Ebene. Moderieren. Es lassen sich mehrere Transkriptstellen identifizieren, die mit der Integration von Schülerinnen und Schülern assoziiert werden können. Allerdings bleibt hierbei offen, ob es D vorwiegend um andere, neue Rückmeldungen geht, oder ob D tatsächlich passiv wirkende Schülerinnen und Schüler einbeziehen möchte.

Was schreiben wir da hin, Linus? [...] Ja, was meinen die anderen? Wie kürzen wir das ab? Ja, Christian? [...] Ja, E durch Q. Was hast du aber vorhin gesagt, sags nochmal? [...] Energie pro Ladungsmenge. (00:04:20) [7–34]

Sozial-kommunikative Ebene. Schülerbeiträge verarbeiten. Es lässt sich eine Vielzahl von Äusserungen identifizieren, bei denen die Subkategorien *Einfordern*, *Paraphrasieren*, *Bewerten* und *Entwickeln* treffend erscheinen. Eines der Beispiele wurde bereits im Zusammenhang mit der Kategorie *Fachimmanenter Kontext* aufgeführt ([7–33]).

Wie viel Energie ist das, wenn wir die Spannung wissen und zum Beispiel die Stromstärke oder die Ladung, ja? [...] Ja, also wenn wir den Widerstand hätten, genau, U gleich R mal I , das stimmt, aber ich möchte jetzt nicht U haben, sondern ich möchte die Energie haben, die dort raus kommt. (00:15:15)

So, und die potentielle Energie am Pluspol ist eben höher als die potentielle Energie am Minuspol und dann wird das, äh, diese Ladung angetrieben und fließt durch die Lampe und was passiert dann in der Lampe? Hm? Was passiert in dieser Lampe, Simon? [...] Licht, ja also, Licht, genau, und was gibts denn noch ausser Licht? Meistens auch noch, Timo? [...] Bitte? [...] Richtig, Licht und Wärme und was ist das, Licht und Wärme, ja? [...] Was denn? [...] Ja, genau, das ist auch Energie, ja. Das heisst die Energie wird wieder frei ja. (00:07:59) [7–35]

Auf die Frage nach der Energie, die im Stromkreis umgesetzt wird, entgegnet eine Schülerin “ U gleich R mal I ”. Obwohl der Beitrag der Schülerin in dem Zusammenhang unverständlich ist, nimmt D ihn auf. Das geschieht zunächst dadurch, dass D den Beitrag bestätigt, ihn paraphrasiert und bewertet, danach jedoch erneut auf die eigentliche Fragestellung hinweist. Wie aus dem zweiten Exzerpt hervorgeht, reihen sich an eine offene Ausgangsfrage öfters weitere Fragen, die in der Tendenz zunehmend geschlossen formuliert werden. Diesbezüglich ist D durchaus vergleichbar mit Lehrperson C.

Sozial-kommunikative Ebene. Vorzeigen. Die Demonstration des Generator-Glühlampen-Systems geschieht eher beiläufig. Sie wird wiederholt unterbrochen, wenn D Erkenntnisse an der Tafel festhält, vorträgt oder die Lernenden befragt, und später wieder aufgenommen. So ordnet sich das Handeln am Objekt bei D den mentalen Handlungen unter.

Wir haben hier eine Spannungsquelle, >hantiert an Geräten auf dem Tisch< ja? So. >kurbelt< Jetzt ist auch Energie durchgetrieben, durch diese Glühbirne und wenn wir mehr viel Energie reinstecken, dann kommt natürlich auch mehr Energie raus, ja? (00:12:56) [7-36]

Der Aussage über die Energie scheint D durch das Kurbeln am Generator besonderes Gewicht zu verleihen.

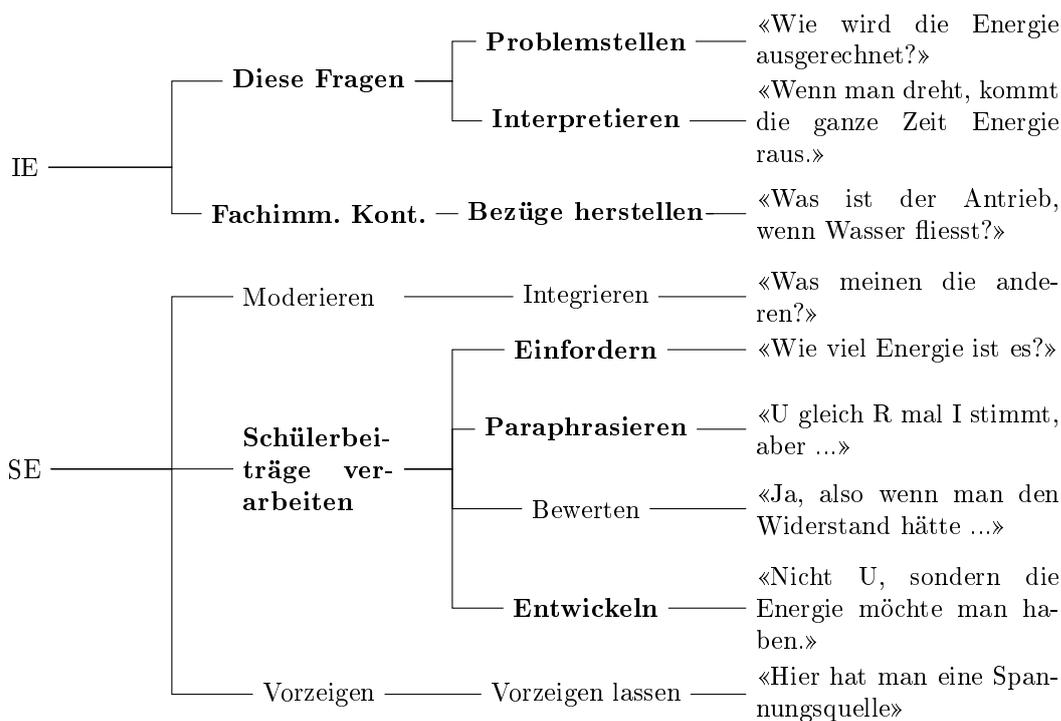


Abbildung 7.9: Strukturdiagramm der Ebenen (Ideenebene IE, Sozial-kommunikative Ebene SE), Kategorien und Subkategorien als Ergebnis der Transkriptanalyse aus Experimentiereinheit 1, S-3-31-01. Fett gedruckt werden jene Kategorien bzw. Subkategorien, die bei der Analyse besonders ausgeprägt hervortreten. In «» ist zu jeder Subkategorie eine erklärende Paraphrase angegeben.

Falldiskussion

D zeigt generell ein grosses Interesse daran, physikalische Sachverhalte sprachlich zu fassen. Das wird einerseits durch die Genauigkeit ersichtlich, mit der D Vorgänge modellhaft beschreibt oder Beiträge von Lernenden kommentiert ([7–32], [7–33] und [7–35]). Das Interesse wird andererseits durch die Art der Durchführung unterstrichen: Die Demonstration ist in einen Lehrervortrag mit fragend-entwickelnden Intervallen eingeflochten ([7–36]). Dabei sind die Handlungen am Objekt durch die mentalen Handlungen von D bestimmt. D versteht es besonders gut, zwischen fachimmanenten Kontexten Bezüge herzustellen und Analogien aufzuzeigen.

Die aus dem Transkript hervorgehenden Merkmale der Experimentiereinheit S-3-31-01 weisen Ähnlichkeiten mit jenen der vorhergehenden Einheit D-1-29-01 auf. Sie beschränken sich nicht nur auf die Gemeinsamkeiten der Sichtstruktur: Beide Einheiten beinhalten Demonstrationsexperimente, welche in fragend-entwickelnde Unterrichtsgespräche eingebettet sind. Die Lehrer C und D stellen beide viele Fragen. Typischerweise beginnt eine Fragesequenz mit einer einleitenden eher offenen Frage, auf die zunehmend geschlossene Fragen folgen. Die Objektebene hat bei beiden Einheiten eine untergeordnete Bedeutung und dient mehr dazu, Äusserungen zu illustrieren, als eigentliche Beobachtungen anzustellen. Trotz der Gemeinsamkeiten hebt sich die aktuelle Einheit deutlich von D-1-29-01 ab: Während C bei Schülerbeiträgen kaum nachhakt und dabei Möglichkeiten der Vertiefung nicht ausschöpft, gelingt es D, sie zu nutzen (vgl. z.B. [7–35]). Dies scheint auch die Qualität der Schülerbeiträge zu verbessern. Bei den Strukturdiagrammen lässt sich das daran erkennen, dass bei S-3-31-01 die Subkategorien *Paraphrasieren*, *Bewerten* und *Entwickeln* vorkommen, während sie bei D-1-29-01 fehlen (vgl. Abbildung 7.9 und 7.8). Bei D fehlt hingegen ein lebensweltlicher Bezug weitgehend.

Vergleich mit den Ratings

Die Ratings der Experimentiereinheit stimmen weitgehend mit den Befunden der Transkriptanalyse überein (vgl. Tabelle 7.9). Das starke Interesse von D, physikalische Vorgänge verbal zu beschreiben, verbunden mit der Hervorhebung von Analogien und dem Denken in Modellen, ist mit den ausgeprägt vorhandenen Perspektiven *Reflexion* und *Theorieleitung* stimmig (PR, KA, TF, BB, 1=*trifft zu* und ER, 2=*trifft grösstenteils zu*). Auch passend ist die durch das kleinschrittige Vorgehen

bedingte mangelnde Zielklarheit (ZK, 4=trifft nicht zu). Einzig der aus Sicht der Transkriptanalyse weniger vorhandene lebensweltliche Bezug passt nicht zum Rating der *Lebensweltlichen Einbettung* (EB2, 2=trifft grösstenteils zu).

Sichtstruktur									
Vorb.		Durchf.		Nachb.		Σ			
12		4 _{Dl}		7		23			
Tiefenstruktur									
FU	OF	EB1	EB2	ZK	PR	ER	KA	TF	BB
6	0000	2	4	3	1	3	1	1	1

Tabelle 7.9: Zusammenfassung der Sicht- und Tiefenstrukturmerkmale des Experiments 1 von Fall S-3-31-01: Angegeben ist die Dauer der Vorbereitung (Vorb.), der Durchführung (Durchf.) und der Nachbereitung (Nachb.) in Minuten. Das Experiment wird als Lehrerdemonstration durchgeführt (Abkürzungen der Facetten, vgl. Abbildungen 7.1).

7.3.5 Übersicht über die Kategorien der Transkriptanalyse

Die Analyse der vier Fälle umfasst einen Prozess (vgl. Abschnitte 7.3.1-7.3.4). Beispielsweise führte die Transkriptanalyse des Falls S-1-05-01 zur Kategorie *Um was geht es?*. Im Verlauf der Analyse von F-0-16-01 erwies sich die allgemeinere Bezeichnung *Problemstellen* als passender. Ähnliche Anpassungen ergaben sich auch bei weiteren Kategorien, bzw. Subkategorien.

Die nachfolgenden Tabellen 7.10-7.12 geben eine Übersicht über verwendete Kategorien und Subkategorien und fassen dabei auch Veränderungen während des Analyseprozesses zusammen. Der Aufbau der drei Tabellen ist identisch: In der linken Spalte werden Subkategorien zu den entsprechenden Ebenen und Kategorien (vgl. Zeilenüberschriften in Tabellen 7.10-7.12) aufgeführt. Die aus dem gesamten Analyseprozess hervorgehenden Subkategorien wurden fett gedruckt. Jeweils darunter sind Fälle aufgeführt, bei denen die entsprechende Subkategorie erstmals zur Anwendung kam. Z.B. lässt sich bei *Problemstellen* (erste Subkategorie, Tabelle 7.10) ablesen, dass diese erstmals bei S-1-05-01 mit der Bezeichnung *Um was geht es?* zur Anwendung kam und im Fall F-0-16-01 zur endgültigen Bezeichnung umbenannt wurde. Die mittlere Spalte beinhaltet kurze Beschreibungen der Subkategorien, die durch charakteristische Paraphrasen ergänzt werden (Spalte links).

Anhand der so aufbereiteten Kategorien, wurden die Transkripte aller vier Fälle (S-1-05-01, F-0-16-01, D-1-29-01 und S-3-31-01) erneut analysiert, wobei alle Ka-

tegorien und Subkategorien unverändert blieben. Tabelle 7.13 zeigt die relativen Häufigkeiten⁵ der Zuordnungen aller Kategorien. Bei S-1-05-01 entfallen somit 30% der Zuordnungen auf *Diese Fragen*, 10% auf *Handlungen* bzw. *Moderieren* und 50% auf *Schülerbeiträge verarbeiten*.

Subkategorie	Beschreibung	Paraphrase
<i>Ideenebene. Diese Fragen.</i>		
Problemstellen ▷ S-1-05-01: Um was geht es? ▷ F-0-16-01	Die LP wirft eine Problemstellung im Zusammenhang mit dem Experiment auf oder verdeutlicht diese.	«Um was geht es?»
Gestalten ▷ S-1-05-01: Was wird gebaut? ▷ F-0-16-01	Die LP äussert sich zum experimentellen Aufbau bzw. begründet oder hinterfragt ihn.	«Was wird gebaut?»
Interpretieren ▷ S-1-05-01: Warum ist man sicher, ...? ▷ F-0-16-01	Die LP interpretiert Beobachtungen oder Erkenntnisse aus dem Experiment oder fordert Schülerinnen und Schüler dazu auf.	«Warum ist man sicher, dass es geht?»
<i>Ideenebene. Lebensweltlicher Kontext.</i>		
Sprachgebrauch ▷ F-0-16-01	Die LP reflektiert Alltags- oder Fachbegriffe und leitet Schülerinnen und Schüler zu einer adäquaten Ausdrucksweise an.	«Verbraucht wird in Anführungszeichen gesetzt.»
Bezüge herstellen ▷ D-1-29-01	Die LP stellt Bezüge zur Lebenswelt der Schülerinnen und Schüler her.	«Man hat es schon betrieben.»
<i>Ideenebene. Fachimmanenter Kontext.</i>		
Bezüge herstellen ▷ F-0-16-01: Erinnern ▷ D-1-29-01	Die LP stellt Bezüge zwischen verschiedenen Fachinhalten her oder greift bereits Behandeltes wieder auf.	«Man will es wieder aufgreifen.»

Tabelle 7.10: Kategorien und Subkategorien der Ideenebene: *Subkategorie:* Bezeichnung der Subkategorien (fett), Fall: Ursprüngliche Bezeichnung; *Beschreibung:* Kurze Beschreibung der Subkategorie; *Paraphrase:* Paraphrase zur entsprechenden Subkategorie.

⁵Wegen Rundungsfehlern ergänzen sich die Prozentangaben nicht immer zu hundert Prozent (Werte gerundet auf eine signifikante Ziffer).

Subkategorie	Beschreibung	Paraphrase
<i>Objektebene. Messungen.</i>		
Erklären ▷ F-0-16-01	Die LP äussert sich im Zusammenhang mit der Messung bzw. Messgeräten: Es geht um Grösse und Betrag der Messung, um die Einstellung und Ablesung des Messgeräts usw.	«Messen, dann darüber nachdenken.»
Vergleichen ▷ S-0-16-01	Die LP vergleicht mehrere Messwerte miteinander oder ruft Schülerinnen und Schüler dazu auf.	«Man schaut, wie die Ergebnisse aussehen.»
<i>Objektebene. Handlungen.</i>		
Anweisen ▷ F-0-16-01	Die LP ruft dazu auf, ein Experiment nach den vorgegebenen Schritten durchzuführen oder formuliert ad hoc einen Ablaufplan.	«Man schaut nach, wie es geht.»

Tabelle 7.11: Kategorien und Subkategorien der Objektebene: *Subkategorie:* Bezeichnung der Subkategorien (fett), Fall: Ursprüngliche Bezeichnung; *Beschreibung:* Kurze Beschreibung der Subkategorie; *Paraphrase:* Paraphrase zur entsprechenden Subkategorie.

Subkategorie	Beschreibung	Paraphrase
<i>Sozial-kommunikative Ebene. Moderieren.</i>		
Anleiten ▷ S-1-05-01	Die LP thematisiert die Zusammenarbeit in der Klasse oder in einer Gruppe: Arbeitsorganisation, Arbeitsteilung oder Konflikte können Gegenstand der Äusserung sein.	«Man empfiehlt, miteinander zu diskutieren.»
Integrieren ▷ S-1-05-01	Die LP bezieht Schülerinnen und Schüler in ein Gespräch mit ein.	«Hat man auch eine Idee?»
Austausch fördern ▷ S-1-05-01	Die LP ruft zum Austausch des Arbeitsstandes unter den Lernenden auf.	«Was hat man gemacht, gebaut, gedacht und probiert.»
<i>Sozial-kommunikative Ebene. Schülerbeiträge verarbeiten.</i>		
Einfordern ▷ S-1-05-01	Die LP stellt eine Frage ans Plenum, an eine Gruppe, einen oder eine Schülerin.	«Wie viel Energie ist es?»
Paraphrasieren ▷ S-1-05-01	Die LP paraphrasiert Schülerbeiträge.	«Will man das sagen?»
Einsehen ▷ S-1-05-01	Die LP bestätigt durch eine Äusserung, dass sie einen Schülerbeitrag verstanden bzw. nicht verstanden hat.	«Man sieht es ein.»
Bewerten ▷ S-1-05-01	Die LP bewertet einen Schülerbeitrag.	«Es wäre schon gut, wenn man es ändern würde.»
Entwickeln ▷ S-1-05-01	Die LP nimmt einen Schülerbeitrag auf und entwickelt ihn weiter.	«Kann man die Lösung ändern?»
<i>Sozial-kommunikative Ebene. Vorzeigen.</i>		
Selbst vorzeigen ▷ D-1-29-01	Die LP demonstriert ein Experiment. Die diesbezüglichen Äusserungen beziehen sich auf die Durchführung und auf Hinweise beim Zuschauen.	«Man guckt zu und soll beschreiben.»
Vorzeigen lassen ▷ D-1-29-01	Die LP ruft zur Demonstration eines Experiments auf.	«Kann man erzählen, was auffällt?»

Tabelle 7.12: Kategorien und Subkategorien der sozial-kommunikativen Ebene: *Subkategorie:* Bezeichnung der Subkategorien (fett), Fall: Ursprüngliche Bezeichnung; *Beschreibung:* Kurze Beschreibung der Subkategorie; *Paraphrase:* Paraphrase zur entsprechenden Subkategorie.

Ebene	Kategorie	Fälle			
		S-1-05-01	F-0-16-01	D-1-29-01	S-3-31-01
Ideenebene	Diese Fragen	30	20	10	20
	Lebensweltlicher Kontext	0	0	10	0
	Fachimmanenter Kontext	0	20	10	20
Objektebene	Messungen	0	20	0	0
	Handlungen	10	20	0	0
Soz.-komm. Ebene	Moderieren	10	0	0	0
	Schülerbeiträge verarbeiten	50	20	40	50
	Vorzeigen	0	10	20	10

Tabelle 7.13: Relative Häufigkeiten der zu einer Kategorie zugeordneten Exzerpte. Angaben in Prozenten bezogen auf das Total der zugeordneten Exzerpte (gerundet auf eine signifikante Ziffer).

Kapitel 8

Experimentelles Handeln und der Zuwachs an Fachwissensleistung

Im Kapitel 4 wurde das systemische Modell von Lipowsky et al. (2005) als theoretische Basis des Gesamtprojekts vorgestellt (vgl. Abbildung 4.1). Wesentliches Element des Modells ist der Zusammenhang zwischen der Ebene des Unterrichts und den Ausgangsgrößen (Outputvariablen). Das vorliegende Kapitel widmet sich der Frage, ob zwischen den experimentellen Handlungsmustern im Unterricht und den Outputvariablen Zusammenhänge nachgewiesen werden können:

Z1 Lässt sich zwischen der im Kontext experimentellen Handelns stehenden Unterrichtszeit und dem Leistungszuwachs ein Zusammenhang nachweisen?

Z2 Lassen sich zwischen den durch die Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung* definierten Qualitätsmerkmalen und dem Leistungszuwachs Zusammenhänge nachweisen?

Einschränkend muss festgehalten werden, dass experimentelles Handeln nur einen Teil des Unterrichts ausmacht. Daher ist davon auszugehen, dass sich im Rahmen des vorgegebenen Designs nur ganz deutliche Zusammenhänge zwischen dem experimentellen Handeln und einer Outputvariable überhaupt nachweisen lassen. Auch wenn der überwiegende Teil der Doppelstunden als typisch für den normalen Unterricht gelten kann¹, ist es des Weiteren problematisch, von Merkmalen einer Doppelstunde auf den Leistungszuwachs eines Halbjahres zu schliessen (vgl. Abbildung 4.2).

