

Chemie bilingual

Eine quasiexperimentelle Vergleichsstudie zur sachfachlichen, (fremd-)sprachlichen und motivationalen Wirksamkeit eines bilingualen Moduls im Fach Chemie

Von der Philosophischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Philosophie genehmigte Dissertation

vorgelegt von

Frank Hülten

Berichter*innen: Frau Univ.-Prof. Dr. phil. Frauke Intemann
Herr apl. Prof. Dr. rer. nat. Michael Schroeder

Tag der mündlichen Prüfung: 11. Oktober 2023

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	V
Tabellenverzeichnis.....	VII
Abkürzungsverzeichnis.....	IX
Vorwort.....	XI
1 Einleitung.....	1
2 Konzept und Umsetzung bilingualen Unterrichts in Deutschland	4
2.1 <i>Content and language integrated learning</i> oder bilingualer Unterricht?	5
2.2 Entwicklung unterschiedlicher Formate bilingualen Unterrichts	6
2.3 Bilinguale Module	8
2.4 Das Fach Chemie im bilingualen Unterricht	10
3 Bilingualer Chemieunterricht	13
3.1 Worum geht es im bilingualen Chemieunterricht?	13
3.2 Lernen im bilingualen Chemieunterricht.....	16
3.2.1 Content and language integrated learning: Integration von Sprach- und Inhaltslernen ..	17
3.2.2 Die sachfachliche Dimension	18
3.2.2.1 Die Verarbeitung und Speicherung von Wissen	18
3.2.2.2 Unterschiedliche Arten von Wissen und kognitiven Prozessen	21
3.2.2.3 Der Einfluss der Fremdsprache auf das sachfachliche Lernen.....	24
3.2.3 Die (fremd-)sprachliche Dimension	26
3.2.3.1 Spracherwerb im Zeichen von BICS und CALP.....	26
3.2.3.2 Das mentale Lexikon bilingualer Lerner	29
3.2.4 Die motivationale Dimension	32
3.2.4.1 Grundbegriffe	32
3.2.4.2 Motivationale Faktoren im bilingualen Unterricht.....	34
3.2.4.3 Motivation in den Naturwissenschaften: Gendereffekte	35
3.3 Gestaltung bilingualen Chemieunterrichts	36
3.3.1 Der Konstruktivismus und seine Konsequenzen für die Unterrichtsgestaltung	37
3.3.2 Diskursfunktionen einüben	41
3.3.3 Rollen von Fremd- und Muttersprache	44
3.3.4 Wortschatzarbeit	45
3.3.5 Planung bilingualen Unterrichts	46
4 Stand der Forschung: Der bilinguale Unterricht in den Naturwissenschaften	50
4.1 Der sachfachlich-kognitive Lernerfolg.....	50

Inhaltsverzeichnis

4.1.1	Erkenntnisse aus Immersionsprogrammen und bilingualen Zweigen	50
4.1.2	Erkenntnisse aus Modulen und anderen diskontinuierlichen Formen	51
4.1.2.1	Qualitative Studien	51
4.1.2.2	Quantitative Vergleichsstudien	53
4.1.3	Kognitive Vorteile durch bilingualen Unterricht	55
4.1.4	Faktoren, die den sachfachlichen Lernerfolg beeinflussen	57
4.2	Der (fremd-)sprachliche Lernerfolg im Bereich der Lexik	60
4.3	Auswirkung auf die Motivation	65
4.3.1	Motivation für den bilingualen Unterricht	65
4.3.2	Motivation für die mathematisch-naturwissenschaftlichen Sachfächer	67
4.3.3	Erste Ergebnisse zur Motivation im bilingualen Chemieunterricht	73
5	Design der Untersuchung: Forschungsfragen und Methodologie	76
5.1	Forschungsgegenstand und Rahmenbedingungen	76
5.2	Methodologischer Ansatz	78
5.2.1	Grundsätzliches	79
5.2.2	Forschungshypothesen und ihre Testung	81
5.2.3	Datenerfassung und Verarbeitung	82
5.2.4	Die Bildung von Interessensgruppen	87
5.3	Fokus der Untersuchung	88
5.3.1	Leistung im Sachfach	89
5.3.1.1	Hypothesen und ihre Testung	89
5.3.1.2	Theoretischer Hintergrund und Aufbau der Tests	91
5.3.1.3	Konstruktion von Testaufgaben	96
5.3.1.4	Auswertung der Ergebnisse	103
5.3.2	(Fremd-)sprachlicher Lernerfolg	105
5.3.2.1	Hypothesen und ihre Testungen	105
5.3.2.2	Theoretischer Hintergrund und Aufbau der Wortschatztests	106
5.3.2.3	Der deutsche Fachwortschatztest	111
5.3.2.4	Der englische Fachwortschatztest	115
5.3.2.5	Tests zum zusätzlich erworbenen englischen Wortschatz	118
5.3.3	Motivation	121
5.3.3.1	Hypothesen und ihre Testung	121
5.3.3.2	Theoretischer Hintergrund der Motivation: KIM	123
5.3.3.3	Aufbau der Fragebögen	127
5.3.3.4	Erstellung von Items	130

Inhaltsverzeichnis

5.3.3.5	Auswertung der Ergebnisse	132
5.4	Erfassung weiterer Kontrollvariablen.....	133
5.4.1	Kognitive Fähigkeiten	134
5.4.2	Lesekompetenz	136
5.4.3	Allgemeine Fremdsprachenkompetenz	137
5.4.4	Allgemeine Leistungsfähigkeit in Chemie	139
5.4.5	Interesse und Selbstkonzept.....	140
5.4.6	Demografische Merkmale	143
5.5	Das Modul: <i>Alkali and alkaline earth metals</i>	144
5.5.1	Themen und curriculare Einbettung	144
5.5.2	Methodische Gestaltung des Unterrichts	147
6	Die Pilotphase.....	149
6.1	Testgüte und Rahmenbedingungen.....	150
6.2	Chemie-Leistungstest	151
6.2.1	Objektivität	152
6.2.2	Validität.....	153
6.2.3	Reliabilität.....	154
6.3	Allgemeiner Leistungstest in Chemie.....	156
6.3.1	Objektivität	156
6.3.2	Validität.....	157
6.3.3	Reliabilität.....	157
6.4	Wortschatztests	159
6.4.1	Güte der Tests: Objektivität, Validität und Reliabilität.....	161
6.4.2	Deutscher Fachwortschatztest	161
6.4.3	Englischer Fachwortschatztest.....	162
6.4.4	Test zum zusätzlich erworbenen Fach- und Allgemeinwortschatz	163
6.5	Fragebögen	164
6.5.1	Aufbau der Fragebögen	165
6.5.2	Objektivität, Validität und Reliabilität.....	168
6.5.3	Das Schülerfazit.....	169
7	Hauptuntersuchung: Ergebnisse und Auswertung.....	170
7.1	Die Stichproben	171
7.1.1	Merkmale der Stichproben	171
7.1.2	Demografische Merkmale	172
7.1.3	Leistungsvermögen.....	173

Inhaltsverzeichnis

7.1.4	Interesse und Selbstkonzept.....	178
7.2	Die Interessensgruppen.....	179
7.3	Der sachfachliche Lernerfolg	183
7.3.1	Vergleich der Testergebnisse in Untersuchungs- und Kontrollgruppe	184
7.3.2	Der Einfluss verschiedener Variablen auf den sachfachlichen Zugewinn	187
7.3.3	Sachfachliche Leistungen im Bereich der <i>higher order thinking skills</i>	196
7.4	Der sprachliche Lernerfolg.....	199
7.4.1	Zum Lernerfolg im deutschen Fachwortschatz	199
7.4.2	Zum Lernerfolg im englischen Fachwortschatz	204
7.4.3	Zum fremdsprachlichen Lernerfolg im Fach- und Allgemeinwortschatz.....	208
7.5	Motivation	214
7.5.1	Vergleich von Untersuchungs- und Kontrollgruppe	215
7.5.2	Vergleich der Interessensgruppen	217
7.5.3	Vergleich von Jungen und Mädchen.....	224
7.5.4	Prädiktoren für die Motivation im bilingualen Unterricht.....	229
7.5.5	Entwicklung der Motivation bilingualer Schüler	230
8	Zusammenfassung der Ergebnisse und Fazit.....	232
	Literaturverzeichnis	246
	Anhang	267

Abbildungsverzeichnis

Abbildung 1: Das <i>Integrated Dynamic Model</i>	30
Abbildung 2: Das 4C-Modell.....	47
Abbildung 3: Die CLIL-Matrix.....	48
Abbildung 4: Die drei zentralen Fragekomplexe der Untersuchung	77
Abbildung 5: Fragebogenitem zur Ermittlung des Interessensschwerpunkts	88
Abbildung 6: Hypothesen zum sachfachlichen Lernerfolg.....	90
Abbildung 7: Operationalisierte Hypothesen zum sachfachlichen Lernerfolg	91
Abbildung 8: Das Kompetenzstrukturmodell des Projekts ESNaS	92
Abbildung 9: Anforderungsbereiche im Kompetenzmodell ESNaS.....	94
Abbildung 10: Merkmale der Testaufgaben zum Kompetenzbereich UMFW	95
Abbildung 11: Beispiel für eine geschlossene Aufgabe aus dem Chemieleistungstest	97
Abbildung 12: Beispiel für eine offene Aufgabe aus dem Chemieleistungstest.....	99
Abbildung 13: Codierleitfaden zu einer offenen Aufgabe aus dem Chemieleistungstest.....	100
Abbildung 14: Beispiel für eine halb-offene Aufgabe aus dem Chemieleistungstest.....	101
Abbildung 15: Beispiel für ein Itempaar aus den Paralleltests	102
Abbildung 16: Administration der Testvarianten A und B.....	104
Abbildung 17: Hypothesen zum (fremd-)sprachlichen Lernerfolg.....	105
Abbildung 18: Operationalisierte Hypothesen zum (fremd-)sprachlichen Lernerfolg	106
Abbildung 19: Multiple-Choice Item zum rezeptiven Fachwortschatz (Typ 1)	113
Abbildung 20: Multiple-Choice Item zum rezeptiven Wortschatz (Typ 2)	113
Abbildung 21: Ergänzungsaufgabe zum produktiven Fachwortschatz.....	114
Abbildung 22: Parallelitem aus dem englischen Fachwortschatztest (Sieblöffel).....	116
Abbildung 23: Hypothesen zur Motivation	122
Abbildung 24: Operationalisierte Hypothesen zur Motivation.....	123
Abbildung 25: Konstruktion der Items zur Subskala Interesse / Vergnügen	131
Abbildung 26: Umcodierung eines negativ gepolten Items.....	132
Abbildung 27: Items zur Erfassung der Variable Interesse am Fach Englisch	141
Abbildung 28: Items zur Erfassung der Variable Selbstkonzept im Fach Englisch.....	143
Abbildung 29: Ausschnitt aus der <i>concept map</i> der <i>alkali metals</i>	146
Abbildung 30: Weitere Konzepte, auf die im Modul zurückgegriffen wird	147
Abbildung 31: Diskursfunktionen zum Modul	147
Abbildung 32: Offene Aufgabe aus dem Chemieleistungstest	152

Abbildungsverzeichnis

Abbildung 33 Überarbeitung von Item 25 des deutschen Fachwortschatztests.....	162
Abbildung 34 Überarbeitung von Item 4 des englischen Fachwortschatztests	163
Abbildung 35 Überarbeitung von Item 6 des ZEFWT	164
Abbildung 36: Aufbau des Prefragebogens der CLIL-Gruppe	165
Abbildung 37: Aufbau des Postfragebogens der CLIL-Gruppe.....	166
Abbildung 38: Ersetzen der Skala Mehrarbeit durch die Skala Druck	167
Abbildung 39: Ergebnisse im Chemieleistungstest.....	184
Abbildung 40: Zugewinn und Vergessen im Chemieleistungstest.....	186
Abbildung 41: Abhängigkeit des Zugewinns von der Leistungsfähigkeit in Chemie	189
Abbildung 42: Abhängigkeit des Zugewinns von der Leistungsfähigkeit in Englisch.....	191
Abbildung 43: Ergebnisse im Anforderungsbereich II/III des Chemieleistungstest.....	197
Abbildung 44: Zugewinn und Vergessen im Anforderungsbereich II/III des CL-Tests	198
Abbildung 45: Ergebnisse im deutschen Fachwortschatztest	201
Abbildung 46: Zugewinn und Vergessen im deutschen Fachwortschatztest	202
Abbildung 47: Zugewinn und Vergessen im deutschen Fachwortschatztest	202
Abbildung 48: Ergebnisse der CLIL-Schüler im deutschen und englischen Fachwortschatztest ...	206
Abbildung 49: Zugewinn und Vergessen der CLIL-Schüler in den Fachwortschatztests	208
Abbildung 50: Ergebnisse im Test zum zusätzlich erworbenen Fachwortschatz	211
Abbildung 51: Ergebnisse im Test zum zusätzlich erworbenen Allgemeinwortschatz	211
Abbildung 52: Zugewinn und Vergessen beim insgesamt zu erwerbenden Wortschatz.....	214
Abbildung 53: Intrinsische Motivation in der CLIL- und in der nicht-CLIL Gruppe	215
Abbildung 54: Interesse am Modul in den vier Interessensgruppen.....	218
Abbildung 55: Nutzen des Moduls in den vier Interessensgruppen	220
Abbildung 56: Kompetenzerleben im Modul in den vier Interessensgruppen	221
Abbildung 57: Empfundener Druck im Modul in den vier Interessensgruppen.....	222
Abbildung 58: Interesse am Modul und Geschlecht.....	225
Abbildung 59: Nutzen des Moduls und Geschlecht.....	226
Abbildung 60: Kompetenzerleben im Modul und Geschlecht.....	227
Abbildung 61: Empfundener Druck im Modul und Geschlecht	228
Abbildung 62: Motivation für den bilingualen Unterricht vor und nach dem Modul	231

Tabellenverzeichnis

Tabelle 1: Kompetenzbereiche im Fach Chemie	15
Tabelle 2: Kognitive Lernziele nach Anderson und Krathwohl (2001)	23
Tabelle 3: Diskursfunktionen nach Vollmer (2011)	42
Tabelle 4: Diskursfunktionen und Operatoren für das Fach Chemie.....	43
Tabelle 5: Auswirkungen bilingualen Unterrichts auf unterschiedliche Sprachkompetenzen.....	60
Tabelle 6: Bildung von operationalisierter Hypothese und Nullhypothese	82
Tabelle 7: Erhebungsinstrumente der Studie	84
Tabelle 8: Zeitlicher Verlauf der Datenerhebung	85
Tabelle 9: Einteilung der Probanden in Interessensgruppen	87
Tabelle 10: Aufgabenmerkmale und ihr Auftreten im Chemieleistungstest	103
Tabelle 11: Inhalt und Umfang der Wortschatztests	108
Tabelle 12: Inhaltsfelder der Wortschatztests.....	110
Tabelle 13: Deutscher Fachwortschatz	112
Tabelle 14: Vergleich von deutschem und englischem Fachwortschatz	115
Tabelle 15: Fachbegriffe im deutschen und englischen Fachwortschatztest	117
Tabelle 16: Umfang des zusätzlich erworbenen englischen Wortschatzes	119
Tabelle 17: Fachbegriffe des Tests zum zusätzlich erworbenen Fachwortschatz	119
Tabelle 18: Vokabeln des Tests zum zusätzlich erworbenen Allgemeinwortschatz.....	120
Tabelle 19: Unterschiedliche Arten der Regulation nach Deci und Ryan.....	124
Tabelle 20: Aufbau des Prefragebogens der CLIL-Schüler	129
Tabelle 21: Aufbau des Postfragebogens der CLIL-Schüler.....	129
Tabelle 22: Entwicklung der Wortschatztests	160
Tabelle 23: Demografische Angaben zu Untersuchungs- und Kontrollgruppe	172
Tabelle 24: Deskriptive Statistik zum kognitiven Fähigkeitstest.....	174
Tabelle 25: Deskriptive Statistik zum Lesegeschwindigkeits- und -verständnistest	175
Tabelle 26: Deskriptive Statistik zum C-Test und zur Englischnote.....	176
Tabelle 27: Deskriptive Statistik zum allgemeinen Chemieleistungstest und zur Chemienote	177
Tabelle 28: Deskriptive Statistik zu Interesse und Selbstkonzept in Englisch und Chemie	178
Tabelle 29: Ausprägung verschiedener Kontrollvariablen in den Interessensgruppen	180
Tabelle 30: Deskriptive Statistiken zu den Chemieleistungstests	184
Tabelle 31: Deskriptive Statistik zu Zugewinn und Vergessen im Chemieleistungstest.....	186
Tabelle 32: Korrelationen zum Lernerfolg.....	187

Tabellenverzeichnis

Tabelle 33: Korrelationen zum unmittelbaren Zugewinn in der CLIL-Gruppe.....	188
Tabelle 34: Zugewinn im CL-Test in den drei ACLT-Leistungsgruppen.....	190
Tabelle 35: Zugewinn im CL-Test in den drei Leistungsgruppen des C-Tests	192
Tabelle 36: Zugewinn weniger sprachbegabter Schüler nach Leistungsstärke in Chemie	193
Tabelle 37: Zugewinn sprachbegabter Schüler differenziert nach Leistungsstärke in Chemie	194
Tabelle 38: β -Koeffizienten für die Prädiktoren des Lernerfolgs im Modul	195
Tabelle 39: Deskriptive Statistik zum Anforderungsbereich II/III des Chemieleistungstests.....	196
Tabelle 40: Zugewinn und Vergessen im Anforderungsbereich II/III des Chemieleistungstest	198
Tabelle 41: Lösungshäufigkeiten und Cronbachs Alpha zum deutschen Fachwortschatztest	200
Tabelle 42: Deskriptive Statistik zu den deutschen Fachwortschatztests	200
Tabelle 43 Korrelationen zum mittelfristigen Zugewinn im deutschen Fachwortschatz.....	203
Tabelle 44: Lösungshäufigkeiten und Cronbachs Alpha des englischen Fachwortschatztests	204
Tabelle 45: Ergebnisse der CLIL-Schüler im deutschen und im englischen Fachwortschatztest....	205
Tabelle 46: Zugewinn und Vergessen der CLIL-Schüler in den Fachwortschatztests	207
Tabelle 47: Lösungshäufigkeiten und Cronbachs Alpha im ZEFWT	209
Tabelle 48: Lösungshäufigkeiten der Items und Cronbachs Alpha des ZEAWT.....	210
Tabelle 49: Deskriptive Statistik zum ZEFWT und zum ZEAWT	210
Tabelle 50: Ergebnisse in Bezug auf den insgesamt zu erwerbenden englischen Fachwortschatz .	212
Tabelle 51: Ergebnisse in Bezug auf den insgesamt zu erwerbenden Allgemeinwortschatz.....	213
Tabelle 52: Zugewinn und Vergessen in Bezug auf den insgesamt zu erwerbenden Wortschatz	213
Tabelle 53: Intrinsische Motivation in der CLIL- und in der nicht-CLIL Gruppe.....	215
Tabelle 54: Korrelationen zur intrinsischen Motivation	217
Tabelle 55: Interesse am Modul in den vier Interessensgruppen	218
Tabelle 56: Nutzen des Moduls in den vier Interessensgruppen.....	219
Tabelle 57: Kompetenzerleben im Modul in den vier Interessensgruppen.....	221
Tabelle 58: Empfundener Druck im Modul in den vier Interessensgruppen	222
Tabelle 59: Intrinsische Motivation der nwo-Schüler	223
Tabelle 60: Intrinsische Motivation der fso-Schüler	224
Tabelle 61: Interesse am Modul und Geschlecht	224
Tabelle 62: Nutzen des Moduls und Geschlecht	226
Tabelle 63 Kompetenzerleben im Modul und Geschlecht	227
Tabelle 64: Empfundener Druck im Modul und Geschlecht	228
Tabelle 65: Korrelationen der Subskalen der intrinsischen Motivation.....	229
Tabelle 66: Motivation für den bilingualen Unterricht vor und nach dem Modul.....	231

Abkürzungsverzeichnis

AB	Arbeitsblatt
ACLT	Allgemeiner Chemieleistungstest
AFB	Anforderungsbereich
ANOVA	Analysis of Variance
BICS	Basic Interpersonal Communicative Skills
BSFU	bilingualer Sachfachunterricht
bso	beidseitig orientiert
bswo	beidseitig wenig orientiert
BU	Bilingualer Unterricht
CA	Cronbachs Alpha
CAE	Chemie auf Englisch
CALP	Cognitive-Academic Language proficiency
CLIL	Content and Language Integrated Learning
CLT	Chemieleistungstest
COMBIH	Competences and Motivatiopon in Bilingual Instruction - History
CR	Chemische Reaktion und Energieumsätze
CT	C-Test
DENOCs	Development of North Rhine-Westfalian CLIL-Students
DESI	Deutsch-Englisch-Schülerleistungen-International
DEZIBEL	Deutsch-Englische Züge in Berlin
DFWT	Deutscher Fachwortschatztest
DMO	Druck im Modul
EDBU	Erwartung Druck im bilingualen Unterricht
EFW65	Gesamter englischer Fachwortschatz
EFWT	Englischer Fachwortschatz
EIBU	Erwartung Interesse am bilingualen Unterricht
EKBU	Erwartung Kompetenzerleben im bilingualen Unterricht
ENBU	Erwartung Nutzen des bilingualen Unterrichts
ESNaS	Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I
FB	Fragebogen
FICH	Fachinteresse Chemie
FIEN	Fachinteresse Englisch
FMKS	Verein für frühe Mehrsprachigkeit an Kitas
fso	fremdsprachlich orientiert
HA	Hausaufgabe
IDM	Integrated Dynamic Model
IM	Immersion
IMO	Interesse am Modul
KESS	Kompetenzen und Einstellungen von Schülerinnen und Schülern
KFT	Kognitiver Fähigkeitstest
KG	Kontrollgruppe
KIM	Kurzskala Intrinsische Motivation
KMK	Kultusministerkonferenz
KMO	Kompetenzerleben im Modul
L1	Erstsprache, Muttersprache
L2	Zweitsprache, Fremdsprache
LAU	Lernausgangslage und Lernentwicklung

Abkürzungserzeichnis

LGN	Lesegenauigkeit
LGS	Lesegeschwindigkeit
LGVT 5-12+	Lesegeschwindigkeits- und Leseverständnistest
LH	Lösungshäufigkeit
LV	Leseverständnis
M	arithmetisches Mittel
MC	Multiple-Choice
MD	Median
MINT	Mathematik, Informatik, Naturwissenschaften und Technik
N	Zahl der Probanden einer Stichprobe (number)
NMO	Nutzen des Moduls
nwo	naturwissenschaftlich orientiert
OECD	Organisation for Economic Co-operation and Development
PBA	Phänomene erklären, Bezüge finden und Abstraktionen nutzen
PISA	Programme for International Student Assessment
PSE	Periodensystem der Elemente
SD	Standardabweichung (standard deviation)
SE	Struktur und Eigenschaften, Standardfehler (standard error)
SKCH	Selbstkonzept Chemie
SKEN	Selbstkonzept Englisch
SPSS	Statistical Package for the Social Sciences
TB	Tafelbild
TIMSS	Trends in International Mathematics and Science Study
TOEFL	Test of English as a Foreign Language
TS	Trennschärfe
UG	Untersuchungsgruppe
UMFW	Umgang mit Fachwissen
ZEAWT	Test zum zusätzlich erworbenen Allgemeinwortschatz
ZEFWT	Test zum zusätzlich erworbenen englischen Fachwortschatz
ZPD	Zone of proximal development
ZU1CLT	Zugewinn im Chemieleistungstest

Vorwort

Die vorliegende Arbeit entstand in den Jahren 2017 bis 2022 im Rahmen meiner Tätigkeit als Lehrer am Bischöflichen Gymnasium St. Ursula in Geilenkirchen; sie wurde im Wintersemester 2022 als Dissertation im Fachbereich Fremdsprachendidaktik der Philosophischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule (RWTH) Aachen eingereicht.