¹Erfahrungen aus anderen Videostudien bestätigen, dass videographierter Unterricht überwiegend als repräsentativ für den normalen Unterricht gelten kann (Stigler et al., 2000).

Die angedeuteten Schwierigkeiten sowie die Fokussierung auf die Entwicklung und den Einsatz von Kategoriensystemen experimentellen Handelns führen dazu, dass die Zusammenhänge zwischen Unterrichtsebene und Outputvariablen im Rahmen der vorliegenden Arbeit nicht erschöpfend abgehandelt werden können. Die Befunde des Abschnitts haben daher hypothesengenerierenden Charakter.

Es werden erste Ansätze skizziert, wie die Daten der Sicht- und Tiefenstruktur mit Outputvariablen verbunden werden können. Das geschieht im Bewusstsein, dass es neben dem angewandten Verfahren noch adäquatere mathematische Modelle gibt, wie z.B. Mehrebenenmodelle. Die Implementierung und Interpretation solcher Modelle geht über den Umfang des Buches hinaus und muss deshalb an anderer Stelle erfolgen (vgl. Forschungsdesiderata, Abschnitt 9.9.2).

8.1 Leistungsdaten

Schülerleistungen wurden im Rahmen des Pre- und Posttests mit dem *Content Knowledge Paper and Pencil Test* (CKPT) erhoben. Das Instrument umfasst offene und geschlossene Items unterschiedlicher Schwierigkeitsgrade zu den Themenbereichen Energie und Elektrizität und wurde von Geller et al. (2010, in Vorb.), Geller (in Vorb.) entwickelt. Die Modellierung der Daten erfolgte nach einem eindimensionalen dichotomen Raschmodell (vgl. Wilson, 2005, Rost, 1996). Eingang in die Berechnungen der Abschnitte 8.2 und 8.3 findet die auf Klassenebene ausgewiesene mittlere Differenz der Schülerfähigkeiten zwischen den beiden Testzeitpunkten (Leistungszuwachs).

Der CKPT wurde inhaltlich auf die Doppelstunde zum Thema “Zusammenhang zwischen elektrischer Energie und Leistung” abgestimmt (vgl. Kapitel 4). Ausserdem umfasst er Themenbereiche, wie die mechanische Energie und Leistung sowie die Beschreibung elektrischer Stromkreise. Die Projektpartner der drei Länder prüften den Test hinsichtlich seiner Curriculumsvalidität. Der Test wurde in den drei Ländern pilotiert.

Insgesamt kamen 54 Aufgaben zum Einsatz, wobei jede Schülerin, jeder Schüler 18 Aufgaben im Pretest, die restlichen 36 Aufgaben im Posttest bearbeitete. Das Ensemble der Aufgaben setzte sich aus 40 Multiple-Choice sowie 14 offenen Aufgabenformaten zusammen. Die Aufgabenkonstruktion erfolgte nach dem Kompetenzmodell von Kauertz (2008), welches die systematische Variation der Aufga-

benschwierigkeit erlaubte. Dabei wurde nur eine Komplexitätsachse berücksichtigt. Durch die Rotation der Aufgaben konnte sichergestellt werden, dass alle Aufgaben sowohl im Pre- als auch im Posttest-Zeitpunkt eingesetzt wurden.

Der Einfluss der theoretischen Komplexität (Kauertz, 2008) auf die empirisch ermittelte Aufgabenschwierigkeit wurde zur Prüfung der Konstruktvalidität untersucht. Hierbei zeigte sich ein signifikanter Effekt der Komplexität auf die Schwierigkeit, was die Validität des Tests belegte.

Die Testergebnisse zeigen, dass sich bei finnischen und Schweizer Schülerinnen und Schüler ein signifikanter Leistungszuwachs ergibt, wobei erstere am besten abschneiden (Abbildung 8.1, Geller et al. (in Vorb.)). Dagegen ist bei der deutschen Stichprobe kein signifikanter Zuwachs zu verzeichnen. Diese Befunde entsprechen den Erwartungen aus den PISA-Studien (OECD, 2001, 2004, 2007, 2010).

Die in Abbildung 8.1 summarisch dargestellten Daten, werden im nächsten Abschnitt mit den Ergebnissen aus der Sicht- und Tiefenstrukturanalyse verbunden.

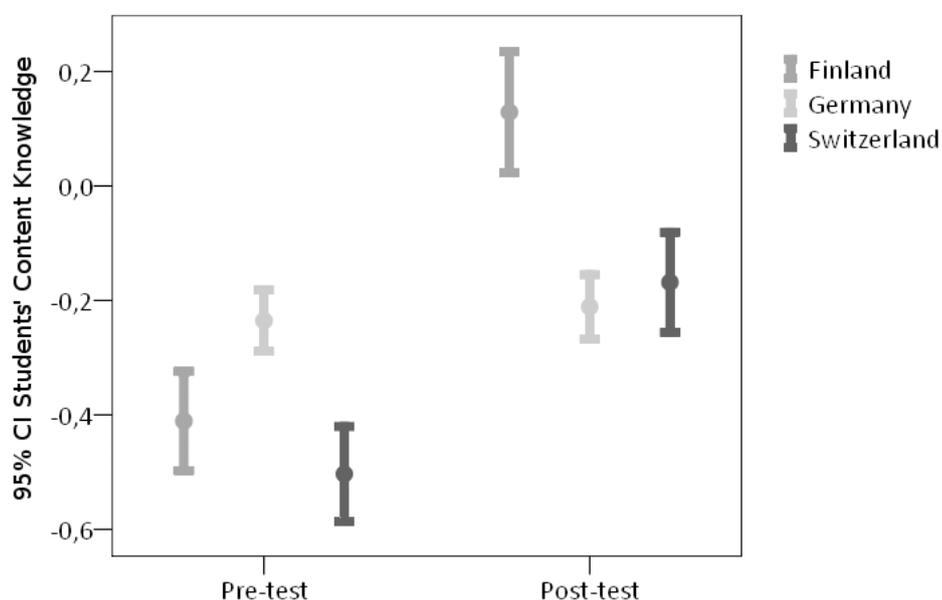


Abbildung 8.1: Das Diagramm aus Geller et al. (in Vorb.) zeigt die Mittelwerte und die 95%-Konfidenzintervalle der Schülerfähigkeiten beim Pre- und beim Posttest in Finnland, Deutschland und der Schweiz. *Students' Content Knowledge*: Der Wert Null entspricht dem Mittelwert der Aufgabenschwierigkeit.

8.2 Der Zusammenhang zwischen der Sichtstruktur experimentellen Handelns und dem Leistungszuwachs

Tesch (2005) vermutet auf der Grundlage der Daten der Projektphase I (vgl. Abschnitt 3.5), dass zwischen der insgesamt im Kontext experimentellen Handelns stehenden Unterrichtszeit und dem Leistungszuwachs ein positiver Zusammenhang besteht. Die Vermutung soll anhand der vorliegenden Daten erneut geprüft werden.

Insgesamt werden drei Modellannahmen miteinander verglichen. Der Leistungszuwachs (unabhängige Variable) wird mit $\bar{\Delta}L$, die insgesamt im Kontext experimentellen Handelns stehende Unterrichtszeit (unabhängige Variable) mit T_{Exp} bezeichnet. Um den Einfluss der verschiedenen Länderstichproben zu berücksichtigen werden zwei Dummy-Variablen eingeführt:

$$d_{FIN} = 1, \text{ falls finnische Klasse, } 0 \text{ sonst}$$

$$d_{CH} = 1, \text{ falls Schweizer Klasse, } 0 \text{ sonst}$$

Folgende Modelle werden verwendet:

$$\text{Modell 1: } \bar{\Delta}L = b_0 + b_1 T_{Exp}$$

$$\text{Modell 2: } \bar{\Delta}L = b_0 + b_1 T_{Exp} + b_2 d_{FIN} + b_3 d_{CH}$$

$$\text{Modell 3: } \bar{\Delta}L = b_0 + b_1 T_{Exp} + b_2 d_{FIN} + b_3 d_{CH} + b_4 T_{Exp} d_{FIN} + b_5 T_{Exp} d_{CH}$$

Das Modell 1 berücksichtigt den Zusammenhang zwischen Experimentierzeit (inkl. Vor- und Nachbereitungszeit) und dem Leistungszuwachs. Nicht einbezogen ist der Umstand, dass verschiedene Populationen (Länderstichproben) vorliegen. In Modell 2 wird dies berücksichtigt. Bei einer Experimentierzeit von 0 sind hier für jedes Land verschiedene Leistungszuwächse möglich. Bei allen Ländern konstant gehalten wird hingegen die Differenz des Leistungszuwachses pro Differenz der Experimentierzeit (d.h. der Differenzenquotient). Geometrisch betrachtet setzt sich das Modell damit aus drei in der Ebene parallel verschobenen Geraden zusammen. Zusätzlich zur Verschiebung der Geraden erlaubt das Modell 3 unterschiedliche Steigungen, d.h. unterschiedliche Änderungsraten des Leistungszuwachses. Für die drei Modelle werden Regressionsanalysen durchgeführt. Anhand eines F-Tests kann zwischen den

Land	Shapiro-Wilk-Test		
	n	W	p
D	43	0.94	.03*
FIN	25	0.94	.12
CH	31	0.96	.23

Tabelle 8.1: Ergebnis des Shapiro-Wilk-Tests für die drei Länderstichproben. Bei der deutschen Stichprobe lässt sich die Annahme der Normalverteilung nicht mehr aufrecht erhalten (H_0 -Hypothese: Es liegt eine Normalverteilung vor). n Stichprobengröße, W Teststatistik, p Irrtumswahrscheinlichkeit.

Modellen verglichen werden. Dabei soll das adäquate Modell die Daten angemessen beschreiben und gleichzeitig von möglichst einfacher Gestalt sein (Parsimonitätsprinzip).

Die Stichproben erfüllen nicht alle Voraussetzungen für die Modellierung nach dem Standardverfahren² (vgl. Kapitel 6). So lässt sich anhand des Shapiro-Wilk-Tests eine Normalverteilung der deutschen Länderstichprobe nicht belegen (vgl. Tabelle 8.1). Dennoch führte die Analyse nach dem Standardverfahren zu den gleichen Ergebnissen wie die Anwendung eines robusten Modells³. Der Einfachheit halber werden im Folgenden Ergebnisse aus dem Standardverfahren präsentiert.

Werden die hierarchischen Modelle 1 bis 3 paarweise miteinander verglichen, ergibt sich folgendes Bild: Von Modell 1 zu Modell 2 kann eine signifikant bessere Modellanpassung nachgewiesen werden ($F(2, 95) = 12.6, p < .001$). Die Verbesserung von Modell 2 zu 3 ist dann nicht mehr signifikant ($F(2, 93) = 0.07, p < .930$). Nach dem Parsimonitätsprinzip ist damit das Modell 3 zu verwerfen und 2 anzunehmen. Aus Modell 2 lässt sich erkennen, dass sich die y -Achsenabschnitte (Achsenabschnitt Leistungszuwachs) der Ländergeraden Finnland und Schweiz signifikant von jener für Deutschland unterscheiden. Dagegen kann zwischen der im Kontext experimentellen Handelns stehenden Unterrichtszeit T_{Exp} und dem Leistungszuwachs ΔL kein Zusammenhang nachgewiesen werden.

²Gemeint ist die Regression nach der Methode der kleinsten Quadrate.

³Verwendet wurde die Bibliothek *robustbase* (Rousseeuw et al., 2011), basierend auf Yohai (1987).

		b_i	SE b_i	Sign.
<i>Modell 1</i>	Konst.	0.482	0.086	***
	T_{Exp}	-0.005	0.002	**
<i>Modell 2</i>	Konst.	0.072	0.113	
	T_{Exp}	-0.001	0.002	
	d_{FIN}	0.477	0.098	***
	d_{CH}	0.272	0.083	**
<i>Modell 3</i>	Konst.	0.047	0.165	
	T_{Exp}	0.000	0.003	
	d_{FIN}	0.531	0.201	**
	d_{CH}	0.275	0.231	
	$T_{Exp} d_{FIN}$	-0.001	0.004	
	$T_{Exp} d_{CH}$	0.000	0.004	

Tabelle 8.2: Ergebnisse der Regressionsanalyse (Modell 1, 2 und 3) nach dem Standardverfahren (Methode der kleinsten Quadrate). b_i Koeffizienten, SE b_i Standardabweichung der Koeffizienten, Sign. Signifikanz der Koeffizienten: * $p < .05$, ** $p < .01$ und *** $p < .001$. Modell 1: $R^2 = .09$, $F(1, 97) = 9.7$, $p = 0.002$; Modell 2: $R^2 = .28$, $F(3, 95) = 12.4$, $p < 0.001$; Modell 3: $R^2 = .28$, $F(5, 93) = 7.3$, $p < 0.001$.

8.3 Der Zusammenhang zwischen der Tiefenstruktur experimentellen Handelns und dem Leistungszuwachs

Ähnlich wie beim vorherigen Abschnitt werden die Facetten der Tiefenstruktur experimentellen Handelns mit dem Leistungszuwachs verbunden. Wegen des ordinalen Skalenniveaus der Facetten kam hierfür eine Kovarianzanalyse (Kovariante Land) ohne Berücksichtigung von Interaktionseffekten zur Anwendung. Zuvor kodierte man die vierstufigen Ratings der Facetten in eine binäre Skala um (Dichotomisierung).

Wie aus Tabelle 8.3 ersichtlich ist, haben die Facetten *Fachimmanente Einbettung* und *Ergebnisreflexion* einen signifikanten Effekt (mittlere Effektstärke) auf den mittleren Zuwachs an Fachwissensleistung. Knapp nicht signifikant ist der Zusammenhang zwischen dem *Kognitiven Anspruch* und der unabhängigen Variablen.

Facette	F	p	Sign.	η^2
Fachimm. Einbettung	4.6	.04	*	.07
Lebensweltl. Einbettung	1.5	.22	-	.03
Zielklarheit	1.0	.31	-	.02
Prozessreflexion	2.8	.10	-	.05
Ergebnisreflexion	8.0	.01	**	.12
Kognitiver Anspruch	3.3	.07	.	.05
Theoretische Fundierung	0.2	.66	-	.00
Begriffsbildung	0.3	.60	-	.00
Land	11.	.00	***	.28

Tabelle 8.3: Kovarianzanalyse zwischen der abhängigen Variablen *Leistungszuwachs* und der unabhängigen Variablen (Facette) sowie der Kovariate *Land*. Spalten: Facette, F (F-Statistik), p (Irrtumswahrscheinlichkeit), Signifikanz (- ($p \geq .1$), . ($p < .1$), * ($p < .05$), ** ($p < .01$), *** ($p < .001$)), η^2 (Effektstärke).

Teil IV

Diskussion und Ausblick

Kapitel 9

Diskussion und Ausblick

Im aktuellen Kapitel werden die Ergebnisse aus den Kapiteln 6, 7 und 8 noch einmal zusammengefasst¹ und anschliessend diskutiert. Gegenstand der Analysen sind Doppelstunden aus dem Physikunterricht des 9. bzw. 10. Schuljahres² aller Schultypen der Länder Deutschland, Finnland und Schweiz zum Thema “Zusammenhang zwischen elektrischer Energie und Leistung”. Die deutsche Stichprobe kann näherungsweise als repräsentativ für das Bundesland Nordrhein-Westfalen, die finnische für Mittelfinnland (Umkreis von 200km um Jyväskylä) und die Schweizer Stichprobe für die Deutschschweiz gelten.

9.1 Zusammenfassung zentraler Ergebnisse der Sichtstruktur experimentellen Handelns

Insgesamt 99 videographierte Doppelstunden (D 43, FIN 25, CH 31) wurden bezüglich ihrer experimentellen Handlungsmuster untersucht. Hierfür wurde das Kategoriensystem von Tesch (2005) adaptiert (vgl. Abbildung 9.1 und Kapitel A). Die Analyse der Videos erfolgte zeitbasiert mit einem Analyseintervall von 20 Sekunden. Bis auf die Kategorie *Sachbegegnung* konnten zwischen den Kodierenden zufriedenstellende Übereinstimmungen ($\kappa > .7$) erzielt werden (vgl. Kapitel 5).

¹Einzelne wichtige Abbildungen aus früheren Kapiteln erscheinen hier nochmals: (Abbildungen 9.1, 9.2, 9.3 und 9.4) entsprechen (Abbildungen 5.1, 6.1, 7.1 und 7.2).

²Deutschland: 10. Schuljahr, Schweiz und Finnland 9. Schuljahr.

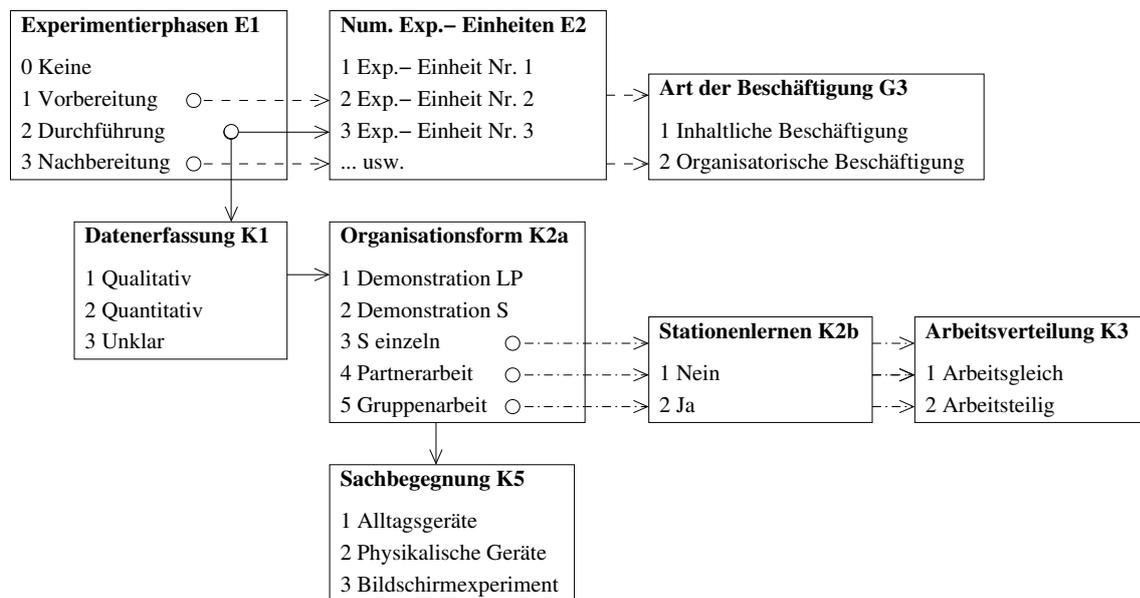


Abbildung 9.1: Das Kategoriensystem der Sichtstruktur experimentellen Handelns basierend auf Tesch (2005) (entspricht Abbildung 5.1).

S1 & S2 Welche experimentellen Handlungsmuster lassen sich im videografierten Unterricht identifizieren? Inwiefern bestehen zwischen den Ländern Unterschiede bei den experimentellen Handlungsmustern?

9.1.1 Unterrichtszeit im Kontext experimentellen Handelns

Die Ergebnisse zeigen, dass rund die Hälfte der Unterrichtszeit (49min) im Kontext experimentellen Handelns stattfindet. Damit sind diejenigen Zeiten gemeint, während denen Experimente durchgeführt, vor- oder nachbereitet³ werden. Bezogen auf die Länderstichproben ergeben sich signifikante Unterschiede (vgl. Abbildung 9.2 und Tabelle 6.2). Während in Deutschland über zwei Drittel der Unterrichtszeit im Kontext experimentellen Handelns stehen, macht der entsprechende Anteil in Finnland nur gerade knapp ein Drittel aus (D 65min, FIN 28min). Der Schweizer Physikunterricht liegt diesbezüglich zwischen dem deutschen und dem finnischen (CH 42min). Bezogen auf die einzelnen Fälle ergeben sich jedoch grosse Unterschiede, wie Abbildung 9.2 zeigt. Die Spannweite reicht dabei vom vollständigen Verzicht auf Experimente (3 Vorkommnisse FIN von insges. 99 Klassen) bis hin zu Fällen,

³Die Phasen *Vorbereitung*, *Durchführung* und *Nachbereitung* werden durch das Kategoriensystem, siehe Anhang A festgelegt.

bei denen der Unterricht vorwiegend aus experimentellem Handeln besteht (zwischen drittem und viertem Quartil 10 Vorkommnisse in Deutschland). Bezüglich ihrer zentralen Tendenz unterscheiden sich die Länderstichproben alle signifikant.