Mein besonderer Dank gilt meiner Betreuerin, Frau Prof. Dr. Frauke Intemann für ihre fachlich kompetente, unkomplizierte und stets freundliche Unterstützung. Außerdem bedanke ich mich bei Prof. Dr. Michael Schröder für die Bereitschaft, das Zweitgutachten zu übernehmen.

Bedanken möchte ich mich auch bei den studentischen Hilfskräften Herrn Tjarko Rahlf und Frau Evelyn Petre für ihren Einsatz bei der Datenerfassung, sowie bei der Schulleitung des Bischöflichen Gymnasiums St. Ursula für die Erlaubnis zur Durchführung der Studie. Meinem Bruder Helmut danke ich für die formalen Verbesserungsvorschläge. Zuletzt gilt mein besonderer Dank den Schülerinnen und Schülern, die an der Untersuchung teilgenommen haben. Ohne ihre Bereitschaft zur Mitarbeit, wäre diese Arbeit nicht zustande gekommen.

Millen, im Dezember 2022

Frank Hülten

Formaler Hinweis

Aufgrund der besseren Lesbarkeit wird auf eine geschlechtsneutrale Differenzierung verzichtet. Entsprechende Begriffe gelten im Sinne der Gleichbehandlung grundsätzlich für alle Geschlechter. Die verkürzte Sprachform hat redaktionelle Gründe und impliziert keine Wertung.

1 Einleitung

Zwar wird bilingualer Chemieunterricht (mit Englisch als Fremdsprache) seit Mitte der 1990er Jahre punktuell praktiziert, er hat aber bis heute keine flächendeckende Verbreitung erfahren und unterliegt gegenüber vielen Lehrerinnen und Lehrern sowie Schülerinnen und Schülern [...] immer noch einem regelrechten Rechtfertigungszwang (Bohrmann-Linde, 2012, S. 183).

Das Fach Chemie zählt nach wie vor zu den weniger verbreiteten Sachfächern des bilingualen Unterrichts, der sich ansonsten nach wie vor großer Beliebtheit erfreut. Die Gründe hierfür sind vielfältig und unter anderem auf den Mangel empirisch abgesicherter Daten zur Wirksamkeit dieser Unterrichtsform im Fach Chemie zurückzuführen. Die vorliegende Studie soll dazu beitragen, diesen Mangel zu beheben. Dazu werden die sachfachlichen und fremdsprachlichen Lernerfolge in einem bilingualen Chemiemodul sowie die Motivation der Schüler untersucht.

Sachfachlicher Lernerfolg

In der Forschung zum bilingualen Unterricht wurde der sachfachliche Lernerfolg lange Zeit vernachlässigt. Immer wieder wurde - nicht nur von Vertretern der Sachfachdidaktiken - beklagt, dass nicht hinreichend geklärt ist, welche sachfachlichen Einbußen durch den Einsatz einer nur unvollkommen beherrschten Fremdsprache zu erwarten sind (Zydati, 2002, S. 58).

Mittlerweile liegen jedoch zahlreiche Forschungsergebnisse zu den verschiedenen Sachfächern vor. Dalton-Puffer (2008, S. 143) fasste die europäischen Forschungsergebnisse zum sachfachlichen Lernen zusammen und kommt zu dem Schluss, dass die Mehrzahl der Studien zeigen, dass die Leistungen der bilingualen Schüler denen der monolingual in der Muttersprache unterrichteten Schüler vergleichbar sind. Für die Naturwissenschaften sind die Ergebnisse jedoch alles andere als eindeutig. Auf der einen Seite gibt es zahlreiche Studien, die zeigen, dass die sachfachlichen Leistungen bilingual und monolingual unterrichteter Schüler durchaus vergleichbar sind (Koch 2005, Rittersbacher 2006, Hegerfeldt 2006, Kondring und Ewig 2006, Luprichova 2013, Haagen-Schützenhöfer 2011, Rodenhauser 2016) oder, dass die bilingualen Schüler den Regelschülern sogar überlegen sind (Bonnet 2004, Osterhage 2007). Auf der anderen Seite jedoch existieren Studien, die auf sachfachliche Defizite der bilingualen Schüler im Vergleich zu den Regelschülern hindeuten (z.B. Hartmannsgruber 2014, Piesche 2016, Lo & Lo 2014, Jäppinen 2005). Für das Fach Chemie ist die Datenlage allerdings besonders schwach. Studien zum sachfachlichen Lernerfolg sind oft qualitativer Natur und beziehen sich auf wenige Probanden (Bonnet 2004, Koch 2005, Rittersbacher

2006, Hegerfeldt 2006, Hülten 2014), großangelegte quantitative Untersuchungen fehlen. Dabei nehmen die bislang vorliegenden Untersuchungen in den Naturwissenschaften vor allem den Wissenserwerb im Sachfach in den Blick und weniger den Umgang mit diesem Wissen auf der Basis höherer Denkopoperationen (vgl. z.B. Hartmannsgruber 2004, Piesche 2016, Haagen-Schützenhöfer 2011, Rodenhauser 2016). Es gibt jedoch Grund zur Annahme, dass solche *higher order thinking skills* im bilingualen Unterricht in bestimmten Fächern sogar besser ausgebildet werden können als im Regelunterricht (Osterhage 2007, Zydati 2012, Heine 2010a, Vollmer 2006). Vorliegende Arbeit nimmt die Chemieleistungen deshalb differenziert nach ihren kognitiven Anforderungen in den Blick, um zu prüfen, ob eine solche Überlegenheit der bilingualen Schüler vorliegt, oder, ob gerade kognitiv anspruchsvollere Aufgaben in einem Unterricht, der überwiegend in der Fremdsprache abläuft, nicht bewältigt werden können. Unter dem Slogan „Bili für alle“ wurden in den vergangenen Jahren weitere Formate des bilingualen Unterrichts entwickelt, die es ermöglichen sollen, möglichst vielen Schülern ein bilinguales Angebot anzubieten (siehe Abschnitt 2.2). Zu diesen Formaten gehören auch die bilingualen Module. Ob sich jedoch bilingualer Unterricht für alle Schüler eignet, insbesondere mit Blick auf das Sachfach Chemie, ist nicht vollständig geklärt. Deshalb werden im Rahmen der vorliegenden Studie auch personenbezogene Faktoren untersucht, die den sachfachlichen Lernerfolg fördern beziehungsweise beeinträchtigen können. Hierzu wurden in der Vergangenheit neben dem Leistungsvermögen im Sachfach und in der Fremdsprache auch motivationale Aspekte wie Interesse oder Selbstkonzept in den Blick genommen (vgl. Zydati 2012, Piesche 2016, Marsh et al. 2000, Farrel 2011). Die Ergebnisse sind jedoch durchaus widersprüchlich, beziehungsweise für das Fach Chemie nicht existent (siehe Abschnitt 4.1.4).

(Fremd-) sprachlicher Lernerfolg: Wortschatz

Die Auswirkungen bilingualen Unterrichts auf die Fremdsprachenkompetenzen wurde in der Vergangenheit für viele Sachfächer intensiv erforscht (siehe Abschnitt 4.2). Für das Fach Chemie liegen dazu jedoch kaum Daten vor. Außerdem mangelt es an Untersuchungen, die den fremdsprachlichen Erfolg bilingualer Kurzformen fokussieren, wie etwa den von Modulen. Ob fremdsprachliche Erfolge auch hier zu erwarten und wie nachhaltig diese gegebenenfalls sind, ist noch wenig erforscht¹. Zumindest die Annahme, dass im Chemieunterricht ausschließlich fachsprachlich kommuniziert wird und dieser deshalb keinen Beitrag zur Förderung der

1 Eine Ausnahme ist Rodenhausers (2016) Untersuchung zur Lesefertigkeit bei Schülern, die an einem bilingualen Labortag im Fach Biologie teilgenommen hatten.

Fremdsprache leisten kann, wurde von Bonnet (2004) in der Vergangenheit widerlegt. Dazu beschrieb er anhand konkreter Unterrichtssituationen den Anteil der Allgemeinsprache im modernen Chemieunterricht. Vorliegende Analyse soll die Anteile von Fach- und Allgemeinsprache für ein bilinguals Chemiemodul erstmals quantifiziert und die tatsächlichen Lernzuwächse der Schüler ermitteln. Dazu wird der fremdsprachliche Wortschatz betrachtet, dem, nach Read (2011, S. 1), eine zentrale Bedeutung in der Fremdsprachenforschung zukommt. Das liegt daran, dass Wörter die Grundbausteine von Sprache sind, auf deren Grundlage sich die Bedeutungen größerer Strukturen wie Satz, Abschnitt oder Text ergeben. Lernerfolge im Bereich Lexik wirken sich deshalb auch auf andere Kompetenzbereiche aus. Rumlich (2016, S. 210) stellt dazu fest:

Furthermore, vocabulary skills are also most closely related to the overall proficiency measures, underlining their being a pivotal asset to language proficiency and performance in general (ebd., S. 210).

Der für den bilingualen Unterricht geforderte Aufbau einer *doppelten Fachliteralität* ist bislang wenig erforscht. Diehr (2016, S. 61f.) stellt dazu fest, dass „Verfahren der Leistungsmessung, die die Zweisprachigkeit berücksichtigen gänzlich fehlen“. Der von Diehr (2016, S. 63f.) gesehene Forschungsbedarf in diesem Bereich erklärt sich aus der Tatsache, dass der bilinguale Unterricht lange Zeit als ein monolingual in der Fremdsprache geführter Unterricht betrachtet wurde (vgl. z.B. Bonnet 2007, Hallet 2002, Niemeier 2005). Die vorliegende Studie dient dazu, diese Lücke zu schließen, indem neben dem englischen auch der Erwerb des deutschen Fachwortschatzes betrachtet wird.

Motivation

Da sich bei PISA 2018 gegenüber PISA 2015 auch die naturwissenschaftliche Kompetenz der Schüler verringert hat, werden immer mehr Stimmen laut, dass Maßnahmen zur Förderung dieser Kompetenz "im Sinne einer mehrdimensionalen Bildung auch motivationale Orientierung und Einstellungen als wichtige Bildungsziele adressieren" (Schiepe-Tiska, Rönnebeck, & Neumann, 2019, S. 238). Da Jugendliche, die ein Zutrauen in ihre eigenen Fähigkeiten haben, selbstbewusster mit naturwissenschaftlichen Fragestellungen umgehen und Lernstrategien effektiver nutzen, fordern die Autoren als wichtige Ziele naturwissenschaftlicher Grundbildung auch "das Erleben von Freude und Interesse an Naturwissenschaften oder das Herausbilden positiver naturwissenschaftlicher Selbstbilder" (Schiepe-Tiska, Simm, & Schmidtner, 2016, S. 100). Auch Becker (2012, S. 359) bemerkt, dass "Emotionen gleichwohl Voraussetzungen für Lernerfolg und elementare Lernmechanismen" sind und stellt fest, dass eine allzu enge "Fixierung auf das Kognitive" in der

Vergangenheit vielleicht "kontraproduktiv" war. Genau hier eröffnet sich eine Chance für den bilingualen Unterricht, wenn es gelingt, positive Emotionen aus dem Gebrauch der Fremdsprache nutzbar zu machen. Umso überraschender ist es, dass Breidbach und Viehbrock 2012 feststellen, dass Motivation in der deutschsprachigen CLIL-Forschung zu den nur am Rande untersuchten Konzepten gehört (Breidbach & Viehbrock, 2012, S. 11). In den letzten Jahren haben jedoch zahlreiche Untersuchungen in diesem Bereich dazu beigetragen, affektiv-motivationale Auswirkungen des bilingualen Unterrichts in den Naturwissenschaften besser zu verstehen. Während für die Fächer Physik (Piesche 2016), Biologie (Hartmannsgruber 2014, Scheersoi 2008, Rodenhauser 2017, Ohlberger 2019) und Mathematik (Verriere 2014, Lipski-Buchholz 2019) erste empirische Studien vorliegen, fehlen vergleichbare Untersuchungen für das Fach Chemie bislang. Von besonderem Interesse für die Naturwissenschaften ist dabei auch die Frage, ob es gelingt, Mädchen, die sich laut zahlreicher Studien (z.B. PISA 2015) weniger für Naturwissenschaften und mehr für Sprachen interessieren, für Chemie zu begeistern (siehe Abschnitt 4.3).