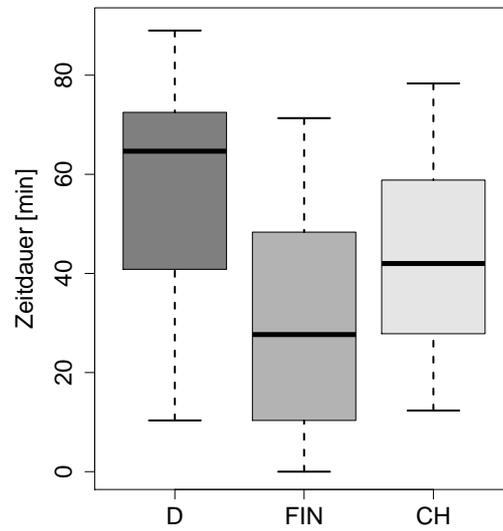


Abbildung 9.2: Die Zeitdauer pro Doppelstunde, die im deutschen, finnischen und Schweizer Physikunterricht im Zusammenhang mit experimentellem Handeln steht (entspricht Abbildung 6.1).

9.1.2 Phasen experimentellen Handelns

Die Phasen *Vorbereitung*, *Durchführung* und *Nachbereitung* machen einen vergleichbaren Anteil an der erwähnten Zeit im Kontext experimentellen Handelns aus. Jedoch unterscheidet sich die Aufteilung der Phasen in den Ländern: Im deutschen Unterricht sind sie mit gleichen Anteilen vertreten, dagegen überwiegt im Schweizer Unterricht die Durchführung (46%⁴) gegenüber der Nachbereitung (30%) und der Vorbereitung (20%) (vgl. Abbildung 6.2). Auch im finnischen Unterricht überwiegt der Anteil der Durchführung gegenüber den übrigen Phasen, allerdings sind die Verhältnisse nicht so deutlich wie in der Schweiz.

Vor- und Nachbereitungsphasen sind dabei hauptsächlich von Fachinhalten bestimmt (vgl. Tabelle 6.4): 85% der Intervalle zu diesen Phasen konnten einer fachinhaltlichen Beschäftigung zugeordnet werden. In 16% der Intervalle wurde eine ausschliesslich organisatorische Beschäftigung festgestellt.

Wie bei der Vor- und Nachbereitungsphase wurde auch die Durchführungsphase auf Spezifika der Phase hin untersucht (vgl. Kategorien K1, K2a, K2b, K3, K5 in Abbildung 9.1). Hierbei zeigt sich folgendes Bild: Es werden sowohl qualitative als auch quantitative Experimente durchgeführt (Abbildung 6.3). In Deutschland überwiegen jedoch quantitative Experimente, sowohl zeitlich (hoch signifikant, vgl. Tabelle 6.5) als auch bezüglich ihrer Häufigkeit (Tabelle 6.6). Auch im Schweizer Unterricht dominieren quantitative Experimente, wobei der Unterschied nicht signifikant ausfällt. Zudem sind Doppelstunden mit qualitativen bzw. quantitativen Experimenten dort ähnlich häufig vertreten. Tendenziell andere Verhältnisse ergeben sich für Finnland: Dort überwiegen qualitative Experimente, sowohl zeitlich als auch bezüglich ihrer Häufigkeit. Jedoch sind auch hier die Unterschiede nicht signifikant.

Die Durchführung der Experimente findet wesentlich in der Gestalt zweier Organisationsformen statt (Abbildung 6.4): Dem Lehrerexperiment und dem Schülerexperiment in Gruppen, bestehend aus jeweils drei oder mehr Lernenden. Andere Organisationsformen machen zusammen genommen nicht mehr als 17% der Doppelstunden aus, bei denen Experimente stattfinden (Tabelle 6.7): Partnerarbeit (9%), Demonstration Schülerin/Schüler (5%), Schülerin/Schüler einzeln (3%). Innerhalb der jeweiligen Subkategorien lassen sich hier keine signifikanten Unterschiede nachweisen. Bezogen auf die gesamte Stichprobe dauert die Durchführung von Schülerex-

⁴Angabe der Mediane: Die Summe der Prozentwerte ergänzt sich nicht notwendigerweise zu 100%.

perimenten in Gruppen im Mittel länger als die von Lehrerexperimenten (Mediane: 9min, 2min). Allerdings treten beide Organisationsformen ähnlich häufig auf, wobei die Aussage auch bezogen auf die einzelnen Länder gilt. Schülerexperimente sind meist arbeitsgleich organisiert: In ausschliesslich 12 der insgesamt 70 Doppelstunden mit Schülerexperimenten kommt eine arbeitsteilige Organisation vor. Das bedeutet, dass im Rahmen von Schülerexperimenten jede Gruppe bzw. jeder Schüler vorwiegend die gleiche Aufgabenstellung bearbeitet.

Für die Durchführung von Experimenten werden überwiegend physikalische Geräte⁵ eingesetzt (Abbildung 6.5). Der Einsatz physikalischer Geräte ist häufiger und dauert im Mittel auch signifikant länger als der Einsatz von Alltagsgeräten (Tabellen 6.9 und 6.8). Letztere beiden Aussagen gelten für alle Länderstichproben. Bildschirmexperimente bzw. Simulationsexperimente sind lediglich in drei Doppelstunden zu beobachten. Unterschiede zwischen den Ländern konnten in der Wahl der Geräte und Materialien nicht festgestellt werden. Die Ergebnisse zum Einsatz von Alltagsgeräten und physikalischen Geräten müssen aufgrund der teilweise mangelhaften Übereinstimmung der beiden Kodierenden mit der nötigen Vorsicht interpretiert werden (vgl. Abschnitt 5.1.2).

9.1.3 Experimentiereinheiten

Mehr als die Hälfte aller deutschen und finnischen Doppelstunden enthalten höchstens zwei Experimentiereinheiten (vgl. Tabelle 6.10). Dabei sind Doppelstunden mit einer Experimentiereinheit in diesen Ländern am häufigsten. Die Schweiz weicht bezüglich der pro Doppelstunde durchgeführten Experimentiereinheiten von den übrigen Ländern ab: Mehr als die Hälfte aller Schweizer Doppelstunden beinhalten wenigstens drei Experimentiereinheiten. Dabei treten Doppelstunden mit drei Einheiten am häufigsten auf. Im Mittel dauert eine Experimentiereinheit in Deutschland 35min, in der Schweiz 18min und in Finnland 15min (Abbildung 6.6). Wie zu erwarten, ergeben sich grosse Unterschiede zwischen verschiedenen Organisationformen: Eine als Gruppenarbeit organisierte Experimentiereinheit dauert zwischen drei (CH) und fünfmal (D) länger als ein Lehrerexperiment⁶.

⁵Dazu zählen Geräte und Materialien, die hauptsächlich für die Verwendung im Physikunterricht bzw. im Labor hergestellt wurden, wie Stativmaterial, Volt- und Amperemeter, Netzgeräte, u.ä.

⁶Finnland liegt diesbezüglich zwischen Deutschland und der Schweiz.

S3 Inwiefern sind zwischen den verschiedenen Schultypen Unterschiede bei den experimentellen Handlungsmustern auszumachen?

9.1.4 Schultypen

Zwischen den Schultypen konnten bezüglich der Sichtstrukturmerkmale experimentellen Handelns keine spezifischen Unterschiede nachgewiesen werden.

9.2 Diskussion Sichtstruktur

Experimentelles Handeln als konstitutives Element des Physikunterrichts.

Die Ergebnisse der Sichtstrukturanalyse zeigen die enorme Bedeutung experimentellen Handelns für den Physikunterricht. Insbesondere in Deutschland und der Schweiz nimmt das Experiment eine den Unterricht dominierende Stellung ein. In Übereinstimmung mit den aktuellen Ergebnissen erhält Tesch (2005) aus der Projektphase 1 den gleichen Anteil experimentellen Handelns für den deutschen Unterricht. Im Sinne einer externen Validierung bestärkt die Übereinstimmung die vorliegenden Ergebnisse. Der zahlenmässig gleiche Anteil resultiert auch als Ergebnis aus Projektphase 2 unter Berücksichtigung einer grösseren Stichprobe in Deutschland und unter Hinzunahme von Fällen aus der Schweiz (Tesch & Gerber, 2005, Dalehefte et al., 2009). Dabei wird, den aktuellen Befunden entgegen, nicht von signifikanten Unterschieden zwischen Deutschland und der Schweiz berichtet. Die Differenzen zwischen den aktuellen Ergebnissen und jenen von Tesch & Gerber (2005), Dalehefte et al. (2009) lassen vermuten: 1) Die Ergebnisse der Sichtstruktur experimentellen Handelns sind abhängig von den unterrichtlichen Inhalten⁷. 2) Länderspezifische Unterschiede treten abhängig von den Inhalten verschieden ausgeprägt auf.

Jonas-Ahrends (2004) Aussage, dass Experimente den Lehrpersonen helfen, ihren Unterricht zu strukturieren, scheint in Anbetracht der erwähnten Anteile, die diese Aktivität ausmachen, naheliegend. Nicht selten besteht der Unterricht aus einer Aneinanderreihung mehrerer Experimentiereinheiten. Inwiefern die Einheiten inhaltlich miteinander verbunden werden lässt sich jedoch auf Basis der vorliegenden Daten nicht feststellen.

⁷Im Unterschied zur aktuellen Studie wurden bei Tesch & Gerber (2005), Dalehefte et al. (2009) Doppelstunden zum Thema ‘Einführung in den Kraftbegriff’ sowie ‘Einführung in die Optik’ verwendet.

Das Ergebnis, dass im Schweizer Unterricht im Vergleich zum deutschen mehr Experimentiereinheiten pro Doppelstunde durchgeführt werden, ist interessant, kann hier aber nicht abschliessend erklärt werden. Einerseits könnte die unterschiedliche Häufigkeit der Experimentiereinheiten ein länderspezifisches Merkmal sein, das die verschiedenen Unterrichtskulturen widerspiegelt. Andererseits ist das Verständnis von dem, was eine Experimentiereinheit ausmacht, multikriterial festgelegt (vgl. Anhang, Abschnitt A.1). Eine Abgrenzung zwischen zwei Einheiten kann damit in Einzelfällen einer hoch-inferenten Beurteilung entsprechen – mit den naheliegenden Konzessionen an die Zuverlässigkeit der Daten. Somit wäre es denkbar, dass die Abgrenzbarkeit von Experimentiereinheiten das Ergebnis beeinflusst. Bezogen auf den deutschen Unterricht bedeutet das: Ein ausgesprochen kleinschrittiges Vorgehen beim experimentellen Handeln macht es schwieriger, einzelne Einheiten experimentellen Handelns voneinander abzugrenzen.

Der vergleichsweise deutlich geringere Anteil experimentellen Handelns an der gesamten Unterrichtszeit in Finnland ist bemerkenswert und kann nicht ausschliesslich der spezifischen Themenwahl der Doppelstunde zugeschrieben werden. Ergebnisse der allgemeinen Sichtstruktur lassen vermuten, dass finnische Lehrende auf Kosten der Experimentierzeit alternative Phasen⁸ des Unterrichts einplanen (Börlin et al., 2010), die im deutschsprachigen Unterricht weniger grosse Bedeutung haben.

Fachsystemantik anstelle von Alltagskontexten. Fast immer werden Experimentiereinheiten dazu eingesetzt, Fachinhalte zu illustrieren, einzuführen oder zu vertiefen. Der Einbezug von Alltagskontexten oder die Orientierung an aktuellen Problemstellungen haben keine zentrale Bedeutung. Sie werden allenfalls als Motivationselemente eingesetzt, wobei die Lehrenden dazu neigen, schnell zu den Fachinhalten überzugehen. Die Feststellung muss durch die nachfolgende Tiefenstrukturanalyse weiter untermauert werden. Gestützt wird sie bislang durch den ausgeprägten Einsatz von physikalischen Geräten anstelle von Alltagsgeräten. Offenbar wird die Zuverlässigkeit der Erzeugung eines Phänomens oder der Generierung von Daten, welche nur mit dem dafür vorgesehenen Gerät erreicht werden kann, gegenüber einer improvisierten Anordnung bevorzugt: Experimente sollen zu eindeutigen Er-

⁸Phasen, die in Finnland gegenüber den übrigen beiden Ländern ausgeprägter auftreten: Wiederholung der Inhalte der vorherigen Stunde zu Beginn einer neuen Stunde, Zusammenfassung am Ende der Stunde und der bedeutend häufigere Einsatz von Prüfungen, Leistungs- und/oder Hausaufgabenkontrollen.

gebnissen führen, so der damit verbundene Wunsch. Der Befund gilt für alle Länder, insbesondere auch für Finnland.

Quantitative Experimente überwiegen qualitative. Die Bevorzugung quantitativer Experimente im deutschsprachigen Unterricht steht im Kontrast zu den Befunden von Tesch (2005) (Projektphase 1). Dort wird berichtet, dass über 90% der Durchführungszeit qualitativ und nicht quantitativ experimentiert wird. Es kann angenommen werden, dass die abweichenden unterrichtlichen Inhalte der beiden Studien für die erwähnte Diskrepanz verantwortlich sind. Zudem handelt es sich bei der aktuellen Studie nicht um Einführungsdoppelstunden wie das bei Tesch (2005), Tesch & Gerber (2005) der Fall war. Ausschlaggebend für das Ergebnis könnte auch der Umstand sein, dass das Thema der Doppelstunde – ‘Der Zusammenhang zwischen elektrischer Energie und Leistung’ – auf einen formalen Zusammenhang verweist. Vor dem Hintergrund ist es eher erstaunlich, dass im finnischen Unterricht den qualitativen Experimenten eine tendenziell grössere Bedeutung beigemessen wird. Es kann vermutet werden, dass dazu die starke Bindung der Lehrenden an das Lehrmittel beigetragen hat: Offenbar verlassen sich Lehrpersonen in Finnland bei der Planung vorwiegend auf eines der beiden finnischen Physiklehrmittel – beide bieten zum entsprechenden Thema ein qualitatives Experiment an. Aufgrund der weitaus grösseren Auswahl an deutschsprachigen Lehrmitteln ist in diesen Ländern eine solche Bindung weniger auffällig bzw. weniger vorhanden. Inwiefern sich die finnischen Lehrmittel insgesamt im Angebot von qualitativen und quantitativen Experimenten gegenüber den deutschsprachigen unterscheiden bleibt hier unbeantwortet.

Das Experiment ist mehr als seine Durchführung. Neben der Durchführung machen Vor- und Nachbereitung wesentliche Anteile der Zeit experimentellen Handelns aus. Das ist erfreulich, denn eine inhaltliche und methodische Hinführung auf das Experiment, die Zusammenführung von Messergebnissen und Erfahrungen oder Erwartungen zu einer Interpretation sind Grundelemente experimentellen Handelns (Harlen & Wake, 1999, Watson, 2000). So ist die Einplanung einer adäquaten Vor- und Nachbereitung eine notwendige Voraussetzung für gelingendes experimentelles Handeln. Ob die Phasen tatsächlich dahingehend genutzt werden, lässt sich vorläufig jedoch nicht feststellen.

Im Vergleich zur IPN-Videostudie (Dalehefte et al., 2009, Tesch & Gerber, 2005) zeigen sich bei den Ergebnissen der deutschen und Schweizer Stichprobe Gemeinsamkeiten: Der Anteil der Durchführungszeit am experimentellen Handeln ist in der Schweiz auf Kosten einer kürzeren Vor- und Nachbereitung grösser als in Deutschland. Zu unterschiedlichen Ergebnissen kommen die Studien bei den Anteilen der Phasen *Vorbereitung*, *Durchführung* und *Nachbereitung* am experimentellen Handeln. In der aktuellen Studie macht die *Durchführung* den grössten Anteil am experimentellen Handeln aus, wogegen bei der IPN-Videostudie die Nachbereitung am grössten ist (Tesch & Gerber, 2005). Es könnte sein, dass der, im Vergleich zur IPN-Videostudie⁹, häufigere Einsatz von Schülerexperimenten zur Vergrösserung der Durchführungsphase geführt hat.

Zwei Organisationsformen experimentellen Handelns. Es dominieren zwei Organisationsformen experimentellen Handelns (vgl. Abschnitt 9.1): das Lehrerexperiment und das Schülerexperiment mit wenigstens drei Lernenden pro Gruppe. Während Lehrerexperimente meist in ein fragend-entwickelndes Unterrichtsgespräch eingebettet sind, erfolgt die Leitung der Schülerexperimente vorwiegend auf der Grundlage von schriftlichen Anleitungen. Dabei erweist sich gerade das Schülerexperiment als komplexes Lernangebot, das durch verschiedene Problemfelder gekennzeichnet ist (vgl. Abschnitt 9.3): Schülerinnen und Schüler haben oft Mühe, die schriftlichen Anleitungen zu verstehen. Die meist vorhandene Unruhe in der Klasse durch Gespräche und Aufbauarbeiten erschwert die Konzentration. Nicht selten werden Lernende wiederholt von der Lehrperson durch Mitteilungen unterbrochen. Weiter ergeben sich Herausforderungen im sozial-kommunikativen Bereich: Fast immer sind die Rollen und Aufgaben der Gruppenmitglieder nicht klar definiert. Experimente, die in Gruppen durchgeführt werden, könnten in praktisch allen Fällen auch alleine bearbeitet werden. Unausgesprochen bleibt auch, dass sich Gruppen gegenseitig unterstützen könnten. Das zeigt: Gerade Schülerexperimente werden oft nicht adäquat eingesetzt.

Das Experiment, ein Mittel zum Zweck? In Anbetracht der teilweise sehr langen Schülerexperimente in Deutschland und der Schweiz stellt sich die Frage nach

⁹Tesch & Gerber (2005) berichten, dass Lehrende aus Projektphase II eher traditionell unterrichteten und fügen an: "That means a lot of teacher talk and demonstrations". Jedoch wird die Aussage nicht quantifiziert.

der Angemessenheit der gewählten Organisationsform in Bezug auf das Unterrichtsziel oder wie Harlen & Wake (1999) es ausdrücken: “Practical work should be seen as means to various ends and not as an end itself.”

9.3 Zusammenfassung zentraler Ergebnisse der Tiefenstruktur experimentellen Handelns

In dem Abschnitt werden Ergebnisse aus den Abschnitten 7.1 und 7.2 zusammengefasst und anschliessend diskutiert.

Bei einer Teilstichprobe, bestehend aus 45 Doppelstunden¹⁰, wurden sämtliche Experimentiereinheiten auf ihre Qualitätsmerkmale hin untersucht. Qualitätsmerkmale aus den nachfolgend aufgeführten drei Perspektiven P1-P3 bildeten die Grundlage für die Beurteilung (vgl. Abschnitte 3.6, 5.2 und 5.3).

Experimentelles Handeln als

P1 **kontextorientierte,**

P2 **reflexive,**

P3 **und theoriegeleitete**

Aktivität.

In Abbildung 9.3 ist die Struktur des Ratinginstruments, bestehend aus den Perspektiven, den untergeordneten Facetten sowie einem allgemeinen Teil illustriert. Letzterer besteht aus den beiden Kontrollvariablen *Funktion* und *Offenheit*. Das Instrument fokussiert auf die Angebotsseite, d.h. auf die Aktionen der Lehrperson. Von den Facetten als auch von den Perspektiven selbst wird angenommen, dass sie für die Qualität der Experimentiereinheiten relevant sind (vgl. Kapitel 3). Die Annahme stützt sich auf theoretische und fachdidaktische Überlegungen. Eine empirische Bestätigung oder Widerlegung dazu kann im Rahmen dieser Studie nicht geleistet werden.

¹⁰Geschichtete Stichprobe mit 15 Doppelstunden pro Land, vgl. Abschnitt 5.2.1.

Die Beurteilung wurde durchgängig von zwei Experten¹¹ nach der Konsensmethode vorgenommen. Sie umfasste die unabhängige Bewertung durch die Experten sowie nachfolgend einen auf der Basis des Ratinginstruments argumentativ ausgetragenen Konsensfindungsprozess. Die unabhängigen Bewertungen zeigen mehrheitlich eine zufriedenstellende Übereinstimmung ($ICC3 > .6$, Kategorien *Fachimmanente Einbettung*, *Lebensweltliche Einbettung*, *Ergebnisreflexion*, *Kognitiver Anspruch* und *Theoretische Fundierung*, vgl. Tabelle 5.6). Diesbezüglich problematischer sind die Facetten *Zielklarheit*, *Begriffsbildung* ($ICC3 > .5$) und *Prozessreflexion* ($ICC3 = .46$). Dabei muss beachtet werden, dass die Koeffizienten aus den unabhängigen Bewertungen der Experten errechnet wurden. Die für die Auswertung relevanten Bewertungen ergeben sich jedoch aus dem anschließenden Konsensfindungsprozess.

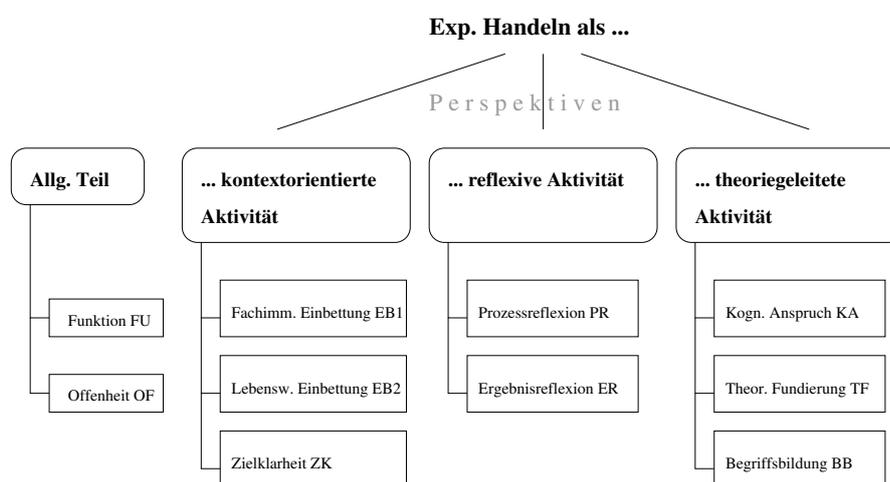


Abbildung 9.3: Die Struktur des Ratinginstruments, bestehend aus den drei Perspektiven experimentellen Handelns und einem allgemeinen Teil (abgerundete Rechtecke). Den Subkategorien bzw. Facetten (Rechtecke) entsprechen Items, die auf einer 4-stufigen Likert-Skala beurteilt werden (entspricht Abbildung 7.1).

Aus dem Verfahren resultierten pro Doppelstunde für jede Experimentiereinheit eine Bewertung, bestehend aus den Ratings aller Facetten und den beiden Kontrollvariablen, sowie eine summarische Bewertung aller Experimentiereinheiten zusammen (Gesamtrating). Die Ergebnisse werden im Folgenden zusammen gefasst.

T1 Findet sich die Qualität experimentellen Handelns, aufgespannt durch die drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung*, in der untersuchten Teilstichprobe wieder?

¹¹Eine Hilfsassistentin sowie der Entwickler des Instruments (J.B.).

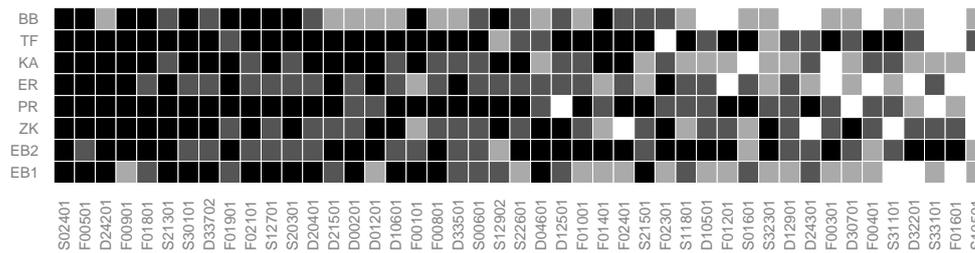


Abbildung 9.4: Überblick über die Ergebnisse der Tiefenstruktur: Zu den acht Facetten der drei Perspektiven *Kontextorientierung*, *Reflexivität* und *Theorieleitung* sind die Ergebnisse für jede Klasse zusammengefasst. *Weiss = trifft zu*, *hellgrau = trifft grösstenteils zu*, *dunkelgrau = trifft teilweise zu*, *schwarz = trifft nicht zu*. Die Klassen sind entsprechend ihres Rating-Summenwerts geordnet (entspricht Abbildung 7.2).

Die mit den drei Perspektiven verbundenen Annahmen über die Qualität der Angebotsseite experimentellen Handelns finden sich nur partiell in der untersuchten Stichprobe. Mit Ausnahme der Facette *Begriffsbildung* wurden in über 50% der Fälle Facetten mit Ausprägungen von *trifft teilweise zu* bis *trifft nicht zu* bewertet.

P1 Experimentelles Handeln als kontextorientierte Aktivität. Die Mediane der Facetten zur Perspektive *Kontextorientierung* betragen (vgl. Tabelle 7.1): *trifft teilweise zu* (*Fachimmanente Einbettung*, *Zielklarheit*) und *trifft nicht zu* (*Lebensweltliche Einbettung*). Insgesamt sind demnach Experimentiereinheiten zu wenig eingebettet. Lehrende greifen zwar relevante Inhalte auf, sowohl vor als auch während der Experimentiereinheit, diese werden aber zu wenig mit den Zielen und Inhalten des Experiments verknüpft. Wenn Zusammenhänge aufgezeigt werden, dann geschieht das meist auf einer eher allgemeinen Ebene. Im Vergleich zur lebensweltlichen Einbettung kommt die fachimmanente Einbettung häufiger und ausgeprägter vor. Weiter sind Experimentiereinheiten nur teilweise zielorientiert. Fragestellungen des Experiments bzw. Hypothesen werden kaum geklärt. Oft unterteilen Lehrpersonen das Experiment in kleine Teilschritte, die sie der Reihe nach durchführen bzw. durchführen lassen (Ablauforientierung). Dabei verlieren Lehrende und Lernende leicht den Überblick über die gesamte Einheit.

P2 Experimentelles Handeln als reflexive Aktivität. Zur Perspektive *Reflexivität* erhält man für die Facetten folgende Mediane (Tabelle 7.1): *trifft teilweise zu*

(*Ergebnisreflexion*), trifft nicht zu (*Prozessreflexion*). Somit gestalten Lehrpersonen Experimentiereinheiten überwiegend nicht reflexiv. Die Reflexion über die Experimentiereinheit als Ganzes gelingt dabei den Lehrpersonen eher als die Initiierung von Reflexionen über Ideen und Vorstellungen während der Aktivität (vgl. Facetten *Prozessreflexion* und *Ergebnisreflexion*, Abschnitt 5.3). Lehrende nehmen im Rahmen der Nachbereitungsphase häufig Bezug auf das durchgeführte Experiment und reflektieren Teile der Durchführung. Mängel, die zu einer ausschliesslich teilweisen Ausprägung der Facette *Ergebnisreflexion* führen, sind: Ungenügender Vergleich zum Ausgangspunkt bzw. zur Fragestellung des Experiments sowie das Fehlen einer “griffigen” Zusammenfassung der Ergebnisse.

Als nicht zutreffend wurde in einer Mehrheit aller Fälle die Initiierung von Reflexionen über Ideen und Vorstellungen (*Prozess Reflexion*) bewertet. Das Frage- und Antwortverhalten der Lehrenden scheint hierbei Reflexionen der Schülerinnen und Schüler nicht zu unterstützen. Fragen, welche die Lehrperson im Rahmen des fragend-entwickelnden Unterrichtsgesprächs äussert, sind vorwiegend geschlossen und geben wenig Anlass zur Reflexion. Darüber hinaus neigen Lehrpersonen dazu, bei einer “falschen” Schülerantwort einen weiteren Schüler aufzurufen. Für die Lernenden besteht somit wenig Gelegenheit ihre Antworten zu reflektieren und gegebenenfalls zu revidieren.

P3 Experimentelles Handeln als theoriegeleitete Aktivität. Die Mediane der Facetten zur Perspektive *Theorieleitung* betragen (Tabelle 7.1): *trifft grösstenteils zu (Begriffsbildung)*, *trifft teilweise zu (Kognitiver Anspruch)* sowie *trifft nicht zu (Theoretische Fundierung)*. Am besten wurden die Experimentiereinheiten bezüglich der Kategorie *Begriffsbildung* bewertet: Lehrende setzen die Sprache grösstenteils adäquat und reflektiert ein. Sie sind sich dabei der unterschiedlichen Bedeutungen von Fach- und Alltagsvokabular bewusst und weisen Lernende wiederholt auf die Unterschiede hin (vgl. Facette *Begriffsbildung*, Abschnitt 5.3). In der Mehrheit der Fälle ermuntern sie die Lernenden zum Gebrauch adäquater Fachtermini und bieten Unterstützung.

Dagegen beobachtet man bei einer Mehrheit der Fälle Experimentiereinheiten, deren kognitiver Anspruch höchstens teilweise zu den Lernenden passt (Facette *Kognitiver Anspruch*). Viele Experimentiereinheiten unterfordern Schülerinnen und Schüler bezüglich der mit dem Experiment verbundenen Ideen und Vorstellungen.

So wurden 25 von insgesamt 68 Experimentiereinheiten der Funktion *Durchführen einer oder mehrerer Messungen* zugeordnet (vgl. Tabelle 7.3). Bei Einheiten dieser Funktion fehlt die Verbindung zu Ideen und Vorstellungen vollständig. Zur Funktion *Ein physikalisches Konzept veranschaulichen*, welche eher die erwähnte Verbindung unterstützen würde, gehören dagegen nur gerade 6 Einheiten, was mit der überwiegend nicht vorhandene *Theoretische Fundierung* zusammen passt (Tabelle 7.1).

Nicht selten, und im Speziellen bei Schülerexperimenten, geht die erwähnte Unterforderung mit einer Überforderung bei der Anwendung von technischen Geräten einher. Die Unterforderung bei den Vorstellungen und Ideen hat damit zu tun, dass diese beim experimentellen Handeln praktisch keine Bedeutung haben. Das zeigt die überwiegend nicht zutreffende Facette *Theoretische Fundierung*: Qualitative, physikalische Überlegungen zu Phänomenen und Vorgängen beim Experimentieren werden selten thematisiert. Meist geht die Absenz qualitativer Vorstellungen mit einer besonderen Betonung quantitativer Zusammenhänge einher (vgl. Ergebnisse zur Kategorie *Funktion*, Ausprägung FU5, Tabelle 7.3 und der Sichtstruktur Kategorie *Datenerfassung*, Abbildung 6.3).

Die Befunde über das Vorhandensein der drei Perspektiven experimentellen Handelns mag einen generell negativen Eindruck erwecken. Es muss jedoch betont werden, dass eine nahnhafte Minderheit von Fällen insgesamt positiv bewertet werden konnte (vgl. Abbildung 9.4, rechtes Drittel). Auf der Basis einer hierarchischen Clusteranalyse konnten neun Experimentiereinheiten von hoher Qualität identifiziert werden (vgl. Abschnitt 5.4 und Anhang, Kapitel C). Ein Teil davon wurde im Rahmen der qualitativen Analyse einzelner Einheiten vorgestellt.

T2 Lassen sich bezüglich der Perspektiven oder Facetten Unterschiede zwischen den Ländern nachweisen?

Signifikante länderspezifische Unterschiede treten ausschliesslich bei der Facette *Begriffsbildung* auf. Sie kommt ausgeprägter im deutschen Unterricht (Median: *trifft grösstenteils zu*) im Vergleich zum Unterricht in der Schweiz und in Finnland (Mediane: *trifft teilweise zu*) vor (Abbildungen 7.3, 7.4 und 7.5).

T3 Wie verhalten sich die Ratings einzelner Experimentiereinheiten innerhalb einer Klasse zu ihrer summativen Beurteilung (Gesamtrating)?

Die Bewertungen (Ratings) einzelner Experimentiereinheiten einer Klasse variieren in den meisten Fällen nur wenig, wie Tabelle 7.2 zeigt. Der Befund unterstützt das Vorgehen, die Ergebnisse der drei Perspektiven auf Klassenebene auszuwerten. Dies könnte als Hinweis interpretiert werden, dass die durch die Perspektiven definierten Qualitätsmerkmale eng mit dem Verhalten der Lehrperson verknüpft sind und möglicherweise weniger mit der inhaltlichen und organisatorischen Ausgestaltung des jeweiligen Experiments zusammenhängen.

T4 Welche Merkmale der Kategorien *Funktion* und *Offenheit* lassen sich in den Experimentiereinheiten identifizieren?

Über 80% aller Experimentiereinheiten erfüllen eine der beiden fachdidaktischen Funktionen: *Durchführen einer oder mehrerer Messungen* (FU3) oder *Eine physikalische Gesetzmässigkeit plausibel machen (quantitativ oder qualitativ)* (FU5). Beide Funktionen treten zu etwa gleichen Teilen auf. Die restlichen knapp 20% der Experimentiereinheiten werden den Funktionen *Das Lösen bzw. Erkennen eines technischen Problems* (FU2), *Das Kennenlernen eines Phänomens* (FU4) sowie *Ein physikalisches Konzept veranschaulichen* (FU6) zugeordnet. Keine Bedeutung haben *Das Kennenlernen eines Laborgerätes* (FU1) und *Das Erlangen eines besseren allgemeineren Verständnisses von den Arbeits- und Funktionsweisen der Naturwissenschaft* (FU7). Es wird festgestellt, dass die Ausprägungen der Facetten von der Funktion abhängig sind: Experimentiereinheiten des Merkmals FU6 weisen häufiger hohe Ausprägungen bei den Facetten *Prozess Reflektion*, *Kognitiver Anspruch* und *Theoretische Fundierung* auf als die der Merkmale FU3 bis FU5.

Die untersuchten Experimentiereinheiten sind mit wenigen Ausnahmen nicht offen gestaltet. Das bedeutet, dass die Einheiten oft auf der Basis eng geführter Handlungsanweisungen ablaufen. Dies trifft für Lehrer- und Schülerexperimente gleichermaßen zu.

9.4 Diskussion Tiefenstruktur

Das Experiment – als zusammenhangslose Aktivität. Die Ergebnisse der Tiefenstrukturanalyse lassen vermuten, dass Einheiten experimentellen Handelns zu wenig mit den Inhalten des Unterrichts verbunden werden (Perspektive *Kontextori-*

entierung). Zwar haben Experimentiereinheiten in der überwiegenden Mehrheit ein inhaltliches Ziel, wie z.B. das Erkennen des Zusammenhangs zwischen elektrischer Energie, Leistung und Stromstärke. Dennoch gelingt es oft nicht, die Einheit in Bezug auf die inhaltliche Bedeutung genügend zu begründen. Es ist zu vermuten, dass gerade Schülerexperimente vorwiegend aus methodischen Überlegungen und weniger aus einer inhaltlichen Notwendigkeit geplant werden. Experimente helfen den Lehrenden, ihren Unterricht zu strukturieren, wie das schon Jonas-Ahrend (2004) feststellt. Dass Vor- und Nachbereitungszeiten alleine als Indikatoren für Einbettung aufgefasst werden können, wie das bei Tesch (2005) vermutet wird, muss aufgrund der Ergebnisse aus Sicht- und Tiefenstruktur bezweifelt werden. Für die mangelnde *Kontextorientierung* lassen sich verschiedene Gründe anführen:

Lehrpersonen orientieren sich stark an Inhalten auf der Ebene von Observablen und Gesetzen (vgl. Ergebnisse zur Facette *Theoretische Fundierung*, Abschnitt 7.1). Werden diese nicht durch übergeordnete Ideen und Konzepte in einen Zusammenhang gestellt, wirken sie als Ansammlung unzähliger Fakten und Einzelergebnissen. Der Eindruck bleibt auch dann bestehen, wenn beispielsweise vor dem Experiment physikalische Größen repetiert werden, was gelegentlich beobachtet wurde. Bezüge zwischen Inhalten und dem Experiment werden meist sehr unverbindlich und allgemein gehalten.

Lehrpersonen äussern sich kaum zu den pädagogisch-didaktischen und inhaltlichen Zielen der Aktivität. Das beobachten auch Abrahams & Millar (2008), die feststellen, dass oft völlig unklar bleibt, wie Lernen auf der Grundlage der Observablen des Experiments stattfinden soll. Die aktuellen Befunde sowie jene von Abrahams & Millar passen zu denen von Hart et al. (2000), die festhalten, dass Schülerinnen und Schüler die Ziele¹² von Experimenten oft nicht verstehen. Sie fordern: “[...] teachers need to make the pedagogical purpose of laboratory work explicit for students”. Der Forderung kann man mit Blick auf die Resultate der Tiefenstrukturanalyse nur zustimmen. Die Kenntnis der Ziele ist eine Voraussetzung für die Übernahme von Verantwortung für das Lernen (Labudde, 2000).

Während Schülerinnen und Schüler öfters Gelegenheit erhalten, ihr Vorwissen mitzuteilen, gelingt es kaum, das Vorwissen auch einzubeziehen. Dies mag daran liegen, dass Lehrpersonen sich stark am geplanten Ablauf der Einheit orientieren

¹²Hart et al. verwenden den Begriff “purpose” und meinen damit: “[...] the intentions the teacher has for the activity when she/he decides to use it with this class at this time”.

(vgl. nachfolgender Paragraph) und dabei die Sicht der Schülerinnen und Schüler zu wenig integrieren. Der mangelnde Bezug auf die Sicht der Lernenden beim experimentellen Handeln stellen auch Jonas-Ahrend (2004), Abrahams & Millar (2008) fest.

Das Experiment – im Korsett der Ablauforientierung. Wie die Ergebnisse der Perspektive *Reflexivität* zeigen, bieten die analysierten Experimentiereinheiten den Lernenden wenig Gelegenheiten zur Reflexion. Die überwiegende Abwesenheit reflexiver Elemente kann möglicherweise mit der von Fischler (2000, S.79) eingeführten Handlungsorientierung, der *Ablauforientierung*, verstanden werden. Unter der Ablauforientierung versteht er die Absicht, “[...] den Unterricht am Fließen zu halten und die im vorgängig entworfenen Gesamtbild vom Unterrichtsablauf enthaltenen Unterrichtsabschnitte auch wirklich zu realisieren”. Es kann vermutet werden, dass besonders Experimente stark mit dieser Absicht verbunden werden, da sie im Sinne der weit verbreiteten induktivistischen Logik (vgl. folgenden Paragraph) erst zu den entsprechenden Inhalten führen.

Mit der Ablauforientierung geht meist ein kleinschrittiges und eng geführtes Vorgehen der Lehrenden einher. Der Befund wird durch die Ergebnisse der Subkategorie *Offenheit* gestützt. So nimmt die Lehrperson den Lernenden einen wesentlichen Teil des kognitiven Anspruchs im Zusammenhang mit dem Experiment ab (vgl. Ergebnisse zur Facette *Kognitiver Anspruch*). Die Ablauforientierung steht in Konkurrenz mit den Gelegenheiten zur Reflexion: So neigen Lehrende, wie beschrieben wurde, dazu, bei ‘falschen’ Schülerantworten einen weiteren Schüler, eine weitere Schülerin aufzurufen, die sie als kompetent in Bezug auf die Frage einstufen. Das Vorgehen ist zwar aus Sicht der Ablauforientierung durchaus folgerichtig, stellt jedoch in Bezug auf die aktive Konstruktion von Wissen (Labudde, 2000) ein erhebliches Problem dar. Als Folge davon wird es für Lernende schwieriger, am Experiment gedanklich teilzuhaben. Dies wäre jedoch eine der Voraussetzungen dafür, dass Schülerinnen und Schüler ein Experiment motivierend erleben (Hodson, 1993). Die Ablauforientierung findet sich sowohl bei Lehrer- als auch bei Schülerexperimenten. Bei letzterer Aktivität unterstützen meist schriftliche Anleitungen die Ablauforientierung während der Schülerarbeitsphasen¹³.

¹³Hodson (1993), Harlen & Wake (1999, u.a.) sprechen in diesem Zusammenhang von “cook-book activities”.