Gliederung

Das erste Kapitel dient dazu, die Organisationsform des bilingualen Moduls von anderen Varianten des bilingualen Unterrichts abzugrenzen und die besonderen Bedingungen des Faches Chemie zu bestimmen. Im Anschluss wird zunächst die Zielsetzung des bilingualen Chemieunterrichts umrissen (Abschnitt 2.1) und der lerntheoretische Hintergrund vorgestellt, aufgrund dessen sich der Erfolg des bilingualen Chemieunterrichts verstehen und bewerten lässt. Dabei werden sowohl die sachfachliche als auch die fremdsprachliche und die motivationale Dimension des Lernens im bilingualen Unterricht beleuchtet (Abschnitt 3.2). Wie bilingualer Chemieunterricht gestaltet sein muss, um die anvisierten Lernerfolge zu ermöglichen, wird in Abschnitt 3.3 erörtert. Bevor im empirischen Teil zunächst das Design der Untersuchung (Kapitel 5) und anschließend die Ergebnisse der Pilotphase (Kapitel 6) sowie der Hauptuntersuchung (Kapitel 7) vorgestellt werden, erfolgt in Kapitel 4 eine ausführliche Darstellung des aktuellen Forschungsstandes. Im letzten Kapitel (Kapitel 8) findet sich eine Zusammenfassung und Diskussion der Ergebnisse vorliegender Studie.

2 Konzept und Umsetzung bilingualen Unterrichts in Deutschland

Dieses Kapitel dient dazu, das Konzept des bilingualen Chemie-Unterrichts in Modulform als unterrichtliches Setting der empirischen Untersuchung vorzustellen. Dazu erscheint es sinnvoll,

zunächst die Begriffe *bilingualer Unterricht* und *CLIL* voneinander abzugrenzen (Abschnitt 2.1), die unterschiedlichen Ausprägungen bilingualer Unterrichtsformen vorzustellen (Abschnitt 2.2) und die Besonderheiten des modularen Charakters (Abschnitt 2.3) sowie des Sachfachs Chemie (Abschnitt 2.4) herauszustellen.

2.1 *Content and language integrated learning* oder bilingualer Unterricht?

Wenn von bilingualem Unterricht im deutschen Kontext die Rede ist, wird häufig die treffendere Bezeichnung CLIL (*Content and language integrated learning*) gewählt². Der Begriff wurde gegen Ende des 20. Jahrhunderts geprägt und bezieht sich auf bilinguale Unterrichtsangebote, die vor dem Hintergrund eines stärker zusammenwachsenden Europas entstanden. So bildeten sich Unterrichtsformen, in denen Lehren und Lernen in einer weiteren Sprache stattfindet.

Dabei stellt der duale Fokus auf Inhalt und Sprache, die integriert unterrichtet werden, das allgemeine didaktische Prinzip von CLIL (EMILE)³ dar (Rumlich, 2016, S. 49).

Es werden also stets sowohl sprachliche als auch inhaltliche Ziele verfolgt und durch gezielte unterrichtsmethodische Maßnahmen vorbereitet (Coyle, Hood, & Marsh, 2012, S. 3):

...[A]chieving this twofold aim calls for the development of a special approach to teaching in that the non-language subject is not taught *in* a foreign language but *with* and *through* a foreign language (Eurydice, 2006, S. 8).

Die Umsetzung von CLIL je nach Schulform und Land kann jedoch sehr unterschiedlich sein und hängt nicht nur von der Bedeutung, der Sprache bzw. Inhalt jeweils zukommt ab, sondern insbesondere vom jeweiligen Schulsystem (Rumlich, 2016, S. 49). Deshalb finden sich heute eine große Vielzahl verschiedener Implementierungsformen von CLIL, was dazu geführt hat, dass die Bezeichnung oft als Oberbegriff verwendet wird, der in Wirklichkeit eine Vielzahl höchst unterschiedlicher Varianten umfasst (Marsh D., 2000, S. 3). Aus den verschiedenen organisatorischen Kontexten lassen sich jedoch gemeinsame Merkmale isolieren. Dazu zählen nach Dalton-Puffer und Smit (2013, S. 546):

2 Der Begriff bilingualer Unterricht (BU) wurde in der Vergangenheit häufig kritisiert, weil er ursprünglich die Fähigkeit bezeichnet, zwei Sprachen nahezu gleich gut zu beherrschen, wie es etwa in zweisprachigen Familien der Fall ist. Damit weckt der Begriff bilingualer Unterricht die Hoffnung, ein vergleichbares Maß an sprachlicher Kompetenz im schulischen Kontext zu erwerben (Weller, 1993).

3 "Enseignement d'une Matière par l'Intégration d'une Langue Etrangère" (Marsh D., 2002, S. 58).

1. Der Gebrauch einer Fremd- oder Verkehrssprache (*lingua franca*), der die Schüler vornehmlich im Klassenraum und nicht in der Gesellschaft, in der sie leben, begegnen, wie bei einer Zweitsprache.
2. CLIL-Unterricht beginnt erst, wenn die Schüler ihre Muttersprache in Wort und Schrift beherrschen.
3. CLIL-Lehrer sind in der Regel keine Muttersprachler, sondern unterrichten das jeweilige Sachfach.
4. CLIL-Stunden sind im Stundenplan als Stunden in einem bestimmten Sachfach ausgewiesen. Der reguläre Unterricht an der Schule läuft parallel weiter.

Auch in Deutschland haben sich verschiedene Varianten des bilingualen Unterrichts entwickelt (vgl. Abschnitt 2.2). Dabei unterscheiden sich die deutschen Varianten von CLIL von den in vielen anderen Ländern praktizierten Formen jedoch in einem entscheidenden Punkt: Der CLIL-Lehrer verfügt nämlich in der Regel sowohl über die Fakultas im Sachfach als auch über die Fakultas in der Fremdsprache. Er ist also gleichzeitig Sachfach- und Fremdsprachenlehrer. Damit ist er in der Lage, den Unterricht sowohl in der Mutter- als auch in der Fremdsprache zu gestalten. Bei der Verwendung der L2 gelingt es ihm deshalb, eher die sprachlichen Defizite der Lerner zu erkennen und die fremdsprachliche Kommunikation aufrecht zu erhalten. Diese besondere Situation hat jedoch in der Vergangenheit dazu beigetragen, dass der Unterricht eben nicht wirklich in zwei Sprachen – also bilingual - geführt wurde, sondern fast ausschließlich in der Fremdsprache - also monolingual fremdsprachig. Da jedoch die doppelte Sachfachliteralität das erklärte Ziel der deutschen Variante von CLIL ist (Vollmer H. , 2005, S. 34), würde der monolingual in der L2 geführte Unterricht der Zielsetzung von CLIL nicht gerecht. Content and language integrated learning im deutschen Kontext bedeutet nämlich, dass nicht nur Inhalt und Fremdsprache, sondern Inhalt, Fremdsprache und Muttersprache integriert unterrichtet werden. Deshalb finden nicht wenige Autoren, dass doch *bilingualer Unterricht* - und nicht CLIL - die zutreffendste Bezeichnung für dieses Lernangebot sei (Diehr, 2012, S. 22).

2.2 Entwicklung unterschiedlicher Formate bilingualen Unterrichts

Der Beginn des bilingualen Unterrichts in Deutschland wird häufig auf das Jahr 1963 datiert, als die ersten bilingualen Programme an deutschen Schulen eingeführt wurden. Sie waren das unmittelbare Resultat des deutsch-französischen Freundschaftsvertrages vom 20. Januar dieses Jahres und zunächst nur auf wenige Schulen in unmittelbarer Grenznähe beschränkt. Gemäß dem Vertrag sollte die Sprache des Nachbarlandes besonders gefördert werden. Anfangs stellen die ausschließlich französischsprachigen Programme noch eine Besonderheit dar und bis 1987 gab es lediglich 25

Schulen, die ein entsprechendes Angebot aufweisen konnten (Wolff D. , 2017). Erst mit Beginn der 70er Jahre entstanden dann die ersten deutsch-englischen Projekte und 1971 richtete das Helene-Lange-Gymnasium in Hamburg den ersten deutsch-englischen Zweig ein (Bonnet, 2004, S. 35). In den 1990er Jahren - mit der Ankündigung des europäischen Binnenmarktes - begannen sich die Verhältnisse zu verschieben und Englisch löste innerhalb weniger Jahre Französisch als häufigste Sprache des bilingualen Unterrichts ab. In nur vier Jahren verdoppelte sich die Zahl der bilingualen Züge und im Jahr 1995/96 gab es mehr als 220 bilinguale Züge in Deutschland (Wolff D. , 1997, S. 171). Ende der 1990er Jahre zählte man bereits über 300 solcher Programme, darunter 80 deutsch-französische und 238 deutsch-englische Züge. Zu den weiteren Sprachen des bilingualen Unterrichts zählten Griechisch, Holländisch, Italienisch, Portugiesisch, Russisch, Spanisch, Tschechisch und Türkisch (Bonnet, 2004, S. 36). Für das Jahr 2013 zählt Rumlich (2016, S. 56) bereits über 1500 Schulen mit bilingualen Angeboten. Entscheidend für den explosionsartigen Anstieg der Zahl bilingualer Angebote dürften dabei auch die entsprechenden Erlasse der Kultusministerkonferenz sein, die, den Forderungen des Europarates folgend, CLIL-Programme immer wieder als wichtigen Beitrag zur Entwicklung fremdsprachlicher Fertigkeiten und interkultureller Kompetenzen gefördert und unterstützt haben. Damit stellt bilingualer Unterricht heute kein regionales Phänomen dar, wenngleich fast ein Viertel der bilingualen Züge an Schulen in Nordrhein-Westfalen eingerichtet wurden (Rumlich, 2016, S. 56). Die Forderung, bilingualen Unterricht allen Schülern zugänglich zu machen, hat auch dazu geführt, dass eine Vielzahl weiterer Formate entwickelt wurden, die auch an anderen Schulformen implementiert wurden. Für das Jahr 2014 bedeutet dies, dass in Nordrhein-Westfalen 26,3 % aller weiterführenden Schulen, mit Ausnahme der Hauptschulen, irgendeine Form bilingualen Unterrichts anbieten (Rumlich, 2016, S. 59). Auch für den Primar- und Vorschulbereich wurden im Laufe der Jahre eine Vielzahl bilingualer Programme entwickelt. Der Verein für frühe Mehrsprachigkeit an Kitas und Grundschulen (FMKS) verzeichnet für das Jahr 2014 insgesamt 1035 bilinguale Angebote an Kindertagesstätten und 287 Angebote an Grundschulen (FMKS, 2022). An deutschen Schulen finden sich heute im Wesentlichen drei Ausprägungen bilingualen Unterrichts, die sich vor allem in den Rahmenbedingungen und im zeitlichen Umfang unterscheiden. Zu diesen gehören die klassischen bilingualen Züge, der bilinguale Sachfachunterricht und bilinguale Module (KMK, 2013, S. 8). Die klassische Variante des bilingualen Unterrichts - die so genannten bilingualen Züge - finden sich heute vornehmlich an Gymnasien. Die Schüler werden dabei in der Regel in maximal drei Sachfächern beginnend ab Klasse 7 in der Fremdsprache unterrichtet⁴. Die Fächer entstammen