Der im Verlauf der Einheit zunehmende Druck der auferlegten Ablauforientierung führt viele Lehrende gegen Ende zu einem meist überstürzten Ziehen von Schlussfolgerungen. Fischler (2000) spricht in dem Zusammenhang von einer Abschlussorientierung der Lehrenden. Unter diesen Umständen gelingt es auch hierbei nicht, die Sicht der Schülerinnen und Schüler in genügendem Masse einzubeziehen. Ein weiterer Aspekt, welcher das nur teilweise Vorhandensein der *Ergebnisreflexion* erklärt, betrifft den Umstand, dass Ziele der Experimentiereinheit oft nicht klar geäußert werden (vgl. vorherigen Paragraph): Auf der Grundlage unklarer Ziele fällt eine Sicherung der Ergebnisse umso schwerer. Hier hängen die beiden Facetten *Zielorientierung* und *Ergebnisreflexion* voneinander ab.

Das Experiment – als Quelle von Erkenntnis. Die Auffassung experimentellen Handelns als theoriegeleitete Aktivität zeigt sich nur in einer Minderheit der untersuchten Fälle. Innerhalb der Perspektive ergeben die Facetten jedoch kein einheitliches Bild ab. Die überwiegend positiv beurteilte Facette *Begriffsbildung* ist sehr erfreulich. Sie zeigt, dass Lehrende bezüglich der Sprache durchaus die Sicht ihrer Schülerinnen und Schüler berücksichtigen. Das bessere Abschneiden der deutschen Lehrpersonen gegenüber ihren finnischen und Schweizer Kolleginnen und Kollegen ist interessant und müsste weiter untersucht werden. Es ist denkbar, dass kulturelle Unterschiede das Ergebnis der Facette *Begriffsbildung* beeinflussen.

Die mehrheitlich positive Bewertung der *Begriffsbildung* steht im Kontrast zum Abschneiden bei den Facetten *kognitiver Anspruch* und *theoretische Fundierung*. Lehrpersonen etablieren kaum qualitative Vorstellungen zu den Vorgängen im Experiment. Nur bei einer kleinen Minderheit der Experimente zogen Lehrende die Wasseranalogie des elektrischen Stromkreises für Erklärungen heran. Auffallend war, dass auch diese Lehrpersonen Schülerinnen und Schüler nicht dazu animierten, die Vorstellung (Wasseranalogie) für ihre eigenen Erklärungen anzuwenden. So blieb es meist bei einem Exkurs der Lehrperson, der kaum nachhaltige Wirkung auf die Interpretation des Experiments entfalten konnte.

Der Befund ist insofern bemerkenswert, weil Lehrende in Befragungen zum Experiment die Verbindung von Theorie und Praxis als wichtiges Ziel herausstreichen (Welzel et al., 1998). Hier besteht zwischen Anspruch und Unterrichtsrealität offenbar eine Diskrepanz. So werden Experimente häufig mit Gesetzmässigkeiten verbunden, jedoch fehlt jenseits der formalen Beschreibung eines Phänomens oder Ver-

haltens die Entwicklung einer qualitativen Vorstellung von den physikalischen Vorgängen. Der Terminologie von Millar et al. (1999) folgend sind die experimentellen Handlungen damit nicht ‘theory-driven’ sondern ‘object-driven’. In der Konsequenz sind Experimentiereinheiten bezüglich Vorstellungen und Ideen überwiegend wenig kognitiv anspruchsvoll, wogegen Schülerinnen und Schüler oft mit dem eigentlichen Aufbau und den Geräten überfordert werden. Dies zeigen auch die Erkenntnisse der Subkategorie *Funktion*, wonach die meisten Experimente dazu dienen, entweder Messungen durchzuführen oder eine Gesetzmässigkeit plausibel zu machen (vgl. Tabelle 7.3).

Die Betonung der Objekte (‘object-drivenness’), die damit verbundene Unterminierung der mit den Objekten verbundenen Vorstellungen und Ideen sowie die bereits besprochene mangelnde Zielorientierung deuten darauf hin, dass Lehrende Experimente nach einer stark induktivistischen Logik organisieren. Vorherrschend ist offenbar die Vorstellung, dass sich Erkenntnis praktisch automatisch mit den gewonnenen Messwerten einstellt. Diese Vorstellungen bestätigen auch Ergebnisse von Tesch (2005), Abrahams & Millar (2008).

9.5 Zusammenfassung zentraler Ergebnisse der Fallanalysen

Die vorgestellten und analysierten Experimentiereinheiten erlauben einen Einblick in das verarbeitete Datenmaterial. Sie zeigen bezüglich des angelegten Massstabes qualitativ hochstehende Experimentiereinheiten (vgl. Abbildung 9.4).

T5 Was zeichnet Experimentiereinheiten hoher Qualität aus? Welche Eigenschaften der Lehrperson sind damit verbunden?

Die vorgestellten Fälle umfassen ein breites Spektrum verschiedener Experimentiereinheiten. Es stellt sich daher die Frage, welche Gemeinsamkeiten diese aufweisen.

Bezüge herstellen und Analogien aufzeigen. Als gemeinsames Merkmal der Fälle F-0-16-01 (Lehrperson B), D-1-29-01 (C) und S-3-31-01 (D) gelingt es den Lehrenden, das Experiment mit verschiedenen fachimmanenten oder lebensweltli-

chen Aspekten zueinander in Beziehung zu setzen¹⁴. Hierbei konnten verschiedene Formen der Bezugnahme identifiziert werden. Eine davon ist das Verbinden zweier Systeme unterschiedlicher Komplexität. Beispielsweise setzt C ein idealisiertes Dynamo-Motor-System (Lehrerexperiment) mit der elektrischen Energieversorgung von Haushalten (komplexes System) in Beziehung. Ganz ähnlich verweist D vom Lehrerexperiment, einem Dynamo-Licht-System, auf die Energieversorgung der Leuchtstoffröhren im Zimmer. Eine weitere Variante der Bezugnahme ist der Vergleich verschiedener idealisierter Systeme. So vergleichen B und D einen einfachen elektrischen Stromkreis mit einem Wasserkreislauf, wogegen C die mechanische Transmission von Energie mit der elektrischen vergleicht.

Dadurch gelingt es den Lehrenden, Eigenschaften von Systemen über die Bezugnahme zu anderen Systemen herauszuarbeiten. Das geschieht meist anhand von Lehrerexperimenten. Lernende können so eher an bereits Bekanntem Anknüpfen und neue Aspekte besser einordnen.

Prozesse sichtbar machen. Lehrende machen kognitive Prozesse im Zusammenhang mit dem Experiment und der Gewinnung von Erkenntnis modellhaft sichtbar. Besonders deutlich wird das bei A und D in der Kategorie *Diese Fragen* (vgl. Tabelle 7.13). So analysiert A sehr genau den Arbeitsstand von Schülergruppen und bespricht darauf die nächsten Entwicklungsphasen. Lernende können so ihre Fortschritte bei der Bearbeitung der Entwicklungsaufgabe besser erfassen. Ähnlich, aber im Rahmen eines fragend-entwickelnden Unterrichtsgesprächs, konfrontiert D die Lernenden immer wieder mit der relevanten Problemstellung. Das prozesshafte Vorgehen zeigt sich dabei in der fortwährenden gegenseitigen Anpassung von Lösungsvorschlägen und Problemstellungen.

Auseinandersetzung mit Schülerbeiträgen. Die analysierten Lehrpersonen (insbesondere A und D) zeigen ein besonderes Interesse an Schülerbeiträgen (vgl. Kategorie *Schülerbeiträge verarbeiten*, Tabelle 7.13). Das äussert sich vor allem darin, dass Lehrpersonen bemüht sind die Beiträge richtig zu verstehen, indem sie diese paraphrasieren und allenfalls Rückfragen stellen. So zeigen sie, dass die Beiträge für den Unterricht wertvoll sind, auch dann, wenn diese nicht unmittelbar in Bezug auf die Problemstellung zielführend sind. Gleichzeitig gelingt es den Lehrenden weit-

¹⁴jeweils 20% der analysierten Exzerpte wurden dem lebensweltlichen oder fachimmanenten Kontext zugeordnet, vgl. Tabelle 7.13

gehend, die Balance zwischen dem Eingehen auf Fragen und einem zielgerichteten Ablauf zu wahren.

Orientierung an Vorstellungen. Die Etablierung von Vorstellungen im Zusammenhang mit dem Experiment zeichnet die vier Fälle besonders aus. Beispielhaft sei das Simulationsexperiment in F-0-16-01 erwähnt, anhand dessen B Analogien zwischen dem elektrischen Stromkreis und dem Wasserkreislauf aufzeigt. Im Rahmen eines Lehrervortrags gelingt es B so, eine tragfähige Vorstellung der Grössen Stromstärke, Spannung und Widerstand anzubieten, die dann auch im weiteren Verlauf des Experiments verwendet werden. Hier bietet das Simulationsexperiment eine theoretische Anschauung, die aus einem realen Experiment nicht ersichtlich wäre.

Die Fokussierung auf qualitative physikalische Vorstellungen oder Ideen geht aus der ausgeprägt vorhandenen *Ideenebene* und der dagegen weniger ausgeprägten *Objektebene* hervor (vgl. Tabelle 7.13).

T6 Inwiefern decken sich die Befunde der Fallanalysen mit den Ergebnissen aus dem Ratingverfahren?

Die Befunde aus der qualitativen Analyse ergeben im Vergleich zu den Ratings ein überwiegend stimmiges Bild. Die Triangulation der Fälle zeigt somit, dass zwei verschiedene Vorgehensweisen zu mehrheitlich übereinstimmenden Befunden führt, was als Hinweis für die Validität des Ratinginstruments gewertet werden kann. Für eine umfassende Validierung müssten jedoch mehr Fälle, insbesondere auch Fälle mit schwachen Qualitätsausprägungen untersucht werden.

Bei den Facetten *Lebensweltliche Einbettung* EB2 und *Prozessreflexion* PR ergab die qualitative Analyse in mehr als einem Fall ein abweichendes Ergebnis. Eine abschliessende Erklärung für die Abweichungen, kann nicht gegeben werden. Es kann vermutet werden, dass die grösstenteils zutreffende *Lebensweltliche Einbettung* in S-1-05-01 und S-3-31-01 deshalb nicht als besonders charakterisierendes Merkmal bei der Transkriptanalyse aufgefallen ist, weil andere Merkmale deutlicher hervortreten (S-1-05-01: *Prozessreflexion*, trifft zu, *Ergebnisreflexion* trifft zu. S-3-31-01: *Theoretische Fundierung*, trifft zu; *Zielklarheit*, trifft nicht zu). Die Beurteilung aufgrund vorgegebener Kategorien scheint sich hier von der Transkriptanalyse, bei der die Kategorien erst aus den Daten hervorgehen, abzuheben.

Vorherige Argumentation lässt sich mit Einschränkungen auch für die Facette *Prozessreflexion* anführen (F-0-16-01). Ausserdem konnte bei der erwähnten Facette nur eine mangelhafte Übereinstimmung zwischen den Ratern nachgewiesen werden (ICC=.46, vgl. Tabelle 5.6), was auf die Schwierigkeit bei der Bewertung der Facette hindeutet. Die Ergebnisse der Facette *Prozessreflexion*, bei der somit Entwicklungsbedarf besteht, sind daher als hypothesengenerierend anzusehen (vgl. Abschnitt 9.8.3).

T7 Welche neuen oder verfeinerten Kategorien ergeben sich aus den Fallanalysen?

Aus der qualitativen Analyse der Transkripte ergeben sich verschiedene Ideen für wichtige Kategorien im Zusammenhang mit verfeinerten Kategoriensystemen (vgl. Tabellen 7.10-7.12). Grundsätzlich unterscheiden sich sämtliche qualitativen Kategorien von den Facetten des Ratinginstruments. Im Folgenden werden drei Erweiterungen des Ratingsystems vorgeschlagen:

1. Im Zusammenhang mit Schülerexperimenten bereitet der experimentelle Aufbau oft grosse Schwierigkeiten (*Objektebene*). Interessant wäre eine genauere Erfassung, inwiefern Lernende bei der Durchführung von Experimenten durch die erwähnten Schwierigkeiten das Ziel der Einheit verfehlen. Eine Gegenüberstellung der Anzahl Schülerfragen zum Aufbau und der Anzahl Schülerfragen zum konzeptionellen Verständnis des Experiments wäre hier ein erster Anhaltspunkt.
2. Wiederum bei Schülerexperimenten stellt sich die Frage, wie Schülerinnen und Schüler innerhalb der bzw. zwischen den Gruppen kommunizieren und welche Orientierung die Lehrperson diesbezüglich bietet (*Sozial-kommunikative Ebene. Moderieren*).
3. Eine differenziertere Erfassung der Lehrer-Schüler-Kommunikation liesse sich durch die Kategorie *Schülerbeiträge verarbeiten* erreichen. Anhand der Subkategorien könnte die Bedeutung der bestehenden Facette *Prozessreflexion* noch genauer erfasst werden.

9.6 Diskussion Fallanalysen

Die Fallanalysen auf der Grundlage von Lehrertranskripten geben Einblicke in Experimentiereinheiten hoher Qualität. So können die Beurteilungen auf der Basis des Ratinginstruments anhand der Fälle besser nachvollzogen werden. Es ist erfreulich, dass die qualitative Fallanalyse zusammen mit den Ratings ein überwiegend stimmiges Bild ergibt. Einzelne Unstimmigkeiten können dazu beitragen, das Ratinginstrument für zukünftige Anwendungen zu verbessern.

Zwischen dem Verhalten der vier analysierten Lehrpersonen gibt es verschiedene Gemeinsamkeiten (vgl. Abschnitt 9.5): 1. Das Herstellen von Bezügen und Analogien, 2. Das Sichtbarmachen von Prozessen, 3. Die Auseinandersetzung mit Schülerbeiträgen sowie 4. Die Orientierung an Vorstellungen. Es ist demnach ein implizites oder explizites Bewusstsein über gelingende Lehr-Lernprozesse, die zusammen mit einer an physikalisch relevanten Konzepten orientierten Gestaltung der Experimentiereinheit, die hohe Qualität ausmachen.

Auffallend ist zudem, dass ausnahmslos alle Aktivitäten aus wissenschaftlicher Sicht nicht als experimentelle Handlungen im engeren Sinne bezeichnet würden. Während es bei S-1-05-01 um eine Entwicklungsaufgabe geht, dienen die übrigen Experimentiereinheiten dazu, ein physikalisches Konzept zu veranschaulichen.

9.7 Diskussion zentraler Ergebnisse zum Zusammenhang zwischen experimentellem Handeln und Fachwissensleistung

Im vorliegenden Abschnitt werden Ergebnisse zum Zusammenhang zwischen experimentellem Handeln und dem Zuwachs an Fachwissensleistung zusammengefasst und diskutiert.

Für die Erhebung des Zuwachses an Fachwissensleistung absolvierten Schülerinnen und Schüler im Rahmen des Pre- und des Posttests einen Test¹⁵. Die über die Klassen gemittelten Testergebnisse wurden sodann als unabhängige Variable in die

¹⁵Der *Content Knowledge Paper and Pencil Test* (CKPT) ist in Geller et al. (in Vorb.) dokumentiert. Er umfasst Items unterschiedlicher Schwierigkeitsgrade zum Thema "Zusammenhang zwischen elektrischer Energie und Leistung".

Analyse einbezogen. Als davon abhängig betrachtete man die Variablen *Experimentierzeit* (Sichtstruktur) und die Facetten der Tiefenstruktur.

Es handelt sich hierbei um eine *erste*, nicht abschliessende Analyse zum Zusammenhang zwischen Variablen experimentellen Handelns im Unterricht und Outputvariablen (vgl. Abschnitt 9.9.2).

Z1 Lässt sich zwischen der im Kontext experimentellen Handelns stehenden Unterrichtszeit und dem Leistungszuwachs ein Zusammenhang nachweisen?

Aufgrund der Ergebnisse aus Kapitel 8 kann kein Zusammenhang zwischen der im Kontext experimentellen Handelns stehenden Zeit und dem Leistungszuwachs nachgewiesen werden, wie er z.B. bei Tesch (2005, S.105) vermutet wird.

Es bestätigt sich auch hier der allgemeine Befund von Hugener et al. (2006, zitiert in Pauli & Reusser (2006)), wonach niedrig-inferente Kodierungen der Sichtstruktur zwar Aufschluss über methodischen Gestaltungsformen des Unterrichts geben, jedoch kaum Rückschlüsse auf die Leistungsentwicklung der Schülerinnen und Schüler zulassen.

Z2 Lassen sich zwischen Qualitätsmerkmalen experimentellen Handelns und dem Leistungszuwachs Zusammenhänge nachweisen?

Die Kovarianzanalyse¹⁶ ergibt, dass sich zwischen den Facetten *Fachimmanente Einbettung* und *Ergebnisreflexion* und dem Zuwachs an Fachwissensleistung ein signifikanter Zusammenhang mittlerer Effektstärke nachweisen lässt. Knapp nicht signifikant ist der Zusammenhang zwischen der Facette *Kognitiver Anspruch* und der Outputvariablen. Für die verbleibenden Facetten lässt sich ein solcher Zusammenhang nicht nachweisen.

Zusammenfassend lässt sich damit feststellen: Fachinhaltlich eingebettete, anspruchsvolle Experimente, deren Ergebnisse sorgfältig reflektiert werden, haben günstige Auswirkungen auf die mittlere Leistungsentwicklung einer Klasse.

¹⁶Abhängige Variable: Klassenmittel des Leistungszuwachses, unabhängige Variablen: Facetten *Fachimmanente Einbettung*, *Lebensweltliche Einbettung*, *Zielklarheit*, *Prozessreflexion*, *Ergebnisreflexion*, *Kognitiver Anspruch*, *Theoretische Fundierung*, *Begriffsbildung*, Kovariate: Land.

9.8 Methodenreflexion

9.8.1 Gesamtstudie

Dieser Teil der Methodenreflexion bezieht sich auf die Gesamtstudie ‘Quality of Instruction in Physics’. Die Studie basiert auf dem systemischen Modell von Lipowsky et al. (2005), wonach das Unterrichtsgeschehen durch eine Vielzahl von Variablen¹⁷ beeinflusst wird (vgl. Kapitel 4). Zusammen mit den Variablen des Unterrichts, z.B. den Perspektiven experimentellen Handelns, sollen Zusammenhänge zwischen dem Unterricht und Outputvariablen, wie Motivation und Interesse oder Fachkompetenz der Schülerinnen und Schüler, abgeleitet werden können. Eine solche Analyse umfasst sowohl die Klassen- als auch die Schülerebene. Während bei letzterer der Stichprobenumfang als genügend eingestuft werden kann, ist er auf der Ebene der Klassen deutlich zu tief. Zu einem identischen Befund kommt bereits Knierim (2008, S.191f) mit Bezug zur Vorgängerstudie ‘Lehr-Lern-Kultur’. Dabei errechnet sie eine Stichprobengröße¹⁸ von 250 Klassen. Mit dem Einbezug aller Schultypen und den drei Länderstichproben stellt die aktuelle Studie eher noch grössere Anforderungen an die Grösse der Stichprobe. Die vorhandene Stichprobengrösse von insgesamt 99 Klassen (43 Deutschland, 25 Finnland und 31 Schweiz, vgl. Tabelle 4.1) ist damit für die Implementierung des Modells von Lipowsky et al. deutlich zu klein bemessen.

9.8.2 Sichtstruktur

Das auf Tesch (2005) basierende Kategoriensystem zur Erfassung der Einheiten experimentellen Handelns und deren Phasen hat sich bewährt. Das zeigt sich auch darin, dass die Ergebnisse überwiegend zu denen im Rahmen der Vorgängerstudie gefundenen Befunde passen, was für die Validität des Systems spricht. Dennoch dürfen die Ergebnisse der Beobachterübereinstimmung (vgl. Tabellen 5.2 und 5.3) nicht überschätzt werden: Bei der zeitbasierten Kodierung von Videos ergeben sich methodische Probleme, die bisher in der Literatur kaum Beachtung gefunden haben (Seidel et al., 2005, Pauli & Reusser, 2006, Stigler et al., 2000). Sämtliche Berech-

¹⁷Professionswissen, Einstellungen, Motivation und Arbeitsbedingungen (Lehrperson) bzw. familiärer Hintergrund, kognitive Fähigkeiten, Kompetenz und Motivation (Schülerinnen und Schüler); vgl. Abbildung 4.1.