4 Manchmal tritt bereits ab Klasse 6 Sport als bilinguales Sachfach dazu (Bonnet, 2004, S. 31).

dabei typischerweise dem gesellschaftswissenschaftlichen Bereich (z.B. Erdkunde, Geschichte oder Politik), aber auch Naturwissenschaften (i.d.R. Biologie) und Kunst, Sport oder Musik befinden sich heute immer häufiger unter den Sachfächern (KMK, 2013, S. 14). Die Schüler werden in separaten Klassen unterrichtet, wobei die Wochenstundenzahl der bilingualen Sachfächer meist von zwei auf drei Stunden erhöht wird. Der reguläre Fremdsprachenunterricht wird parallel weitergeführt. Ab Klasse 8 tritt dann i.d.R. ein zweites Sachfach hinzu und später ein weiteres. Dabei werden meist zwei oder drei Sachfächer, jedoch niemals mehr als drei Sachfächer gleichzeitig bilingual unterrichtet. In den meisten Fällen ist der BU-Lehrer sowohl Sachfach- als auch Fremdsprachenlehrer. Dieser sogenannte integrative Typ des bilingualen Unterrichts (Krechel, 2013, S. 75) ermöglicht eine integrierte Entwicklung fremdsprachlicher und sachfachlicher Kompetenzen und wird in der Regel empfohlen (KMK, 2013, S. 16). Beim additiven Typ werden zwei von insgesamt drei Stunden von einem Zielsprachen-Muttersprachler und eine weitere von einem deutschsprachigen Lehrer unterrichtet. Zur Vorbereitung auf die erhöhten sprachlichen Anforderungen erhalten die Schüler in den Klassen 5 und 6 zwei Wochenstunden mehr Fremdsprachenunterricht, also meist statt fünf dann sieben Stunden. In der Oberstufe wird üblicherweise ein bilinguals Sachfach als Grundkurs weitergeführt und die Fremdsprache zusätzlich als Leistungskurs belegt. Die Teilnahme am bilingualen Unterricht wird auf dem Zeugnis vermerkt.

2.3 Bilinguale Module

Die fortschreitende Globalisierung und ein immer weiter zusammenwachsendes Europa bescherten bilingualet Unterricht mit Beginn der 1990er Jahre einen regelrechten Boom. Das Bemühen, möglichst vielen Schülern einen Zugang zu ermöglichen, führte zur Diversifizierung der Angebote (Breidbach & Viebrock, 2012). Die Fremdsprache wird dabei meist als Arbeitssprache im Sachfachunterricht eingesetzt. Dieser kann kontinuierlich, d.h. in der Regel über die Dauer von einem Jahr, bzw. diskontinuierlich, also z.B. in Form kurzer bilingualer Module stattfinden. Beim kontinuierlichen Arbeitssprachenunterricht wird der bilinguale Unterricht durchgehend in der Mittel- oder Oberstufe für mindestens ein Jahr erteilt. Dabei entfällt in der Regel der erweiterte Fremdsprachenunterricht in Klasse 5 und 6, die Fremdsprache wird jedoch meist in vollem bzw. in reduziertem Stundenumfang parallel weitergeführt⁵. Die Ziele dieser Form des bilingualen Unterrichts werden von der KMK wie folgt formuliert:

5 Zur Ausprägung des bilingualen Sachfachunterrichts an Gesamtschulen siehe Hallet & Königs (2013, S. 77 f.).

Der Einsatz der Fremdsprachen als Arbeitssprachen intensiviert fachliches und sprachliches Lernen und dient der Vorbereitung auf die zunehmende Internationalisierung in Ausbildung, Studium und Berufsleben. (KMK, 2011)

Die diskontinuierlichen Formen des Arbeitssprachenunterrichts erfolgen in sehr unterschiedlichen Organisationsformen. Dazu gehören der vorübergehende Einsatz im regulären Sachfach- oder Fremdsprachenunterricht sowie in Arbeitsgemeinschaften. Auch Projekte eignen sich etwa im Rahmen von Austauschmaßnahmen, internationalen Begegnungen oder Betriebspraktika (Krechel, 2013, S. 79). Obwohl für den bilingualen Unterricht in Modulform keine offiziellen Zahlen vorliegen, so beobachtet Krechel (Krechel, 2013, S. 78) in den letzten Jahren in fast allen Bundesländern einen kontinuierlichen Zuwachs für die einzelnen Schulformen. Bilinguale Module dienen oft dazu, den kontinuierlichen Arbeitssprachenunterricht vorzubereiten. Gerade die zeitliche Begrenztheit und die Möglichkeit, fremd- und muttersprachliche Anteile im Hinblick auf die jeweilige Lerngruppe frei zu wählen, haben dazu geführt, dass auch zahlreiche neue Sachfächer für den bilingualen Unterricht erschlossen werden konnten. Zu diesen gehören heute alle nichtsprachlichen Fächer, also auch Mathematik, die Naturwissenschaften, Sport und die Fächer des musisch-künstlerischen Aufgabenbereichs. Krechel (2013) weist in diesem Zusammenhang insbesondere auf das innovative Potenzial der Module hin:

Durch das Angebot an bilingualen Modulen werden neue Wege beschritten, die das bilinguale Lernen und Arbeiten in Zukunft wesentlich erweitern und sich der jeweiligen Lerngruppe flexibel anpassen lassen. Sie sind nicht an bestimmte Schuljahre und Fächer gebunden und für Lernende und Lehrkräfte ein Experimentierfeld, in dem sie neue Möglichkeiten des Lernens und des unterrichtlichen Arbeitens entdecken (ebd., S. 78).

Dabei muss der Einsatz bilingualer Module jedoch auch den weniger sprachinteressierten Lernern plausibel sein. Er bietet sich deshalb besonders bei solchen Themen an, die einen landeskundlichen Bezug zum Partnerland haben oder deren fremdsprachliche Erarbeitung große berufliche Bedeutung hat. Bilinguale Module zielen vor allem darauf ab, möglichst allen Schülern Erfahrungen im Gebrauch der Fremdsprache als Arbeitssprache zu ermöglichen. Dabei geht es nach Krechel (2013, S. 79) im Wesentlichen um die "Ausweitung fremdsprachlichen Handelns auf andere fachbezogene Anwendungsbereiche" sowie "die Erweiterung der fremdsprachlichen Kompetenz bei der Bewältigung fachspezifischer Situationen und Anforderungen" insbesondere im Hinblick auf eine verbesserte Berufsvorbereitung.

2.4 Das Fach Chemie im bilingualen Unterricht

Aktuelle Angebote

Während der bilinguale Unterricht boomt und die Zahl der Angebote immer weiter zugenommen hat, gehört Chemie nach wie vor zu den Fächern, die an deutschen Schulen eher selten vertreten sind. So verzeichnet die KMK in ihrer Länderabfrage von 2013 zwar einen Zuwachs bei der Nutzung der Fremdsprache in den MINT-Fächern seit 2006 (KMK, 2013, S. 10), Chemie und Physik werden jedoch insgesamt nur in der Hälfte der Bundesländer bilingual unterrichtet. Es dominiert im Bereich der Naturwissenschaften weiterhin das Fach Biologie und bezogen auf alle Unterrichtsfächer nach wie vor die gesellschaftswissenschaftlichen Fächer, deren Verbreitung in den einzelnen Bundesländern bei 100 % liegt (KMK, 2013, S. 14). Immerhin hat die Verbreitung des Fachs Chemie seit 2006 zugenommen und wird statt in drei nun in acht Bundesländern angeboten (KMK, 2013, S. 10). Was aus diesen Angaben nicht hervorgeht ist, an wie vielen Schulen Chemie bilingual unterrichtet wird, wie viele Schüler diese Angebote wahrnehmen und in welchem zeitlichen Umfang der Unterricht angeboten wird. Die Situation wird auch dadurch erschwert, dass bilinguale Module, deren Anzahl laut KMK seit 2006 gestiegen ist, häufig statistisch nicht erfasst werden (KMK, 2013, S. 10ff.). Aktuell lassen sich über die Suchmaschine des Ministeriums für Schule und Bildung in Nordrhein-Westfalen 512 Schulen mit bilinguaalem Angebot eruieren, 80 % davon offerieren bilingualen Unterricht in Englisch an (MSW-NRW, 2022). Dabei findet sich deutschlandweit nur ein Gymnasium mit einem bilingualen Angebot in Chemie (FMKS, 2022)⁶.

Für und Wider bilingualen Chemieunterrichts

Die Gründe für die geringe Verbreitung des bilingualen Chemieunterrichts in Deutschland liegt vor allem am Mangel an qualifizierten Lehrern, die über die Doppelfakultas Chemie und Fremdsprache verfügen. Wenngleich dies keine zwingende Voraussetzung für die Erteilung von bilinguaalem Unterricht ist, sind in der Praxis Lehrer, die über diese Doppelfakultas verfügen, deutlich häufiger im bilingualen Unterricht anzutreffen (Mentz, 2010). Der Vorwurf, dass Chemie sich für bilingualen Unterricht aufgrund eines zu geringen Spracherwerbspotentials, des insgesamt zu hohen Schwierigkeitsgrades sowie des Fehlens von Möglichkeiten zu interkulturellem Lernen nicht eigne, konnte in der Vergangenheit überzeugend entkräftet werden (vgl. Bonnet, 2002, S. 126ff.). Die Vorstellung, dass im Chemieunterricht überwiegend fachsprachlich-abstrakt und kontext-reduziert

6 Den gleichen Wert fand Bohrmann-Linde bereits 2013 (Bohrmann-Linde, Chemie, 2013, S. 296).

kommuniziert wird, konnte ebenfalls korrigiert werden (Maxis-Gehrke & Bonnet, 2001, S. 155ff.). Während der naturwissenschaftliche Diskurs unter Experten oft "abstrakt und sprachenübergreifend" ist, so gilt gleiches doch nicht für den unterrichtlichen Diskurs in der Schule (Gnutzmann & Jakisch, 2013, S. 169). Gerade ein handlungsorientierter Unterricht, der die Lebenswirklichkeit der Schüler miteinbezieht und auf den Alltagskonzepten aufbaut, bietet eine Vielzahl von Anlässen für allgemeinsprachliche Kommunikation, gerade wenn experimentiert wird, Handlungsabläufe abgesprochen und Beobachtungen versprachlicht werden (vgl. Bohrmann-Linde, 2012, S. 183; Bonnet, 2002, S. 127; Butzkamm, 1993). Dabei kann sich insbesondere das Verbinden von Sprache und Handeln (*binding*) spracherwerbsfördernd auswirken (vgl. Bohn & Doff, 2010, S. 78; Koch, 2005, S. 34). Bonnet (2002, S. 127ff.) bestimmte das Spracherwerbspotential des bilingualen Chemieunterrichts, indem er eine 23-stündige Unterrichtsreihe videografierte und anschließend auf solche Abschnitte hin untersuchte, die aus spracherwerbstheoretischer Sicht als spracherwerbsfördernd gelten. Den so ermittelten Anteil an spracherwerbsrelevanter Interaktion beziffert er auf 47 %. Gerade die Vielzahl unterschiedlicher Darstellungs- und Symbolisierungsformen (Leisen J., 2005, S. 9ff.) sowie die Anzahl der angesprochenen Sinneskanäle (Rymarczyk, 2003, S. 86) unterstützen im Chemieunterricht die Rezeption und die Produktion im sprachlichen Bereich. Selbst die hohe Dichte an Fachtermini erweisen sich beim Erwerb der Fachsprache nicht als zu große Hürde, da sie oft lateinischen bzw. griechischen Ursprungs sind und damit den deutschen Fachbegriffen sehr ähnlich, wie z.B. *oxidation*, *redoxreaction*, *base*, *electron*, *atom ion* etc. (vgl. Bohn & Doff, 2010, S. 79; Bohrmann-Linde, 2012, S. 183). Dabei erleichtert gerade die starke Normierung der naturwissenschaftlichen Fachsprache ihr Erlernen, da sie sich durch eine größere Präzision, eine einfachere Syntax und den Verzicht auf rhetorische Mittel auszeichnet (Laupenmühlen, 2012, S. 164). Die Vermutung, dass das große Abstraktionsniveau eine zu große Hürde für die fremdsprachliche Behandlung darstellt, konnte bereits früh von Bonnet (2004, S. 288) widerlegt werden. Er beobachtete im bilingualen Unterricht ein ähnlich hohes Kompetenzniveau bei den Konzepten *Stoff*, *Reaktion* und *kleinste Teilchen* wie im monolingual mehrheitssprachlichen Unterricht. Außerdem konnte er - wie später auch Hegerfeldt (2006) - zeigen, dass auch ein Wechsel in die Muttersprache bei Verständnisschwierigkeiten oft nicht weiterhilft, weil die Probleme konzeptueller und nicht sprachlicher Natur sind. Kulturspezifische Sichtweisen werden laut Mäsch (1993, S. 162) im naturwissenschaftlichen Unterricht nicht erkennbar. Die Sprache der Naturwissenschaften gilt als "kulturunabhängig bzw. kulturübergreifend" und der wissenschaftliche Diskurs "als universelle, von den Einzelsprachen losgelöste Kommunikation", weshalb naturwissenschaftliche Fächer "für sprachreflexive wie auch interkulturelle Zugänge weniger in Frage kommen" (Gnutzmann &