¹⁸Gemeint ist die Stichprobengrösse pro Land.

nungen der Beobachterübereinstimmung¹⁹ setzen voraus, dass aufeinanderfolgende Zeitintervalle *unabhängig* voneinander beurteilt werden können. Das ist jedoch nicht möglich, da Kodierende über Wissen der bereits bearbeitete Intervalle verfügen. Hat sich beispielsweise eine Kodierende dafür entschieden, dass ein Intervall neu nicht mehr der *Vorbereitung* zugeordnet wird, sondern der *Durchführung in Gruppen*, so wird die Kodierung nachfolgender Intervalle stark von dieser *einen* Entscheidung beeinflusst. In der Folge wird der Grad der Übereinstimmung, ausgedrückt durch eine der erwähnten Kennzahlen, massiv überschätzt.

Aus vorhergehenden Überlegungen kann angenommen werden, dass der begangene Fehler von der mittleren Zeit zwischen zwei Vorkommnissen und dem gewählten Beobachtungsintervall abhängt, wobei kürzere Intervalle zu einer Vergrößerung des Fehlers führen. Es besteht somit Forschungsbedarf zur Methodik der Videoanalyse.

9.8.3 Tiefenstruktur

Die Fokussierung des Ratinginstruments auf Qualitätsmerkmale experimentellen Handelns erweist sich aus folgenden Gründen als sinnvoll: Empirische Untersuchungen von Zusammenhängen müssen theoretisch begründet werden. Das gelingt durch die konsequente Verbindung des Ratingsystems mit den Qualitätsmerkmalen. Die Ansprüche an die Qualität werden durch die Ausrichtung des Instruments transparent gemacht. Der gewählte Ansatz ist für die Analyse experimentellen Handelns im Unterricht neuartig. Bisherige Kategoriensysteme beschränkten sich entweder auf die Sichtstruktur oder vermischten diese mit Tiefenstrukturmerkmalen, wobei Qualitätsansprüche nicht immer klar ersichtlich waren (vgl. Millar et al., 1999, Millar, 2009, Tesch, 2005).

Die in Bezug auf experimentelles Handeln durchgängig doppelt vorgenommene Bewertung der Einheiten erwies sich als adäquat. Diskussionen, die im Rahmen des Konsensfindungsprozesses stattfanden, trugen zur Entwicklung des Verständnisses der Facetten bei.

Nach den ersten Erfahrungen mit dem Ratinginstrument ergeben sich verschiedene Entwicklungsmöglichkeiten, die einerseits das Instrument selbst, andererseits aber auch die Anwendung betreffen:

¹⁹Z.B. die Konstrukte: Prozentuale Übereinstimmung, Cohens Kappa, Intraklassenkoeffizient, u.a.m.

- Das jetzige Ratingsystem fokussiert auf positive Items (Facetten). Um die Aussagekraft von Ergebnissen weiter zu erhöhen, wäre es sinnvoll, ebenfalls negative Items in das System aufzunehmen.
- Aus methodischen Gründen sollte die Itemzahl pro Kategorie grundsätzlich erhöht werden. Bestehende Facetten lassen sich teilweise mehrfach aufteilen; z.B. könnte die Facette *Prozessreflexion* in zwei Aspekte aufgetrennt werden: das Frageverhalten der Lehrperson sowie das Antizipieren kognitiver Prozesse.
- Als Folge von letzterem Punkt liesse sich die interne Konsistenz²⁰ der Perspektiven nachweisen.
- Bisher wurde nicht untersucht, inwiefern die im Rahmen der Experimentiereinheiten untersuchten Verhaltensweisen der Lehrperson sich einem individuellen Unterrichtsstil zuordnen lassen. Es ist anzunehmen, dass beispielsweise die Perspektiven *Kontextorientierung* und *Reflexivität* sich auch auf andere Einheiten des Unterrichts übertragen lassen.
- Um den spezifischen Chancen der verschiedenen Organisationsformen experimentellen Handelns besser Rechnung zu tragen, wäre es angebracht, das System mit speziellen Facetten zum Lehrer- und Schülerexperiment zu erweitern.

9.9 Ausblick

9.9.1 Implikationen für die Schulpraxis

Die Befunde der aktuellen Studie, Ergebnisse anderer Studien zur Praxis experimentellen Handelns²¹ sowie Metaanalysen zum Experiment im Schulunterricht²² geben Anlass, die angewandte unterrichtliche Praxis zu überdenken. Nachfolgend werden zu wichtigen Aspekten der Planung und Durchführung einer Experimentiereinheit Anregungen gegeben. Dabei wird kein Anspruch auf Vollständigkeit erhoben. Die Anregungen basieren auf den Erfahrungen der Tiefenstruktur- und den Fallanalysen (vgl. Abschnitte 9.3 und 9.5).

²⁰Z.B. ausgedrückt durch Cronbachs Alpha.

²¹Tesch (2005), Tesch & Gerber (2005), Abrahams & Millar (2008), Tiberghien et al. (2001, u.a).

²²vgl. z.B. Hodson (1993), Harlen & Wake (1999), Hofstein & Lunetta (2003), Jenkins (1999), Watson (2000).

Ziele. Die Idee, ein Experiment im Physikunterricht einzusetzen, wird bezüglich der pädagogisch-didaktischen und der fachinhaltlichen Ziele hinterfragt. Dabei lassen sich die Ziele auch für die Lernenden einfach und klar formulieren. So können Schülerinnen und Schüler verstehen, warum das Experiment genau an eine bestimmte Stelle des Unterrichts passt. Es ist hilfreich, die Relevanz der Ziele vor dem Hintergrund der naturwissenschaftlichen Bildung einzuschätzen.

Bezüge. Weiter wird untersucht, welche Beziehungen zwischen dem Experiment und verschiedenen Kontexten bestehen. Es ist sinnvoll, hier zwischen Bezügen zum unmittelbaren Kontext der betreffenden Unterrichtseinheit und Bezügen zu anderen, bereits behandelten Kontexten, zu unterscheiden. Bezüge werden nicht nur bei der Sachstruktur gesucht, sondern gerade auch bei übergeordneten Konzepten (vgl. nachfolgender Paragraph *Konzepte etablieren*) oder in der Lebenswelt ausserhalb des Physikunterrichts. An der Stelle der Planung müssen auch Alternativen zum Experiment geprüft werden, wobei auch an unkonventionelle Methoden gedacht werden sollten.

Lehrer- vs. Schülerexperiment. Auf Basis der relevanten Ziele eines Experiments und der verschiedenen Bezügen zu anderen Kontexten kann die Organisationsform gewählt werden: Es muss betont werden, dass die Formeln ‘Schülerexperiment gleich gutes konstruktivistisches Experimentieren’, ‘Lehrerexperiment gleich veraltetes lehrerzentriertes Experimentieren’ wissenschaftlich gesehen nicht haltbar sind²³. Vielmehr haben beide Organisationsformen experimentellen Handelns ihre spezifischen Chancen, die sich aus den verschiedenen Settings ergeben (vgl. Tabelle 9.1). So eignen sich Schülerexperimente vorwiegend dann, wenn Lernende ihre Ideen einbringen können, also nicht nur ‘Ausführende’ sind (für ein gelungenes Bsp. vgl. Fall S-1-05-01, Abschnitt 7.3.1). Das ist nur dann möglich, wenn Schülerinnen und Schüler mit dem Material vertraut sind. Lange Anleitungen zur Durchführung von Schülerexperimenten sind ein Hinweis dafür, dass andere Organisationsformen, z.B. das Lehrerexperiment, vorzuziehen sind. Unter den erwähnten Organisationsformen sind verschiedene Varianten²⁴ subsumiert. Für jede dieser Varianten ergibt sich eine etwas andere Gewichtung der in Tabelle 9.1 aufgelisteten Chancen. Eine weiterführende Diskussion der Organisationsformen muss an anderer Stelle erfolgen.

²³Für weitere fachdidaktische Stereotypen, vgl. Kind & Kind (2007).

²⁴Schülerexperimente einzeln, zu Zweit, in Gruppen, arbeitsteilig, arbeitsgleich, ...

Spezifische Chancen zweier Organisationsformen	
<i>Lehrerexperiment</i>	<i>Schülerexperiment</i>
<ul style="list-style-type: none"> • Fachinhaltliches Ziel im Vordergrund • Phänomen kennen lernen • Kognitive Prozesse antizipieren • Angeleitet Lernen (geschlossen) • Weitere Möglichkeit: Aktivität in eine Erzählung integrieren²⁵ 	<ul style="list-style-type: none"> • Pädagogisch-didaktisches Ziel im Vordergrund • Experimentelles Handeln erfahren • Eigene Vorstellungen, Ideen und Konzepte entwickeln • Selbstgesteuertes Lernen (offen) • Weitere Möglichkeit: Peers lernen voneinander, z.B. durch Präsentationen

Tabelle 9.1: Spezifische Chancen von Lehrer- und Schülerexperimenten.

Schülerbeiträge wertschätzen. Die Wertschätzung von Schülerbeiträgen ist immer wichtig, also nicht nur im Kontext experimentellen Handelns. So kommt den Schülerbeiträgen im Zusammenhang mit dem Experiment eine besondere Bedeutung zu, da die Lernenden gerade durch Phänomene herausgefordert und zum Mitmachen animiert werden. Wertschätzung kann ausgedrückt werden, indem Lehrende Zeit geben für Schülerbeiträge, zuhören, Rückfragen stellen, Schülerbeiträge überdenken lassen, Ideen aufnehmen und weiterentwickeln sowie auf Schülerbeiträge verweisen. Dabei ist auf die Balance zwischen der Erreichung des vordefinierten unterrichtlichen Ziels und dem Eingehen auf Schülerbeiträge zu achten. Die Qualität beim Umgang mit Schülerbeiträgen scheint wichtiger als die Quantität der verarbeiteten Beiträge zu sein.

Konzepte etablieren. Meist ist es sinnvoll, Aktivitäten experimentellen Handelns bereits bei Beginn mit tragfähigen Konzepten bzw. Theorien²⁶ zu verbinden. Keine Untersuchung kann frei von Erwartungen, Vorstellungen und Annahmen angegangen werden. Daher ist es naheliegend, die Erwartungen, Vorstellungen und

²⁵Bsp.: Eine fiktive Auseinandersetzung zwischen Aristoteles und Newton zur beschleunigten Bewegung eines Körpers vortragen. Dazu passende Beispiele einer beschleunigten Bewegung demonstrieren.

²⁶Der Begriff der Theorie soll hier offen aufgefasst werden: Er ist eine Antwort auf eine "what's going on here?"-Frage (Kosso, 1992). Es lassen sich daher folgende Begriffe subsumieren: Vorstellungen, Ideen, Konzepte, Modelle. Bsp. für Theorien: Teilchenvorstellung, Vorstellung vom Kontinuum, Ladungskonzept, Feldkonzept, Superpositionsprinzip, Wasseranalogie des elektrischen Stromkreises, Transportphänomene, ...

Annahmen im Sinne einer gedanklichen, theoretischen Vorarbeit zu artikulieren. Auch Vorstellungen der Lernenden sollen einbezogen werden. Das Argumentieren in Theorien muss geübt werden. Somit ist es wichtig, dass diese im Verlauf des Experiments gebraucht und immer wieder aufgegriffen werden.

9.9.2 Forschungsdesiderata

Die Auseinandersetzung mit einem Forschungsgegenstand wirft neue und weiterführende Fragen auf. Aus den Themenfeldern Länderspezifika, Qualität experimentellen Handelns und Methodik werden Vorschläge für neue oder weiterführende Untersuchungen gegeben. Es handelt sich hierbei lediglich um eine Auswahl, aus Sicht des Autors wichtiger Vorschläge.

Länderspezifische Unterschiede experimentellen Handelns. Der Umstand, dass in Finnland im Vergleich zum Physikunterricht in den deutschsprachigen Ländern, weniger Zeit für experimentelles Handeln eingesetzt wird, ist bemerkenswert. Es wäre allerdings wichtig zu erfahren, inwiefern sich die Unterschiede auch bei anderen unterrichtlichen Themen manifestieren. Hier könnte als Ergänzung eine, im finnischen Physikunterricht durchgeführte, Längsschnittstudie zu wertvollen Erkenntnissen kommen.

Es wurde auf die Bindung an zwei Lehrmittel in Finnland hingewiesen. Der Eindruck aus den Videoanalysen müsste wissenschaftlich weiter untermauert werden. In dem Zusammenhang wäre eine genaue Analyse deutschsprachiger und finnischer Lehrmittel angezeigt. Ergäben sich Unterschiede bezüglich der Art oder Häufigkeit der vorgeschlagenen Experimente, den Befunden der Unterrichtspraxis entsprechend, wäre dies aufschlussreich und bestätigte erneut die grosse Bedeutung der Lehrmittel bei der Planung von Unterricht.

Qualitätsmerkmale experimentellen Handelns. Die heuristisch abgeleiteten Qualitätsmerkmale wurden bezüglich ihres Effekts auf die Schülerleistung untersucht (vgl. Kapitel 8). Für eine empirische Untermauerung der Qualitätsmerkmale genügt das jedoch nicht. Zu den mit experimentellem Handeln verbundenen Zielen gehören auch affektive Variablen, wie Motivation und Interesse (vgl. Abschnitt 3.3). Daten zu derartigen Variablen wurden erhoben und bereits deskriptiv ausgewertet (Helaakoski, 2009). Offen bleibt bisher die Frage, inwiefern die vorgeschlagenen Qua-

litätsmerkmale experimentellen Handelns Prädiktoren für einen, die Motivation und das Interesse fördernden, Unterricht sind. Eine entsprechende Kovarianzanalyse ist geplant.

Methodische Weiterentwicklungen. Die Videographierung von Unterricht ist eine wertvolle Methode, wenn es darum geht, wichtige Merkmale experimentellen Handelns zu untersuchen. Sie stösst jedoch in der angewandten Form an ihre Grenzen, wenn Schülerhandlungen, insbesondere auch die Interaktion zwischen Lernenden, detailliert untersucht werden sollen. Hier könnten Schülerexperimente unter stärker kontrollierten Bedingungen in der Ergänzung zur komplexen Unterrichtssituation wichtige Beiträge zum Verhalten der Lernenden liefern.

Literaturverzeichnis

- Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1945–1969.
- Armstrong, H. E. (1898). The heuristic method of teaching or the art of making children discover things for themselves. In *The teaching of scientific method and other papers on education*. London: Macmillan and Co., Limited, 1903.
- Beatty, J. W., & Woolnough, B. E. (1982). Practical work in 11-13 science: The context, type and aims of current practice. *British Educational Research Journal*, 8(1), 23–30.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Springer, 4., vollst. überarb. Aufl.
- Börlin, J., Junge, C., & Labudde, P. (2010). Charakteristika des Physikunterrichts: Ein Ländervergleich. In D. Höttecke (Ed.) *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie*. Gesellschaft für Didaktik der Chemie und Physik (GDChP) Jahrestagung in Potsdam 2010, Lit Verlag, Berlin.
- Campanario, J. M. (2002). The parallelism between scientists' and students' resistance to new scientific ideas. *International Journal of Science Education*, 24(10), 1095–1110.
- Dalehefte, I. M., Rimmel, R., Prenzel, M., Seidel, T., Labudde, P., & Herweg, C. (2009). Observing instruction “next-door”: A video study about science teaching and learning in Germany and Switzerland. In T. Janik, & T. Seidel (Eds.) *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*, chap. 5, (pp. 83–99). Waxmann.

- DeBoer, G. E. (1991). *A history of ideas in science education: implication for practice*. New York and London: Teachers College Press.
- Domisch, R. (2009). Keine Mythen, sondern fundierte Schulreformen – Die Lernerfolge finnischer Schüler aus der Perspektive des Finnischen Zentralamtes für Unterrichtswesen. *Zeitschrift für Erziehungswissenschaft*, (pp. 617–632).
- Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, 23(7), 5–12.
- Duit, R., & Tesch, M. (in Vorb.). Eigenständiges Experimentieren im naturwissenschaftlichen Unterricht - Theorie, empirische Forschungsergebnisse, Unterrichtspraxis. In *Sonderband zur DPG-Tagung 2005 in Kassel*.
- EDK (2011). *Grundkompetenzen für die Naturwissenschaften*. Schweizerische Konferenz der Kantonalen Erziehungsdirektoren.
- Feyerabend, P. (1976, 1983). *Wider den Methodenzwang*. Frankfurt am Main: Suhrkamp Verlag.
- Fischer, H., Borowski, A., Kauertz, A., & Neumann, K. (2010). Fachdidaktische Unterrichtsforschung – Unterrichtsmodelle und die Analyse von Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 59–75.
- Fischler, H. (1994). Concerning the difference between intention and action: Teachers' conceptions and actions in physics teaching. In *Teachers' minds and actions: Research on teachers' thinking and practice*. London: Falmer Press.
- Fischler, H. (2000). Über den Einfluss von Unterrichtserfahrungen auf die Vorstellungen vom Lehren und Lernen bei Lehrerstudenten der Physik. Teil 2: Ergebnisse der Untersuchung. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 79–95.
- Geller, C. (in Vorb.). *Sequenzierung im Physikunterricht – Eine ländervergleichende Analyse von Mustern und ihren Wirkungen (Arbeitstitel)*. Ph.D. thesis, Universität Duisburg-Essen (Deutschland).
- Geller, C., Neumann, K., Boone, W., & Fischer, H. (in Vorb.). What makes the difference? An in-depth comparison of students' growth in three countries (working title).

- Geller, C., Neumann, K., & Fischer, H. (2010). Was Mittelstufenschüler über Elektrizität wissen – ein Ländervergleich. In D. Höttecke (Ed.) *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik*. Gesellschaft für Didaktik der Chemie und Physik (GDGP) Jahrestagung in Dresden 2009, Lit Verlag, Berlin.
- Gerber, B. (2007). *Strukturierung von Lehr-Lern-Sequenzen im Physikunterricht*. Ph.D. thesis, Universität Bern, Philosophisch-naturwissenschaftliche Fakultät.
- Giere, N., Roland (1991). *Understanding Scientific Reasoning*. Fort Worth, TX: Rolt, Rinehart and Winston.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857–871.
- Hammann, M., Phan, T. T. H., & Bayrhuber, H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? *Zeitschrift für Erziehungswissenschaft*, 10(Sonderheft 8), 33–49.
- Harlen, W., & Wake, R. (1999). *Effective Teaching of Science*. The Scottish Council for Research in Education.
- Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching*, 37(7), 655–675.
- Helaakoski, J. (2009). Effects of content structure and motivation on students' learning outcomes. Istanbul: European Science Education Research Association (ESERA).
- Helaakoski, J. (in Vorb.). *Content structure of physics lessons (working title)*. Ph.D. thesis, University of Jyväskylä, Finland.
- Hiebert, J., Gallimore, R., Garnier, H., Bogard Givvin, K., Hollingsworth, H., Jacobs, J., Chui, A. M.-Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries : results from the TIMSS 1999 video study*. Washington, DC: National Center for Education Statistics.

- Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education*, 22(1), 85–142.
- Hodson, D. (1996). Laboratory work as scientific method: three decades of confusion and distortion. *Journal of Curriculum Studies*, 28(2), 115–135.
- Hofstein, A., & Lunetta, V. N. (2003). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54.
- Hugener, I., Rakoczy, K., Pauli, C., & Reusser, K. (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lern-Prozesse. In S. Rahm, I. Mammes, & M. Schrantz (Eds.) *Schulpädagogische Forschung. Unterrichtsforschung. Perspektiven innovativer Ansätze*. Innsbruck: Studien Verlag.
- Janik, T., Mikova, M., Najvar, P., & Najvarova, V. (2006). Unterrichtsformen und -phasen im tschechischen Physikunterricht: Design und Ergebnisse der CPV Videostudie Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 219–238.
- Jenkins, E. W. (1999). Practical work in school science – some questions to be answered. In *Practical Work in Science Education - Recent Research Studies*, vol. 1., (pp. 33–59). Roskilde University Press.
- Jonas-Ahrend, G. (2004). *Physiklehrervorstellungen zum Experiment im Physikunterricht*. Logos: Berlin.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Logos: Berlin.
- Keller, M. (2011). *Teacher Enthusiasm in Physics Instruction*. Ph.D. thesis, Universität Duisburg-Essen.
- Kerr, J. F. (1964). *Practical Work in School Science*. Leicester: Leicester University Press.
- Kind, P. M., & Kind, V. (2007). Creativity in science education: Perspectives and challenges for developing school science. *Studies in Science Education*, 43, 1–37.
- Kircher, E., Girwidz, R., & Häußler, P. (2001). *Physikdidaktik: eine Einführung*. 2. Auflage. Berlin: Springer.