Jakisch, 2013, S. 169). Historisch gesehen ist der Erwerb interkultureller Kompetenzen als Zielvorgabe für den bilingualen Unterricht auf den Wunsch nach besserer Völkerverständigung im Zuge des deutsch-französischen Freundschaftsvertrages von 1963 zurückzuführen. Er war vor allem für die Anfänge des damals noch überwiegend deutsch-französischen bilingualen Unterrichts kennzeichnend und führte zu einer Fokussierung auf gesellschaftswissenschaftliche Fächer, welche ein breites Spektrum an kulturspezifischen Aspekten bieten (Wolff D. , 2011, S. 75). Entgegen der Vermutung, dass die Naturwissenschaften keinen Beitrag zum interkulturellen Lernen leisten (vgl. auch Rymarczyk, 2003, S. 87 und Hallet, 1998, S. 118), können im bilingualen Chemieunterricht jedoch sehr wohl eine Reihe von Themen aufgegriffen werden, die Anlässe zu interkulturellem Lernen bieten, weil der gesellschaftliche Kontext Informationen bereithält, die zum Perspektivwechsel und zur Reflexion Anlass geben. So etwa, wenn die gesellschaftliche Relevanz naturwissenschaftlich-chemischer Themen reflektiert wird, wie beim Treibhauseffekt (vgl. Bohrmann-Linde, 2012, S. 185), beim Recycling oder beim Thema Alkohol, wenn etwa die Zeit der Prohibition und die Praxis der *Moonshiners* angesprochen werden. Dabei können historische Themen aber auch zukunftsweisende Entwicklungen wie die moderne Wasserstofftechnologie thematisiert werden. Deren Bedeutung ist, je nach Wahrnehmung des Klimawandels und der Begrenztheit fossiler Ressourcen, von Land zu Land durchaus unterschiedlich. Konkrete politische Maßnahmen können deshalb divers ausfallen, wie auch die nationalen Maßnahmen zum Schutz vor dem Corona-Virus oder zur Bekämpfung der Erderwärmung zeigen. Hier liegt eine Chance zur Reflexion, die sich so im Regelunterricht nicht bietet. Auch die Arbeiten englischsprachiger Forscherpersönlichkeiten wie Ernest Rutherford, John Dalton oder Robert Boyle bieten Anlässe für die Einbeziehung der zielsprachigen Kultur, etwa anhand von Originaltexten aus der Zeit (vgl. Bohrmann-Linde, 2012, S. 194). Bonnet (2004, S. 53), der einen erweiterten Kulturbegriff verwendet, argumentiert, dass die „Aspekthaftigkeit des chemischen Zugriffs auf die Welt“ ebenfalls für viele Schüler eine „Fremdheitserfahrung“ darstellt, welche ein zentrales Element des interkulturellen Lernens ist. Die Lerner müssen sich dabei nämlich an die Denk- und Sichtweise der Naturwissenschaften gewöhnen. Diese umfasst z.B. die Fachterminologie oder spezifische Methoden und Arbeitsweisen, wie das hypothesengeleitete Experimentieren und das sachlich-präzise Protokollieren der Ergebnisse. In diesem Sinne spricht er von einer spezifischen "Kultur der Naturwissenschaften", die sich durch eine rational-analytische Sicht auf die Welt auszeichnet im Gegensatz zur Literatur, welche eher durch einen ästhetisch-holistischen Blick gekennzeichnet ist. Für ihn bedeutet die Auseinandersetzung mit naturwissenschaftlicher Denk- und Sprechweise selbst interkulturelles Lernen. Dieses kommt zwar auch im monolingual-deutschsprachigen Unterricht zum Tragen, wird jedoch durch den Gebrauch der Fremdsprache deutlicher in das Bewusstsein

gerückt. Somit ergeben sich im bilingualen Chemieunterricht, anders als im Regelunterricht, ganz natürlich immer wieder Anlässe, die Besonderheiten der naturwissenschaftlichen Sicht auf die Welt zu reflektieren (Bonnet, 2004). Im Zuge der Globalisierung tritt der Erwerb von Englischkenntnissen für die berufliche Bildung zunehmend in den Vordergrund. Gerade in den Naturwissenschaften ist die Ausgangsmotivation für einen in Englisch geführten bilingualen Unterricht aufgrund der Bedeutung des Englischen als *lingua franca* in der Forschung besonders günstig (vgl. Appel, 2003; Klingauf, 2002; Otten & Wildhage, 2003). Ein großer Teil naturwissenschaftlicher Erkenntnisse wird heute auf Englisch verfasst und derart einer internationalen Forschergemeinschaft zugänglich gemacht. Eine rein monolinguale Schulbildung kann, so die Überzeugung, trotz zusätzlichem Englischunterricht diesen deutlich gestiegenen sprachlichen Anforderungen in Zukunft nicht mehr gerecht werden (Hallet, 2007, S. 27).

3 Bilingualer Chemieunterricht

Im folgenden Kapitel sollen die besonderen Merkmale des bilingualen Chemieunterrichts mit Blick auf seine Zielsetzung (Abschnitt 3.1), die Art und Weise wie gelernt wird (Abschnitt 3.2) und die methodische Unterrichtsgestaltung (Abschnitt 3.3) vorgestellt werden. Die drei Abschnitte liefern damit den didaktisch-methodischen Hintergrund des hier eingesetzten Moduls, das in Abschnitt 5.5 vorgestellt wird.

3.1 Worum geht es im bilingualen Chemieunterricht?

Im bilingualen Unterricht geht es in erster Linie um das Sachfach. Dies betrifft gleichermaßen Inhalte, Ziele und Methoden des Unterrichts. In den Handreichungen zum BU, die in Nordrhein-Westfalen für die Fächer Erdkunde, Geschichte und Biologie erschienen sind und die Lehrpläne des muttersprachlichen Unterrichts ergänzen, heißt es dazu:

Bilingualer Unterricht ist in seinen Anforderungen, Zielen, Inhalten und Methoden grundsätzlich an die geltenden Richtlinien und Kernlehrpläne der Sachfächer gebunden. [...] Themen und Inhalte für die bilingualen Sachfächer werden so ausgewählt, dass sie den allgemeinen curricularen Anforderungen der einzelnen Fächer entsprechen (MSW-NRW, 2012, S. 6).

Da in Bezug auf das bilinguale Sachfach die gleichen curricularen Vorgaben gelten wie im Regelunterricht, muss sich die Beschreibung der Ziele des bilingualen Chemieunterrichts an den Bildungsstandards im Fach Chemie und dem zugrunde liegenden Bildungsbegriff bzw. den dort

formulierten Kompetenzerwartungen orientieren. Der Kernlehrplan für das Fach Chemie in Nordrhein-Westfalen, welcher auf den für alle Bundesländer gleichen gemeinsamen Bildungsstandards basiert, beschreibt den Gegenstand der naturwissenschaftlichen Fächer als die „empirisch erfassbare, in formalen Strukturen beschreibbare und durch Technik gestaltbare Wirklichkeit sowie die Verfahrens- und Erkenntnisweisen, die ihrer Erschließung und Gestaltung dienen“ (Qualis-NRW, 2022). Chemie soll dabei einen Beitrag zu einer „vertieften naturwissenschaftlichen Grundbildung“ (*scientific literacy*) leisten. Darunter wird die Fähigkeit verstanden, chemische Phänomene zu erkennen, die spezifischen Methoden der Erkenntnisgewinnung anzuwenden, eigene Erkenntnisse zu kommunizieren und dabei die Sprache der Naturwissenschaften zu benutzen (Qualis-NRW, 2022)⁷. Außerdem soll die Grundbildung die "Basis für ein selbstgesteuertes und lebenslanges Lernen im Bereich der Naturwissenschaften" liefern (Nerdel, 2017). Dabei bedeutet *scientific literacy* auch, die Grenzen der gewählten Methoden zu erkennen und die Erkenntnisse im Hinblick auf ihre gesellschaftliche Relevanz kritisch zu reflektieren. Den Beitrag des Faches Chemie zur naturwissenschaftlichen Grundbildung definiert Nerdel (2017, S. 6) wie folgt:

1. Untersuchung der stofflichen Welt unter Berücksichtigung der chemischen Reaktion als Einheit aus Stoff- und Energieumwandlung durch Teilchen- und Strukturveränderungen sowie Umbau chemischer Bindungen.
2. Verantwortungsbewusster Umgang mit Chemikalien und Gerätschaften aus Haushalt, Labor und Umwelt sowie das sicherheitsbewusste Experimentieren.
3. Bedeutung der Wissenschaft Chemie, der chemischen Industrie und der chemierelevanten Berufe für Gesellschaft, Wirtschaft und Umwelt; nachhaltige Nutzung von Ressourcen.

Im Kernlehrplan des Landes Nordrhein-Westfalen für das Fach Chemie werden die zu erwerbenden Kompetenzen nach vier Dimensionen unterschieden, den sogenannten *Kompetenzbereichen*. Diese sind: *Fachwissen*, *Erkenntnisgewinnung*, *Kommunikation* und *Bewertung* (Qualis-NRW, 2022).

⁷ Der Begriff *scientific literacy* hat im anglo-amerikanischen Sprachraum eine lange Tradition. Er wurde ab den 1970er Jahren in den USA und Großbritannien entwickelt und soll die "Teilhabe an einer durch Naturwissenschaft und Technik geprägten Gesellschaft ermöglichen" (Bybee, 2002).

Kompetenzbereiche im Fach Chemie	
Fachwissen	Chemische Phänomene, Begriffe, Gesetzmäßigkeiten kennen und Basiskonzepten zuordnen
Erkenntnisgewinnung	Experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen
Kommunikation	Informationen sach- und fachbezogen erschließen und austauschen
Bewertung	Chemische Sachverhalte in verschiedenen Kontexten erkennen und bewerten

Tabelle 1: Kompetenzbereiche im Fach Chemie

(entnommen aus KMK, 2005, S. 7)

Die Inhalte orientieren sich wiederum an vier *Basiskonzepten*, die zentrale Begriffe, Modellvorstellungen und Methoden des Faches betreffen. Es sind: *Stoff-Teilchen-Beziehung*, *Struktur-Eigenschaft-Beziehung*, *chemische Reaktion* und *energetische Betrachtung bei Stoffumwandlungen*. Diese übergeordneten Strukturen erlauben es, chemisches Fachwissen systematisch zu entwickeln und zu vernetzen sowie Fragestellungen situationsübergreifend aus einer fachspezifischen Perspektive zu betrachten (Qualis-NRW, 2022). Lernen findet dabei in Kontexten statt, die authentisch sind und reale Situationen beschreiben, die für die Schüler relevant und mit ihren Möglichkeiten lösbar sind (Qualis-NRW, 2022). Damit wird deutlich, dass sich Chemieunterricht nicht nur ausschließlich an der Wissenschaft Chemie, sondern insbesondere an den Alltagserfahrungen der Schüler orientiert. Inwieweit lassen sich die hier vorgestellten Kompetenzbereiche für das Fach Chemie mit den im bilingualen Unterricht zu erwerbenden Kompetenzen in Einklang bringen? Die meisten Zielbestimmungen zum bilingualen Unterricht formulieren sachfachliche, fremdsprachliche und reflexiv-emanzipatorische Ziele, zu denen auch interkulturelle Kompetenzen gehören (Bonnet & Breidbach, 2013, S. 26). In ihrem Kompetenzmodell zum bilingualen Unterricht unterscheiden Bonnet und Breidbach fünf Dimensionen:

1. Fremdsprachliche Kompetenz
2. Sachfachliche Kompetenz
3. Interkulturelle Kompetenz
4. Metakognitive und selbstbezogene Kompetenzen
5. Emotional-attitudinale Effekte (Bonnet & Breidbach, 2013, S. 28)

Die Ähnlichkeiten der Kompetenzbereiche des Fachs Chemie mit den Kompetenzdimensionen des bilingualen Unterrichts sind nicht zu übersehen. Auch der Chemieunterricht zielt unter den Überschriften Kommunikation und Bewertung auf den Erwerb reflexiver und sprachlicher Kompetenzen. Sachfachliche Kompetenzen werden in den Kernlehrplänen für das Fach Chemie noch weiter ausgeführt und explizit für die Bereiche Fachwissen und Erkenntnisgewinnung gefordert. Hallet (2007:19ff.) weist in diesem Zusammenhang sogar drauf hin, dass sich der dem bilingualen Unterricht laut Bonnet und Breidbach (2013, S. 29f.) zugrunde liegende reflexiv-emanzipatorische Bildungsbegriff viel deutlicher in den Bildungsstandards der Naturwissenschaften wiederfindet als in denen der Fremdsprachen. Während kommunikative Kompetenzen in den Standards der Fremdsprachen nämlich oft nur an „triviale Inhalte und Alltagsanforderungen“ gebunden sind und deshalb „abgetrennt werden vom Weltverstehen und von der Verhandlung zentraler gesellschaftlicher Fragen“ (Hallet, 2007, S. 21), zielt naturwissenschaftliche Bildung im Sinne einer *scientific literacy* ausdrücklich darauf ab „dem Individuum eine aktive Teilhabe an gesellschaftlicher Kommunikation und Meinungsbildung über technische Entwicklung und naturwissenschaftliche Forschung“ zu ermöglichen (KMK, 2005, S. 6). Bonnet resümiert im Hinblick auf die Anschlussfähigkeit zwischen Chemiedidaktik und fächerübergreifender CLIL-Diskussion deshalb:

Die Konzeptualisierung von Chemiekompetenz als *scientific literacy* deckt sich mit der Zielvorgabe, CLIL als Ort des Erwerbs domänenspezifischer Diskurskompetenz zu verstehen. Lediglich die Fremdsprachlichkeit kommt bei CLIL hinzu (Bonnet, 2012, S. 205).

3.2 Lernen im bilingualen Chemieunterricht

Messen und Interpretieren von Lernerfolgen gelingen im Rahmen der vorliegenden Untersuchung nur dann, wenn klar ist, was und wie im bilingualen Unterricht gelernt wird. Im Mittelpunkt der Studie stehen dabei neben sachfachlichen und (fremd-)sprachlichen Lernerfolgen auch motivationale Aspekte des Lernens. Lernerfolge im Bereich des interkulturellen und des reflexiven Lernens werden zwar im Rahmen der Unterrichtsreihe angestrebt, sind jedoch nicht Gegenstand der Auswertung und werden daher nicht weiter thematisiert. Auch wenn die Trennung von sprachlichem und sachfachlichem Lernen in der Praxis des bilingualen Unterrichtes nicht möglich ist (Abschnitt 3.2.1), werden die unterschiedlichen Dimensionen hier aus Gründen der Übersichtlichkeit separat betrachtet. Gegenstand der Studie ist nicht nur der Erwerb, sondern auch der Umgang mit chemischem Fachwissen. Deshalb werden im Bereich des sachfachlichen Lernens neben Prozessen der Verarbeitung und Speicherung von Wissen (Abschnitt 3.2.2.1) auch die

unterschiedlichen Arten von Wissen und kognitiven Operationen (Abschnitt 3.2.2.2) betrachtet. Aus der Tatsache, dass das sachfachliche Lernen im bilingualen Unterricht in der Fremdsprache erfolgt, ergeben sich Grenzen dessen, was gelernt werden kann, aber auch neue Möglichkeiten. Diese Grenzen und Möglichkeiten stehen im Zentrum der Untersuchung zum sachfachlichen Lernerfolg, weshalb der Einfluss der Fremdsprache anschließend thematisiert wird (Abschnitt 3.2.2.3). Inwieweit sprachliche Zugewinne erreicht werden können, wird hier in Bezug auf den englischen und deutschen Fach- und Allgemeinwortschatz untersucht. Eine theoretische Fundierung zu den Mechanismen des Sprachlernens und des Spracherwerbs sowie zum mentalen Lexikon bilingualer Lerner liefern die Abschnitte 3.2.3.1 und 3.2.3.2. Anschließend folgt eine kurze Einführung in die Begriffe Motiv und Motivation (Abschnitt 3.2.4.1) sowie eine Präzisierung der motivationalen Faktoren und deren Einfluss im bilingualen Chemieunterricht (Abschnitt 3.2.4.2). Die Effekte des Geschlechts und der unterschiedlichen Fachinteressen auf die Motivation im Fach Chemie stehen im Mittelpunkt von Abschnitt 3.2.4.3.

3.2.1 Content and language integrated learning: Integration von Sprach- und Inhaltslernen

"Two for the price of one" (Vollmer H., 2010, S. 54) - die Formel bringt auf den Punkt, welche Hoffnung mit dem bilingualen Unterricht verbunden ist: Nämlich Schülern in einem Sachfach zu unterrichten und ihnen gleichzeitig die Möglichkeit zu geben, ihre rezeptiven und produktiven Fertigkeiten in der Fremdsprache weiterzuentwickeln. Dabei ist die Fremdsprache mehr als nur das Medium, durch welches sachfachliche Inhalte vermittelt werden, sondern sie ist auch Gegenstand des Unterrichts. Sprache und Inhalt sind stets aufeinander bezogen, sie werden - wie der Begriff CLIL nahelegt - integriert unterrichtet. Spracherwerb findet dabei im Kontext des Sachfaches statt und sachfachliche Kompetenzen werden auf der Grundlage fremdsprachlicher Kompetenzen erarbeitet (Vollmer H., 2013, S. 124). Sprache und Denken stehen somit in einem engen, wechselseitigen Abhängigkeitsverhältnis. Eine Differenzierung zwischen sachfachlichen und sprachlichen Lernprozessen ist deshalb im bilingualen Unterricht nur schwer möglich.

Da zentrale Denkopoperationen sprachlich gefasst sind [...] und fremdsprachlicher Kompetenzerwerb in den Sachfächern über Konzepterwerb verläuft [...] erscheint es nahe liegend, dass sprachlicher und sachfachlicher Kompetenzerwerb in enger wechselseitiger Abhängigkeit stehen (Bonnet, 2012, S. 202).

Dieses gewissermaßen symbiotische Verhältnis ist auch der Grund dafür, dass Lernfortschritte in dem einen Bereich sich lernfördernd auf den anderen Bereich auswirken können und vice versa

(Rumlich, 2016, S. 97). Solche wechselseitigen, katalytischen Effekte durch gezielte Unterrichtsplanung fruchtbar zu machen - darin besteht wohl eine zentrale Herausforderung des bilingualen Unterrichts. Dass Sprache und Denken sich wechselseitig bedingen, ist im Übrigen keine neue Erkenntnis und gilt für auch für den Regelunterricht. Leisens Forderung nach sprachsensiblen Unterricht macht dies deutlich (Leisen J. , 2013). Im Übrigen ist die Förderung der fachsprachlichen Kompetenzen immer schon ein zentrales Anliegen der Sachfächer und findet sich in den neuen Kernlehrplänen unter der Forderung nach fachspezifischer Diskursfähigkeit wieder (Qualis-NRW, 2022). Das Neue im Setting des bilingualen Unterrichts ist jedoch, dass diese Diskursfähigkeit in einer Sprache entwickelt wird, welche nur eingeschränkt beherrscht wird. Sprachliche Defizite und die neue Perspektive auf die Unterrichtsinhalte, die sich aus dem Gebrauch der Fremdsprache ergibt, beeinflussen das sachfachliche Lernen auf eine völlig neue Art und Weise und können dieses fördern und behindern.

3.2.2 Die sachfachliche Dimension

3.2.2.1 Die Verarbeitung und Speicherung von Wissen

Um die Verarbeitung und Speicherung von Wissen im bilingualen Unterricht zu beschreiben, wird häufig auf die Ideen des Konstruktivismus und der Informationsverarbeitung zurückgegriffen.

Konstruktivismus

Im bilingualen Unterricht wird Lernen im Sinne eines konstruktivistischen Prozesses als individueller Konstruktionsvorgang gesehen, bei dem neu erworbenes Wissen sowohl auf sprachlicher wie auf sachfachlicher Ebene, auf vorhandenen Wissensbeständen aufbauend, entwickelt wird. Wissen kann - so die Theorie - nicht vom Lehrer auf den Schüler übertragen werden, so dass letzterer es in einem dafür vorgesehenen Speicher ablegt, sondern muss vom Schüler aufgenommen, verarbeitet und neu konstruiert werden. Dabei werden neue Wissensstrukturen auf der Basis alter konstruiert; sie knüpfen an diese an und erweitern oder ersetzen diese. Lernen ist damit, gemäß der Vorstellung des Konstruktivismus, ein aktiver Vorgang, bei dem der Lerner kontinuierlich Wissensstrukturen umbaut. Damit ist bilingualer Unterricht vom klassischen, lehrerzentrierten instruktiven Unterrichtsformen zu unterscheiden (Wolff D. , 1997). Im Bereich der Naturwissenschaften entwickeln Schüler zunächst ausgehend von ihren Alltagsvorstellungen Wissensstrukturen, die im Laufe des schulischen Lernens fachwissenschaftlich umstrukturiert werden. In der Didaktik der Naturwissenschaften wird hierfür der Begriff des Konzeptwechsels (*conceptual change*) verwendet (Posner , Strike, Hewson, & Gertzog, 1982). Er

beschreibt, wie fachwissenschaftliche Konzepte sich langsam aus den Alltagsvorstellungen der Schüler entwickeln.

Pupils come to science lessons with ideas about the natural world. Effective science teaching takes account of these ideas and provides activities which enable pupils to make the journey from their current understandings to a more scientific view (Driver, Squires, Rushworth, & Wood-Robinson, 1994).