- KMK (2004). *Bildungsstandards in Physik für den Mittleren Schulabschluss*. Bonn: Kultusministerkonferenz KMK.
- Knierim, B. (2008). *Lerngelegenheiten anbieten - Lernangebote nutzen Eine Videostudie im Schweizer Physikunterricht*. Hamburg: Verlag Dr. Kovac.
- Kobarg, M., & Prenzel, M. (2009). Stichwort: Der Mythos der nordischen Bildungssysteme. *Zeitschrift für Erziehungswissenschaft*, (pp. 597–615).
- Koponen, I. T., & Mäntylä, T. (2006). Generative role of experiments in physics and in teaching physics: A suggestion for epistemological reconstruction. *Science & Education*, 15, 31–54.
- Kosso, P. (1992). *Reading the Book of Nature: An Introduction to the Philosophy of Science*. Cambridge University Press.
- Kuhn, T. S. (1962, 1970). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
- Labudde, P. (2000). *Konstruktivismus im Physikunterricht der Sekundarstufe II*. Bern: Haupt.
- Labudde, P. (2006). *Lernen und Lehren der Physik: ein Vergleich des Physikunterrichts und seiner Wirkungen in Deutschland, Finnland und der Schweiz (Projektantrag)*. Bern: PHBern.
- Lakin, S., & Wellington, J. (1991). *Teaching the Nature of Science. A study of Teachers' Views of Science and their Implications for Science Education*. Sheffield: Division of Education, University of Sheffield.
- Laudan, L., Donovan, A., Laudan, R., Barker, P., Brown, H., Leplin, J., Thagard, P., & Wykstra, S. (1986). Scientific change: Philosophical models and historical research. *Synthese*, 69, 141–223. 10.1007/BF00413981.
- Lipowsky, F., Rakoczy, K., Klieme, E., Reusser, K., & Pauli, C. (2005). Unterrichtsqualität im Schnittpunkt unterschiedlicher Perspektiven. In *Schulentwicklung und Schulwirksamkeit*, (pp. 223–238). Weinheim und München: Juventa Verlag.

- Lunetta, V. N. (1998). The school science laboratory: Historical perspectives and contexts for contemporary teaching. In *International handbook of science education*, (pp. 249–262). Springer.
- Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and teaching in the school science laboratory: An analysis of research, theory and practice. In *Research on Science Education*, (pp. 393–441). Mahwah, New Jersey: Erlbaum.
- Mayring, P. (2002). *Qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken*. 5., überarb. Auflage. BELTZ.
- Millar, R. (1989). What is ‘scientific method’ and can it be taught? In J. Wellington (Ed.) *Skills and processes in science education: a critical analysis*. Routledge.
- Millar, R. (2004). The role of practical work in the teaching and learning of science. In *High School Science Laboratories: Role and Vision*. National Academy of Sciences, Washington, DC: The University of York.
- Millar, R. (2009). *Analysing practical activities to assess and improve effectiveness: The Practical Activity Analysis Inventory (PAAI)*. York: Centre for Innovation and Research in Science Education, University of York.
- Millar, R., Le Maréchal, J.-F., & Tiberghien, A. (1999). ‘mapping’ the domain: Varieties of practical work. In *Practical Work in Science Education - Recent Research Studies*, vol. 1., (pp. 33–59). Roskilde University Press.
- Millar, R., Tiberghien, A., & Le Maréchal, J.-F. (2002). Varieties of labwork: A way of profiling labwork tasks. In *Teaching and Learning in the Science Laboratory*. Kluwer.
- Monk, M., & Dillon, J. (2000). The nature of scientific knowledge. In *Good practice in science teaching: what research has to say*. Open University Press.
- Muckenfuss, H. (2004). Themen und Kontexte als Strukturelemente des naturwissenschaftlichen Unterrichts: Zu den Schwierigkeiten, systematisches Physiklernen zu organisieren. *Physik und Didaktik in Schule und Hochschule*, 2/3, 57–66.
- Nersessian, N. J. (1991). Conceptual change in science and in science education. In M. R. Matthews (Ed.) *History, Philosophy, and Science Teaching: Selected Readings*, (pp. 133–148). Oise Press.

- Neumann, K., Fischer, H., Labudde, P., & Viiri, J. (2009). Postersymposium Physikunterricht im Vergleich: Unterrichtsqualität in Deutschland, Finnland und der Schweiz. In D. Höttecke (Ed.) *Chemie- und Physikdidaktik für die Lehramtsausbildung*. Gesellschaft für Didaktik der Chemie und Physik (GDGP) Jahrestagung in Schwäbisch Gmünd 2008, Lit Verlag, Berlin.
- Neumann, K., Fischer, H., Labudde, P., & Viiri, J. (2010). Physikunterricht: Deutschland, Finnland und die Schweiz im Vergleich. In D. Höttecke (Ed.) *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik*. Gesellschaft für Didaktik der Chemie und Physik (GDGP) Jahrestagung in Dresden 2009, Lit Verlag, Berlin.
- Ntombela, G. M. (1999). A marriage of inconvenience? school science practical work and the nature of science. In J. Leach, & A. C. Paulsen (Eds.) *Practical work in science education - Recent research studies*, (pp. 118–133). Roskilde University Press.
- OECD (2001). *Knowledge and Skills for Life – first results from the OECD programme for international student assessment (PISA) 2000*. OECD Publishing.
- OECD (2004). *Learning for Tomorrow's World - First results from PISA 2003*. OECD Publishing.
- OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World: Executive Summary*. OECD Publishing.
- OECD (2010). *PISA 2009 at a Glance*. OECD Publishing.
- Olszewski, J. (2010). *The Impact of Physics Teachers' Pedagogical Content Knowledge on Teacher Actions and Students Outcomes*. Logos: Berlin.
- Pauli, C., & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *52. Jahrgang*(Heft 6), 774–798.
- Petko, D., Waldis, M., Pauli, C., & Reusser, K. (2003). Methodische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. Ansätze der TIMSS 1999 Video Studie und ihrer schweizerischen Erweiterung. *Zentralblatt für Didaktik der Mathematik*, 35(6), 265–280.

- Pomeroy, D. (2006). Implications of teachers' beliefs about the nature of science: Comparison of the beliefs of scientists, secondary science teachers, and elementary teachers. *Science Education*, 77(3), 261–278.
- Psillos, D., & Niedderer, H. (2003). *Teaching and learning in the science laboratory*. Berlin: Springer.
- Röhneck, C. (1986). Vorstellungen vom elektrischen Stromkreis und zu den Begriffen Strom, Spannung und Widerstand. *Naturw. im Unterricht Physik/Chemie, Heft 13*, 10–14.
- Rost, J. (1996). *Testtheorie Testkonstruktion*. 1. Auflage. Verlag Hans Huber.
- Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., & Maechler, M. (2011). *robustbase: Basic Robust Statistics*. R package version 0.7-3.
- Sachs, L., & Hedderich, J. (2006). *Angewandte Statistik*. 12. Auflage. Springer.
- Seidel, T., Prenzel, M., & Kobarg, M. (2005). *How to run a video study; Technical report of the IPN Video study*. Münster: Waxmann.
- Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classroom and teaching across cultures: Examples and lessons from the timss video studies. *Educational Psychologist*, 35(2), 87–100.
- Suzuki, R., & Shimodaira, H. (2009). *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. R package version 1.2-1.
- Swain, J., Monk, M., & Sally, J. (2000). Development in science teachers' attitudes to aims for practical work: continuity and change. *Teacher Development*, 4(2), 281–292.
- Szlovák, B. (2005). *Lehrplanvergleich-Naturwissenschaften Oktober 2005*. EDK.
- Tesch, M. (2005). *Das Experiment im Physikunterricht: Didaktische Konzepte und Ergebnisse einer Videostudie*. Berlin: Logos: Berlin.
- Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht - Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51–69.

- Tesch, M., & Gerber, B. (2005). On the role of experiments in physics instruction – different emphasis in different instructional traditions. Paper presented at the European Science Education Research Association (ESERA).
- Tiberghien, A., Veillard, L., Le Maréchal, J.-F., Buty, C., & Millar, R. (2001). An analysis of labwork tasks used in science teaching at upper secondary school and university levels in several European countries. *Science Education*, *85*, 483–508.
- Trumper, R. (2003). The physics laboratory ? A historical overview and future perspectives. *Science & Education*, *12*(7), 645–670.
- van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Clarendon Press.
- Watson, R. (2000). The role of practical work. In *Good practice in science teaching: what research has to say*. Open University Press.
- Weinert, F. E. (2001). Vergleichende Leistungsmessungen in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.) *Leistungsmessungen in Schulen*, (pp. 17–31). Weinheim: Beltz.
- Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., Paulsen, A., Robinault, K., & Aufschnaiter, S. v. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden - Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, *4*, 29–44.
- White, R. T. (1996). The link between the laboratory and learning. *International Journal of Science Education*, *18*(7), 761–774.
- Wilson, M. (2005). *Constructing measures: an item response modeling approach*. No. Bd. 1 in Applied Psychology Series. Lawrence Erlbaum Associates.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Freiburg: Hogrefe.
- Wissenschaftliches Konsortium HarmoS (2008). *HarmoS Naturwissenschaften+: Kompetenzmodell und Vorschläge für Bildungsstandards*. Bern: PHBern.
- Wong, S. L., & Hodson, D. (2009). More from the horse's mouth: What scientists say about science as social practice. *International Journal of Educational Research*.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2), 642–656.

Zberg, U. (in Vorb.). *Experimentierkompetenz im tri-nationalen Vergleich (Arbeitstitel)*. Master's thesis, PH-FHNW.

Anhang A

Sichtstruktur experimentellen Handelns

A.1 Das Kategoriensystem (vgl. Tesch, 2005)

Aus jeder Kategorie kann jeweils nur eine Subkategorie gewählt werden. Einige Kategorien werden in bestimmten Unterrichtsabschnitten nicht kodiert. Es wird dann die Kategorie kein/keine kodiert. Wechselt eine Subkategorie innerhalb eines Beobachtungsintervalls, so erfolgt die Zuordnung der Subkategorie nach der folgenden Vorschrift:

1. Spezifische Kodierregel (Vgl. Beschreibung der Kategorie)
2. Bei Wechsel zwischen Kategorien ungleich keine: Wahl jener Subkategorie, der der grössere Teil des Intervalls zugeordnet werden kann.
3. Bei Wechsel zwischen keine und der Kategorie x: Wahl der Kategorie x.

Experimentierphasen E1

Die Kodierung der Experimentierphasen erfolgt disjunkt in folgende Kategorien:

Experimentierphasen E1
0: keine
1: Vorbereitung Experiment
2: Durchführung
3: Nachbereitung Experiment

Alle weiteren Kodierungsdurchgänge beziehen sich dann nur noch auf Unterrichtssequenzen, die mit Subkategorien 1 bis 3 der Kategorie Experimentierphasen kodiert wurden.

0: keine

Inhaltliche Bestimmung Der Unterricht hat keinerlei Bezüge zu einem Experiment, das noch stattfinden soll oder stattgefunden hat. Berichte über Experimente, die nicht im Unterricht durchgeführt werden oder wurden, werden als keine kodiert.

Beschreibung auf der Beobachtungsebene Es wird nicht Experimentieren kodiert, wenn z.B. ein Gerät gezeigt und erklärt, aber nicht vorgeführt wird.

Spezifische Kodierungsregeln Wird ein Gerät gezeigt und erklärt, das später in einem Experiment vorkommt, so wird es bereits in *Vorbereitung Experiment* kodiert.

1: Vorbereitung

Inhaltliche Bestimmung Der Unterricht steht im Kontext eines bevorstehenden Experimentes. Die Fragestellung des Experimentes wird bekannt gegeben, über den möglichen Ausgang eines Experiments gesprochen, Vorstellungen werden erfragt oder Hypothesen aus einer physikalischen Theorie abgeleitet. Der Aufbau wird erklärt, die Durchführung geplant, eine Anleitung wird entwickelt oder mitgeteilt. Dazu gehören auch das Aufbauen des Experimentes sowie alle anderen praktischen Vorbereitungsarbeiten. Auch ein mit dem bevorstehenden Experiment verbundener Ortswechsel der S oder der LP gehören dazu.

Beschreibung auf der Beobachtungsebene Zum Beispiel kündigt die Lehrperson einen Versuch an, oder die Schülerinnen und Schüler werden aufgefordert, Ideen für einen Versuch zu äußern; der Aufbau eines Versuchs wird an die Tafel gezeichnet; die Lehrperson gibt praktische Hinweise zur Durchführung des Versuchs.

Spezifische Kodierungsregeln Die Vorbereitung eines Experiments beginnt häufig mit Äusserungen wie z.B. "Wir wollen jetzt einen Versuch durchführen" und

endet z.B. mit “Dann fangt bitte an”. Schülerexperimentierphasen: Es wird Experimentieren kodiert, wenn eine Mehrheit der S am Experiment hantiert. Das Holen der Experimentierkästen gehört jedoch zur Vorbereitung.

2: Durchführung

Inhaltliche Bestimmung Das Experiment wird durchgeführt, es wird beobachtet oder gemessen. Es kann hier zu kurzen Planungsphasen zwischen den eigentlichen Aktivitäten kommen. Diese werden auch als *Durchführung* kodiert.

Beschreibung auf der Beobachtungsebene Es findet eine Interaktion mit dem Experiment (bzw. Materialien) statt, mit dem Ziel, die zuvor gestellte Fragestellung zu beantworten. Allgemeine Einstell- und Aufbauarbeiten gehören jedoch zur Vorbereitung. Beispiel: Eine Schülerin baut ein Demonstrationsexperiment auf und berät zwischendurch mit der Lehrperson am Schaltbild, das an die Tafel gezeichnet wurde, was als Nächstes zu tun ist.

Spezifische Kodierungsregeln Die Intervalle, in denen das Experiment stattfindet, werden auch dann zur *Durchführung* gezählt, wenn andere Unterrichtsphasen zeitlich überwiegen. Entscheidend ist eine Unterscheidung von *Durchführung* und Nachbereitung. Wenn die Lehrkraft das Experiment unterbricht, um etwas zu diskutieren, wird *Nachbereitung* kodiert. Wenn die Lehrperson etwa bei Problemen sagt: “Hört mal eben auf. Wir müssen etwas besprechen...”, so wird *Nachbereitung* kodiert für die Besprechungsphase. Im Gegensatz zu Planungsphasen, die während der Durchführung auftreten können, beziehen sich diese Besprechungsphasen auf die Interpretation des Experiments. 17. Juli 2009, D-1-25-01: Situation: Verschiedene Geräte werden nacheinander an ein Leistungsmessgerät angeschlossen. Die Leistung wird abgelesen und aufgeschrieben, danach wird ein nächstes Gerät angeschlossen. Intervalle, während derer ausschliesslich Werte abgelesen, aufgeschrieben oder ein neues Gerät angeschlossen werden, zählen wir zur *Durchführung*.

3: Nachbereitung

Inhaltliche Bestimmung Wird kodiert, wenn die *Durchführung* endet bzw. unterbrochen wird mit dem Ziel, Erkenntnisse aus dem Experiment auszuwerten, zu interpretieren oder zu diskutieren. Zur Nachbereitung gehören ausserdem: Das

Aufräumen des Experimentes sowie ein durch das Experimentieren bedingter Ortswechsel.

Beschreibung auf der Beobachtungsebene Die Phase beginnt oft mit: 1) “Hört jetzt bitte auf zu experimentieren. Wir wollen jetzt mal die Ergebnisse an der Tafel zusammentragen...” 2) Die Schülerinnen und Schüler gehen nach der Vorführung des Experimentes zu ihrem Platz zurück.

Spezifische Kodierungsregeln Die Phase der Nachbereitung endet, wenn weiter experimentiert wird oder wenn kein expliziter Zusammenhang des Unterrichts zum Experiment mehr erkennbar ist.

Nummerierung der Experimentiereinheiten E2

Diese Kategorie wird immer dann kodiert, falls bei E1 die Kategorie 1, 2 oder 3 gewählt wird. Die Experimentierphasen werden nummeriert:

<p>Experimentiereinheiten E1</p> <p>1: Experimentiereinheit Nr. 1</p> <p>2: Experimentiereinheit Nr. 2</p> <p>3: Experimentiereinheit Nr. 3</p> <p>usw.</p>
--

Inhaltliche Bestimmung Die Experimente werden fortlaufend bei 1 beginnend über die ganze Doppelstunde nummeriert. Dabei werden alle Experimentiereinheiten (also auch die Vor- und Nachbereitungsphase) mit der jeweiligen Nummer versehen. Die Vorbereitung, die Durchführung und die Nachbereitung eines Experiments werden mit derselben Nummer belegt. Experimente, die wiederholt werden, bekommen wieder die gleiche Nummer. Werden mehrere Experimente gleichzeitig vor- oder nachbereitet, so wird bei der Vorbereitung die Nummer des ersten Experiments, bei der Nachbereitung die Nummer des letzten Experiments kodiert.

Beschreibung auf der Beobachtungsebene Die Frage, wann ein neues Experiment kodiert werden soll, wird nach folgenden Entscheidungskriterien beurteilt (geordnet nach Wichtigkeit):

1. Ankündigung eines neuen Experiments der Lehrperson
2. Beim Wechsel zu einer neuen Fragestellung
3. Falls die Fragestellung nicht bekannt: Bei einem Wechsel der Messinstrumente, evtl. auch der Untersuchungsobjekte

Spezifische Kodierungsregeln Die Wiederholung eines Experimentes wird nicht als neue Experimentiereinheit kodiert.

Wird derselbe Aufbau eines Experiments benutzt, um einen anderen Sachverhalt experimentell zu veranschaulichen, wird die Experimentiereinheiten-Nr. erhöht.

Art der Beschäftigung G3

Wird dann kodiert, wenn für E1 1 oder 3 gesetzt wird.

Art der Beschäftigung G3

- 1: Inhaltliche Beschäftigung
- 2: Organisatorische Beschäftigung

Alle weiteren Kodierungsdurchgänge beziehen sich dann nur noch auf Unterrichtssequenzen, die mit Subkategorien 1 bis 3 der Kategorie Experimentierphasen kodiert wurden.

1: Inhaltliche Beschäftigung

Inhaltliche Bestimmung Das Experiment wird inhaltlich vor- oder nachbereitet.

Dabei kann es um physikalische Inhalte, Methoden, Anwendungen oder historische Zusammenhänge gehen.

Beschreibung auf der Beobachtungsebene Die Lehrperson stellt zum Beispiel ein physikalisches Problem vor, Experimentiererergebnisse werden diskutiert oder es werden Anwendungen besprochen.

Spezifische Kodierungsregeln Wird organisiert und gleichzeitig über einen physikalischen Inhalt gesprochen, so wird "inhaltliche Beschäftigung" kodiert.

2: Organisatorische Beschäftigung

Inhaltliche Bestimmung Der Arbeitsablauf eines Experiments wird organisiert.

Beschreibung auf der Beobachtungsebene Es werden Experimentiermaterialien verteilt. Experimente werden auf- und abgebaut.

Aufgrund eines bevorstehenden Experiments ändern die Schülerinnen und Schüler ihre Sitzordnung, verlassen das Zimmer oder teilen sich in Gruppen auf.

Spezifische Kodierungsregeln Wird organisiert und gleichzeitig über einen physikalischen Inhalt gesprochen, so wird "inhaltliche Beschäftigung" kodiert.

Datenerfassung K1

Wird kodiert, wenn E1 gleich 2 gewählt ist.

Datenerfassung K1

- 1: Qualitativ
- 2: Quantitativ
- 3: Unklar

Alle weiteren Kodierungsdurchgänge beziehen sich dann nur noch auf Unterrichtssequenzen, die mit Subkategorien 1 bis 3 der Kategorie Experimentierphasen kodiert wurden.

1: Qualitativ

Inhaltliche Bestimmung Das Experiment wird ausschliesslich über die Sinne wahrgenommen. Es handelt sich um ein Experiment ohne Aufnahme von Messdaten in Form von Zahlen oder Graphen.

Beschreibung auf der Beobachtungsebene Ein Wagen rollt beispielsweise gegen eine Mauer, ein Fussball wird getreten und es soll beschrieben werden, was man gesehen hat.

Spezifische Kodierungsregeln Ein Experiment wird als qualitativ kodiert, wenn keine Messdaten aufgeschrieben werden.

2: Quantitativ

Inhaltliche Bestimmung Als quantitativ bezeichnet man Experimente mit Messdatenaufnahme. Das Ziel des Versuchs wird erst durch eine Auswertung der numerischen Daten erreicht.

Beschreibung auf der Beobachtungsebene Beispielsweise wird ein Gummiband gedehnt und die Daten der Ausdehnungslänge und der dafür benötigten Kraft werden notiert.

Spezifische Kodierungsregeln Vorsicht bei Zeigerinstrumenten! Sie zeigen zwar numerische Werte an, werden aber häufig eingesetzt, um qualitative Zusammenhänge zu zeigen. Ein Experiment wird als quantitativ kodiert, wenn Messdaten aufgeschrieben werden.

3: Unklar Die Kategorie unklar wird dann kodiert, wenn verschiedene Schülerexperimente durchgeführt werden, die sowohl qualitativ als auch quantitativ sind.

Organisationsform K2a

Wird kodiert, wenn E1 gleich 2 gewählt ist.

Organisationsform K2a

- | |
|---|
| <ul style="list-style-type: none"> 1: Demonstration Lehrperson 2: Demonstration Schülerin/Schüler 3: Schülerin/Schüler einzeln 4: Partnerarbeit 5: Gruppenarbeit |
|---|

1: Demonstration Lehrperson

Inhaltliche Bestimmung Die Lehrperson demonstriert im Plenum Experimente. Dabei können auch einzelne Schüler als Helfer auftreten.

Beschreibung auf der Beobachtungsebene Die Lehrperson steht zum Beispiel vor der Klasse und betätigt einen Expander.

Spezifische Kodierungsregeln Wenn eine Schülerin oder ein Schüler nur kurz etwas ablesen soll, wird weiter “Demonstration Lehrperson” kodiert.

2: Demonstration Schülerin/ Schüler

Inhaltliche Bestimmung Experimente werden durch einzelne oder mehrere Schülerinnen und Schüler im Plenum vorgeführt. Verantwortung für die Durchführung der Experimente liegt bei den Schülerinnen und Schülern. Die Lehrperson nimmt eine Helferrolle ein bzw. verhält sich passiv.

Beschreibung auf der Beobachtungsebene Oft wird eine Schülerin oder ein Schüler nach vorn an das Pult gebeten, wo schon Experimentiermaterialien bereit liegen.

Spezifische Kodierungsregeln Leistet die Lehrperson während eines Schüler-Demonstrationsexperiments Hilfestellung, wird weiter “Demonstration Schülerin/ Schüler” kodiert.

3: Schülerin/Schüler einzeln

Inhaltliche Bestimmung Schülerinnen und Schüler arbeiten alleine an einem Experiment.

Beschreibung auf der Beobachtungsebene Zum Beispiel soll jede Schülerin und jeder Schüler allein aus einem Haushaltspapier eine “Schneeflocke” basteln und fliegen lassen.

Spezifische Kodierungsregeln Es ist darauf zu achten, wie der Arbeitsauftrag von der Lehrperson formuliert wurde.

4: Partnerarbeit

Inhaltliche Bestimmung Schülerinnen und Schüler arbeiten zu zweit an einem Experiment.

Spezifische Kodierungsregeln Treten neben den Zweier-Gruppen auch Gruppen mit mehr als zwei Schülern auf, wird *Gruppenarbeit* kodiert.

5: Gruppenarbeit

Inhaltliche Bestimmung Schülerinnen und Schüler arbeiten in Gruppen. Eine Gruppe besteht aus drei oder mehr Mitgliedern.

Beschreibung auf der Beobachtungsebene Oft sollen die Schülerinnen und Schüler in 3er bis 4er Gruppen, die bereits an ihren Tischen sitzen, ein Experiment aufbauen und durchführen.

Spezifische Kodierungsregeln Die Lehrperson kündigt zu Beginn des Experiments meist die Organisationsform an.

Experimente an Stationen K2b

Wird kodiert, wenn E1 gleich 2 und K2a gleich 3, 4 oder 5 gewählt sind.

Experimente an Stationen K2b

1: Nein

2: Ja

1: Nein

Inhaltliche Bestimmung Schülerinnen und Schüler arbeiten an einem vorgegebenen Experiment oder arbeiten mehrere vorgegebene Experimente in bestimmter Reihenfolge ab. Es besteht keine Wahlmöglichkeit aus einer grösseren Anzahl Experimente.

2: Ja

Inhaltliche Bestimmung Schülerinnen und Schüler arbeiten an verschiedenen Experimenten. Bezüglich der Reihenfolge der Durchführung und der Wahl der Experimente haben die Schülerinnen und Schüler Freiheit. Schülerinnen und Schüler bearbeiten nur einen Teil aller zur Verfügung stehenden Experimente.

Beschreibung auf der Beobachtungsebene Lernen an Stationen wird häufig angekündigt: "Ihr sollt heute an Stationen einige Experimente durchführen."

Arbeitsverteilung Schülerexperiment K3

Wird kodiert, wenn E1 gleich 2 und K2a gleich 3, 4 oder 5 gewählt sind.

Arbeitsverteilung Schülerexperiment K3

1: Arbeitsgleich

2: Arbeitsteilig

1: Arbeitsgleich

Inhaltliche Bestimmung Alle Schülerinnen und Schüler bzw. alle Gruppen erhalten die gleiche Experimentieraufgabe.

Beschreibung auf der Beobachtungsebene Zum Beispiel sollen die Schülerinnen und Schüler aus einem mitgebrachten Karton eine Kamera Obscura bauen.

2: Arbeitsteilig

Inhaltliche Bestimmung Die Schülerinnen und Schüler bzw. Gruppen arbeiten an verschiedenen Experimentieraufgaben.

Beschreibung auf der Beobachtungsebene Zum Beispiel sollen verschiedene Materialien auf elektrische Leitfähigkeit untersucht werden. Die Materialien werden an verschiedene Gruppen/Personen verteilt.

Sachbegegnung K5

Wird kodiert, wenn E1 gleich 2 gewählt ist.

Sachbegegnung K5

1: Alltagsgeräte

2: Physikalische Geräte

3: Bildschirmexperiment

1: Alltagsgeräte

Inhaltliche Bestimmung Es werden Alltagsgeräte und -materialien beim Experimentieren eingesetzt. Das Gerät bzw. Material kann dabei eine andere Funktion einnehmen als im Alltag.

Beschreibung auf der Beobachtungsebene Einige Beispiele: Strohalm als Stativmaterial, Wäscheklammer als Schalter, Skateboard für Versuche zur Trägheit, Auto, Fahrrad...

Spezifische Kodierungsregeln Wird zusätzlich zu Alltagsgeräten z.B. eine Spannungsquelle eingesetzt, wird dennoch *Alltagsgeräte* kodiert. Dominieren hingegen die physikalischen Geräte, wird *Physikalische Geräte* kodiert.

2: Physikalische Geräte

Inhaltliche Bestimmung Es werden Geräte und Materialien eingesetzt, die funktional auf die Versuchssituation abgestimmt sind. Die Geräte bzw. Materialien sind für den Physikunterricht hergestellt worden.

Beschreibung auf der Beobachtungsebene Einige Beispiele: vorgefertigte Kabel, Spannungsquellen, Stativmaterial, Massestücke...

Spezifische Kodierungsregeln Man achte auf die handelsüblichen Lehrmittel.

3: Bildschirmexperiment

Inhaltliche Bestimmung Mit Hilfe eines Computerprogramms werden virtuelle Experimente durchgeführt.

Beschreibung auf der Beobachtungsebene Es wird ein Computer genutzt und ein Modell-Experiment projiziert.

A.2 Beobachterübereinstimmung der Sichtstruktur

A.2.1 Übereinstimmungsmasse Phase 1

In den Kreuztabellen Tabellen A.1-A.8 sind die Ergebnisse der doppelten Kodierung der ersten 6 Doppelstunden aufgeführt (Phase 1). Diese wurden zufällig aus der Teilstichprobe aller deutschen und Schweizer Klassen gezogen:

S-1-05-01, S-2-13-01, S-2-28-01, S-3-09-01, D-1-25-01, D-2-42-01.

Dabei bezeichnen die Kürzel fs und lt die beiden Kodierenden. Die Anzahl der beobachteten Intervalle variiert, da einzelne Kategorien aufgrund von Abhängigkeiten

nur für eine Teilmenge aller Intervalle zugeordnet werden können. In der Tabelle 5.2 sind das Cohens κ und die prozentuale Übereinstimmung für jede Kategorie aufgeführt. Mit Ausnahme einer Kategorie werden sehr gute Übereinstimmungen $\kappa > .80$ erzielt. Der Wert von $\kappa = .67$ für die Organisationsform rührt daher, dass fs aufgrund eines Versehens anstelle von Gruppenarbeit (5) für 55 Intervalle Partnerarbeit (4) kodiert hat (vgl. Tabelle A.5). Wie aus der Tabelle A.6 hervorgeht, wird der Kategorie Organisationsform Schülerexperiment von fs und lt konstant 1 zugeordnet. Aus diesem Grund lässt sich κ hier nicht berechnen.

A.2.2 Übereinstimmungsmasse Phase 2

Anhand der Kreuztabellen Tabellen A.9-A.16 lassen sich die Ergebnisse der doppelten Kodierungen der Phase 2 nachvollziehen. Auch diese Doppelstunden wurden zufällig aus allen deutschen und Schweizer Klassen gezogen:

D-0-33-01, D-1-29-01, D-2-23-01, D-3-32-01, S-1-33-01.

Für die Diskussion der Ergebnisse sei auf Kapitel 5 verwiesen.

fs \ lt	0	1	2	3	Σ
0	720	7	1	27	755
1	4	278	36	24	342
2	0	12	246	25	283
3	8	2	17	277	304
Σ	732	299	300	353	1684

Tabelle A.1: Kategorie Experimentierphasen E1, Analyseprozess Phase 1

fs \ lt	0	1	2	Σ
0	0	3	0	3
1	0	376	7	383
2	0	1	314	315
Σ	0	380	321	707

Tabelle A.2: Nummerierung Experimentierphasen E2, Analyseprozess Phase 1

fs \ lt	lt			Σ
	0	1	2	
0	0	1	0	1
1	0	432	11	443
2	0	14	73	87
Σ	0	447	84	531

Tabelle A.3: Art der Beschäftigung G3, Analyseprozess Phase 1

fs \ lt	lt				Σ
	0	1	2	3	
0	0	0	1	0	1
1	0	159	9	0	168
2	0	7	70	0	77
3	0	0	0	0	0
Σ	0	166	80	0	246

Tabelle A.4: Datenerfassung K1, Analyseprozess Phase 1

fs \ lt	lt						Σ
	0	1	2	3	4	5	
0	0	0	0	0	0	0	0
1	0	73	0	0	0	0	73
2	0	0	2	0	0	0	2
3	0	0	0	0	0	0	0
4	0	0	0	0	31	55	86
5	0	0	0	0	0	85	85
Σ	0	73	2	0	31	140	246

Tabelle A.5: Organisationsform K2a, Analyseprozess Phase 1

fs \ lt	lt			Σ
	0	1	2	
0	0	0	0	0
1	0	171	0	171
2	0	0	0	0
Σ	0	171	0	171

Tabelle A.6: Organisationsform Schülerexperiment K2b, Analyseprozess Phase 1

fs \ lt	0	1	2	Σ
0	0	0	0	0
1	0	116	0	116
2	0	1	54	55
Σ	0	117	54	171

Tabelle A.7: Arbeitsverteilung beim Schülerexperiment K3, Analyseprozess Phase 1

fs \ lt	0	1	2	3	Σ
0	0	0	0	0	0
1	0	0	10	0	10
2	1	0	204	0	205
3	0	0	0	31	31
Σ	1	0	214	31	246

Tabelle A.8: Sachbegegnung K5, Analyseprozess Phase 1

fs \ lt	0	1	2	3	Σ
0	658	17	0	18	693
1	1	130	10	4	145
2	1	8	281	23	313
3	24	9	1	152	186
Σ	684	164	292	197	1337

Tabelle A.9: Experimentierphasen E1, Analyseprozess Phase 2

fs \ lt	0	1	2	Σ
0	0	3	0	3
1	0	280	18	298
2	0	9	43	52
Σ	0	292	61	353

Tabelle A.10: Nummerierung Experimentierphasen E2, Analyseprozess Phase 2

fs \ lt	lt			Σ
	0	1	2	
0	0	1	0	1
1	0	238	8	246
2	0	6	42	48
Σ	0	245	50	295

Tabelle A.11: Art der Beschäftigung G3, Analyseprozess Phase 2

fs \ lt	lt				Σ
	0	1	2	3	
0	0	1	0	0	1
1	0	55	14	0	69
2	0	7	204	0	211
3	0	0	0	0	0
Σ	0	63	218	0	281

Tabelle A.12: Datenerfassung K1, Analyseprozess Phase 2

fs \ lt	lt						Σ
	0	1	2	3	4	5	
0	0	0	0	0	0	0	0
1	0	120	0	0	0	0	120
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	161	161
Σ	0	120	0	0	0	161	281

Tabelle A.13: Organisationsform K2a, Analyseprozess Phase 2

fs \ lt	lt			Σ
	0	1	2	
0	0	0	0	0
1	0	161	0	161
2	0	0	0	0
Σ	0	161	0	161

Tabelle A.14: Organisationsform Schülerexperiment K2b, Analyseprozess Phase 2

fs \ lt	0	1	2	Σ
0	0	0	0	0
1	0	161	0	161
2	0	0	0	0
Σ	0	161	0	161

Tabelle A.15: Arbeitsverteilung beim Schülerexperiment K3, Analyseprozess Phase 2

fs \ lt	0	1	2	3	Σ
0	0	0	1	0	1
1	0	2	161	0	163
2	0	31	86	0	117
3	0	0	0	0	0
Σ	0	33	248	0	281

Tabelle A.16: Sachbegegnung K5, Analyseprozess Phase 2

Anhang B

Tiefenstruktur experimentellen Handelns

Die Teilstichprobe

<hr/>			
	Land		
	D	FIN	CH
<hr/>			
D-0-02-01	F-0-01-01	S-0-06-01	
D-0-12-01	F-0-03-01	S-0-16-01	
D-0-46-01	F-0-04-01	S-0-24-01	
D-1-05-01	F-0-05-01	S-1-05-01	
D-1-06-01	F-0-08-01	S-1-18-01	
D-1-25-01	F-0-09-01	S-1-27-01	
D-1-29-01	F-0-10-01	S-1-29-02	
D-2-04-01	F-0-12-01	S-2-03-01	
D-2-15-01	F-0-14-01	S-2-13-01	
D-2-42-01	F-0-16-01	S-2-15-01	
D-2-43-01	F-0-18-01	S-2-26-01	
D-3-07-01	F-0-19-01	S-3-01-01	
D-3-22-01	F-0-21-01	S-3-11-01	
D-3-35-01	F-0-23-01	S-3-23-01	
D-3-37-02	F-0-24-01	S-3-31-01	

Tabelle B.1: Die Identifikationsnummern (ID) der Fälle (Klassen) der gezogenen Teilstichprobe. Die ID enthält folgende Angaben: *Land-Schultyp-Schulnummer-Klassennummer*

Anhang C

Fallanalysen

Ergebnisse der hierarchischen Clusteranalyse nach dem *Ward*-Verfahren. Als Abstandsmaß wurde die *Gower*-Metrik verwendet. Cluster mit $p > .95$ (approximately unbiased probability value, pvclust) werden im folgenden aufgeführt:

ID	Exp. Nr.	EB1	EB2	ZK	PR	ER	KA	TF	BB
<i>Cluster I</i>									
F-0-21-01	2	4	4	4	4	4	4	4	4
S-0-24-01	1	4	4	4	4	4	4	4	4
<i>Cluster II</i>									
D-2-15-01	3	3	3	3	4	4	4	4	2
D-3-35-01	1	3	3	3	4	4	3	4	2
<i>Cluster III</i>									
F-0-09-01	1	2	4	4	4	4	4	4	4
F-0-18-01	1	3	4	4	4	3	4	4	4
<i>Cluster IV</i>									
D-1-06-01	1	3	2	4	3	3	3	4	1
F-0-21-01	1	4	2	4	3	3	2	4	2
<i>Cluster V</i>									
D-3-37-02	1	4	3	4	4	3	4	4	3

Fortsetzung auf der nächsten Seite

ID	Exp. Nr.	EB1	EB2	ZK	PR	ER	KA	TF	BB
F-0-05-01	1	4	3	3	3	4	4	4	4
F-0-05-01	2	4	3	4	4	4	4	4	4
S-1-27-01	2	4	3	4	4	3	4	4	4
S-3-01-01	2	4	3	3	4	3	4	4	4
<i>Cluster VI</i>									
F-0-10-01	1	3	4	3	4	3	3	4	2
S-0-06-01	1	3	3	3	4	3	3	4	3
S-3-23-01	3	2	4	4	4	3	3	3	2
<i>Cluster VII</i>									
F-0-01-01	1	4	3	2	4	2	4	4	4
S-2-03-01	2	4	4	4	4	3	3	4	4
S-2-15-01	1	4	3	3	4	2	2	4	3
S-3-01-01	1	4	3	4	4	2	3	4	4
<i>Cluster VIII</i>									
D-0-02-01	1	4	4	3	3	3	4	4	2
D-0-12-01	1	2	4	4	3	4	4	4	2
D-1-06-01	2	4	4	4	4	3	4	4	2
D-1-06-01	3	4	4	4	4	4	2	4	2
D-2-04-01	2	4	4	3	4	3	4	4	3
D-2-15-01	1	3	4	4	3	4	4	4	2
D-2-15-01	2	4	4	3	4	4	4	4	3
D-2-42-01	2	4	4	4	4	4	4	4	2
F-0-04-01	2	4	4	4	4	4	4	4	1
F-0-08-01	2	4	4	3	4	4	4	4	2
S-1-18-01	2	4	4	3	4	4	4	4	2
S-2-13-01	1	4	4	4	4	4	3	4	3
<i>Cluster IX</i>									
D-1-29-01	1	1	2	4	1	2	1	2	1
D-3-07-01	1	2	3	4	1	2	2	3	2

Fortsetzung auf der nächsten Seite

ID	Exp. Nr.	EB1	EB2	ZK	PR	ER	KA	TF	BB
D-3-22-01	1	1	4	3	2	1	2	3	2
F-0-03-01	1	2	4	3	3	1	1	3	2
F-0-12-01	1	2	4	3	4	1	2	4	1
F-0-16-01	1	1	4	3	2	1	2	1	1
S-1-05-01	1	2	2	1	1	1	1	3	2
S-3-23-01	2	2	4	4	2	2	1	2	2
S-3-31-01	1	2	4	3	1	3	1	1	1

Tabelle C.1: Den Clustern I-IX zugeordnete Experimentiereinheiten. Innerhalb eines Clusters sind die Ratings der Facetten homogen bzw. ähnlich. EB1 Fachimmanente Einbettung, EB2 Lebensweltliche Einbettung, ZK Zielklarheit, PR Prozessreflexion, ER Ergebnisreflexion, KA Kognitiver Anspruch, TF Theoretische Fundierung und BB Begriffsbildung. *1=trifft zu, 2=trifft grösstenteils zu, 3=trifft teilweise zu, 4 trifft nicht zu.*

Lebenslauf: Johannes Börlin

Personalien

Name: Johannes Börlin
 Geburtsdatum: 12. Mai 1973
 Heimatort: Bubendorf BL

Ausbildung

09/2003 – 04/2006 Hochschule für Pädagogik und Soziale Arbeit beider Basel,
 Ausbildung zum Höheren Lehramt

11/1997 – 04/2002 Universität Basel, Studium der Physik, Nebenfach Mathe-
 matik,
 Diplomarbeit bei Prof. Dr. Ch. Bruder und Prof. Dr. W.
 Belzig,
 (Theorie der kondensierten Materie)

11/1994 – 11/1997 Höhere Technische Lehranstalt beider Basel, Studium zum
 Maschinenbauingenieur, Diplomarbeit bei Dr. P. von Böckh,
 (Thermische Energietechnik)

09/1989 – 09/1992 Lehre als Fahrrad- und Motorfahrradmechaniker

04/1984 – 07/1989 Realschule Basel

04/1980 – 04/1984 Primarschule Riehen

Berufliche Tätigkeiten

07/2007 – heute Wissenschaftlicher Mitarbeiter am Zentrum Naturwissen-
 schaften- und Technikdidaktik der PH-FHNW

08/2004 – 07/2007 Gymnasium Oberwil, Physik- und Mathematiklehrer

10/2003 – 10/2004 Hochschule für Technik und Architektur Luzern,
 Physikdozent

08/2002 – 01/2003 Freies Gymnasium Basel, Physik- und Mathematiklehrer

09/1992 – 11/1994 Fahrrad- und Motorfahrradmechaniker