Auch wenn diese Alltagskonzepte oft nicht falsch sind, so weichen sie doch von wissenschaftlich belastbaren Konzepten ab und müssen deshalb modifiziert werden. Solche Konzeptwechsel stehen im Mittelpunkt von *Conceptual-Change-Theorien*. Der Umbau vorhandener Wissensstrukturen wird von Vertretern des radikalen Konstruktivismus anhand der Lerntheorie von Piaget beschrieben (Glaserfeld von, 1997). Piaget verwendet für kognitive Strukturen, auf die ein Individuum zurückgreift, um seine Umwelt zu verstehen und aus diesem Verständnis Handlungsanweisungen abzuleiten den Begriff des *Schemas*. Dabei werden neue Situationen mit den vorhandenen Schemata verglichen und interpretiert. Gelingt es, ein Erlebnis anhand eines vorhandenen Schemas zu erklären, so wird dieses Erlebnis als Bestätigung des vorhandenen Schemas verstanden (*Assimilation*). Lässt sich die neue Situation nicht anhand eines vorhandenen Schemas erklären, so wird dieses entsprechend den neuen Informationen verändert (*Akkommodation*) (Krapp & Seidel, 2014, S. 249f.). Der Konzeptwechsel vollzieht sich schrittweise auf der Basis neuer Informationen, die in der Regel sprachlich übermittelt werden. Vygotsky (1978) hingegen betont die Bedeutung sozialer Interaktion für Lernprozesse, die sich im Gespräch mit sachkundigen Personen entwickeln. Dabei kommt der Sprache eine zentrale Stellung zu, da Wissen über diese dargestellt und transportiert wird. Dem Lehrer kommt nach Vygotsky die Aufgabe zu, herausfordernde Lernangebote zu machen, die Schüler mit Hilfe seiner Unterstützung weitgehend selbstständig lösen können⁸. Bonnet sieht in der sprachlichen Interaktion den häufigsten Grund für Akkommodation und den Ausgangspunkt für Kompetenzerwerb:

Wenn es gelingt, Lernende zu einer gemeinsamen sensomotorischen und verbalen Arbeit an stofflichen Phänomenen zu bewegen, die ihnen sinnvoll erscheint, kann davon ausgegangen werden, dass in dieser Interaktion Kompetenzerwerb stattfindet (Bonnet, 2004, S. 93).

⁸ Vygotsky (1978) spricht in diesem Zusammenhang von der "*zone of proximal development*" (ZPD) und beschreibt damit ein Lernen, das an Aufgaben orientiert ist, die den Schüler fordern aber nicht überfordern.

Für Bonnet steht die Interaktion der Schüler untereinander und mit dem Lehrer im Zentrum des Kompetenzerwerbs im Chemieunterricht. Diese vollzieht sich im Prozess der Bedeutungsaushandlung der Lernenden untereinander oder mit dem Lehrenden. Neues Wissen entwickelt sich dabei in einem interaktiven Prozess, bei dem Bedeutungen ausgehandelt werden. Lerner entwickeln immer neue Bedeutungen, die sie handelnd prüfen und aufgrund des erhaltenen Feedbacks weiter modifizieren. Der Konzeptwechsel beginnt mit alltagsweltlichen Vorstellungen, welche über Begriffe der Alltagssprache transportiert werden. Somit unterscheidet sich auch die Sprache der Lerner von der zu erwerbenden Fachsprache. Dabei benutzen sowohl Lernende als auch Lehrende im Unterricht ein *interlanguage*, die zwischen einer vollständig allgemeinsprachlichen und einer vollständig fachsprachlichen Redeweise liegt (Selinker, 1972). Im Unterricht geht es dann darum, diese *interlanguage* in Richtung einer fachsprachlich angemessenen Redeweise zu entwickeln. Dazu müssen fachsprachliche Terminologien samt ihrer fachspezifischen Konzeptualisierung gelernt werden.

Informationsverarbeitung

Seit den 1970er Jahren hat sich die Informationsverarbeitung in der Lernforschung etabliert. Ähnlich wie bei einem Computer geht man bei diesem Modell davon aus, dass Informationen vom Lernenden aufgenommen, weiterverarbeitet und abgespeichert werden. Dabei unterscheidet man zwischen Kurz- und Langzeitgedächtnis, je nachdem, wie lange die Information gespeichert wird⁹. Informationen, die über die Sinnesorgane aufgenommen werden, liegen zunächst nur für sehr kurze Zeit im Kurzzeit- oder Arbeitsgedächtnis vor. Um in das Langzeitgedächtnis zu gelangen, müssen diese Informationen tiefer verarbeitet werden. Prinzipiell sieht das Modell der Verarbeitungstiefe nach Craik und Lockhart (1972) drei Verarbeitungsprozesse vor, über die Informationen aus dem Kurz- bzw. Arbeitsgedächtnis in das Langzeitgedächtnis wechseln können: Die visuelle, die phonetische und die semantische Verarbeitung (ebd., 1972). Die semantische, also die bedeutungshaltige Verarbeitung, stellt dabei die tiefste Form der Verarbeitung dar und es wird angenommen, dass diese Form der Verarbeitung zu besseren Erinnerungsleistungen führt. Bei der semantischen Verarbeitung werden die Informationen enkodiert, d.h. in interne Repräsentationen überführt, welche dann in einen Bedeutungszusammenhang mit bereits vorhandenen Informationen treten. Je dichter die neue Information mit den bereits etablierten

9 Das klassische Dreispeichermodeil des Gedächtnisses berücksichtigt noch ein weiteres Gedächtnis, das sogenannte sensorische Gedächtnis, das unterhalb der Bewusstseinssebene liegt, für den Erwerb von schulischem Wissen jedoch weniger relevant ist (vgl. Atkinson & Shiffrin, 1968).

Wissensstrukturen vernetzt ist, umso leichter kann sie aus dem Langzeitgedächtnis wieder abgerufen werden (Siegler, DeLoache, & Eisenberg, 2008). Weil Aufgaben im bilingualen Unterricht authentisch sind, fordern sie die Schüler kognitiv und sprachlich in besonderem Maße. Eine intensive kognitive Aktivität bewirkt aus Sicht der Neurowissenschaft eine tiefere Verarbeitung, die sich in einer größeren Zahl neuronaler Verknüpfungen im Gehirn niederschlägt (van de Craen, Mondt, & Surmont, 2012, S. 34). Zur Beschreibung der Art und Weise, wie Wissen im Arbeitsgedächtnis verarbeitet wird, greift Rodenhauser (2016, S. 37ff.) auf die Theorien von Chandler und Sweller (1991) zurück. Diese gehen davon aus, dass die Kapazität des Arbeitsgedächtnisses begrenzt ist. Für die Belastung des Arbeitsgedächtnisses durch die Verarbeitung neuer Informationen führten Chandler und Sweller (1991, S. 37) den Begriff des *Cognitive Loads* ein. Dabei unterscheiden sie zwischen drei Arten: *Intrinsic*, *Extraneous* und *Germane Load* (Chandler & Sweller, 1991). Der *Load*, welcher durch die Lernprozesse zum eigentlichen Aufbau von Wissensstrukturen hervorgerufen wird, wird als Germane Load bezeichnet. Während Intrinsic Load von der Schwierigkeit des Lerninhalts abhängt und somit nicht beeinflussbar ist, bezeichnet Extraneous Load die zusätzliche Belastung, die durch die Art der Informationsdarstellung hervorgerufen wird. Der Extrinsic Load ist beeinflussbar und kann besonders hoch sein, weil Lerner dann verschiedene Informationen parallel verarbeiten müssen. Zu einem *Cognitive Overload* kann es kommen, wenn Lerner nicht mehr in der Lage sind, relevante Informationen zu verarbeiten, weil die Verarbeitungskapazität des Arbeitsgedächtnisses dies nicht zulässt. Eine solche kognitive Überforderung der Lerner kann im bilingualen Unterricht leicht auftreten, wenn fremdsprachliche und sachfachliche Anforderungen insgesamt zu hoch sind.

3.2.2.2 Unterschiedliche Arten von Wissen und kognitiven Prozessen

Unterrichtliche Lernprozesse werden häufig über den Erwerb konzeptuellen Wissens beschrieben. Beim Lernen spielen jedoch ganz unterschiedliche Arten von Wissen eine Rolle. In der lerntheoretischen Literatur mangelt es nicht an Klassifikationen. Bloom (1956) unterscheidet in seiner Taxonomie zehn Arten von Wissen. Diese ordnet er drei Kategorien zu: Detailwissen; Wissen, wie Details und Fakten behandelt werden können und allgemeines Wissen zu einem Fachgebiet. Anderson und Krathwohl (2001) entwickelten daraus die vier Wissensdimensionen ihrer Lernzieltaxonomie: *Faktenwissen*, *konzeptionelles Wissen*, *prozedurales Wissen* und *metakognitives Wissen*. Konzeptionelles Wissen ist, da es semantisch abgespeichert ist, besonders tief verarbeitet und damit besser reproduzierbar als visuell oder phonetisch verarbeitetes Wissen (vgl. Craik und Lockhart 1972). Dabei wird von der Vorstellung ausgegangen, dass konzeptionell-semantisches Wissen in Form semantischer Netzwerke organisiert ist, sogenannten *Strukturen*.

Zusammengehörende Inhalte sind auf vielfache Art und Weise miteinander verknüpft. Solche Verbindungen können z.B. *Über-Unterordnungs-Relationen*, *Kategoriezugehörigkeiten*, *Eigenschaften* oder *Assoziationen* betreffen. Konzeptuelles Wissen ist somit nicht bildlich-analog, sondern eher propositional-analytisch repräsentiert. Häufig werden solche semantischen Netze mit Hilfe von *concept maps* wiedergegeben. Man geht bei diesem Modell davon aus, dass die Begriffe, die Knoten im semantischen Netz darstellen, ein unterschiedliches Aktivierungspotential haben, d.h. unterschiedlich leicht aus dem Langzeitgedächtnis abgerufen und ins Arbeitsgedächtnis überführt werden können. Diese Aktivierung wird dann entlang der Verbindungen, den Kanten im semantischen Netz, von einem Knoten zum nächsten weitergegeben, wodurch andere Wissens Elemente erinnert werden. Es wird angenommen, dass diese Verbindungen durch häufige Aktivierung weiter verstärkt werden und die Aktivierung zukünftig noch leichter erfolgt. Gleichzeitig können selten aktivierte Verbindungen geschwächt bzw. aufgelöst werden (Krapp & Seidel, 2014, S. 186). Damit ist das Gedächtnis weniger als ein Speicher, sondern als ein Zustand des Gehirns zu verstehen, der sich laufend verändert (Bruegelmann, 2005, S. 100). Für den bilingualen Unterricht wird angenommen, dass solche Aktivierungsprozesse sehr viel häufiger durchlaufen werden, etwa weil aufgrund von Vokabellücken verwandte Begriffe abgerufen und mit dem zu erlernenden Konzept in Beziehung gesetzt werden (Heine, 2010b, S. 209f.).

Das Modell erklärt damit u.a. den Effekt von Üben, Vergessen bzw. Verlernen sowie die Bedeutung, die das Vorwissen für das weitere Lernen hat. Verschiedene empirische Studien zeigen, dass Vorwissen ein guter Prädiktor für Lernerfolg ist, besser noch als eher unspezifische kognitive Potentiale wie Intelligenz (Krapp & Seidel, 2014, S. 189). Je umfangreicher das Vorwissen ist und je besser es strukturiert ist, umso leichter gelingt es, neue Inhalte zu integrieren (Hasselhorn & Mähler, 1998). Weil im bilingualen Unterricht der Konzepterwerb in zwei Sprachen stattfindet und L1- und die L2-Konzepte nicht immer vollständig äquivalent sind (vgl. Abschnitt 3.2.3.2), können bilinguale Lerner in einem kontrastiv angelegten Unterricht detaillierteres Vorwissen erwerben, was sich langfristig positiv auf die Lernleistung im Sachfach auswirken kann. Die im Unterricht zu erwerbenden kognitiven Kompetenzen beschränken sich jedoch nicht nur auf das Memorieren von Wissen, sondern zielen auch auf höhere kognitive Operationen ab. Bloom (1956) unterscheidet sechs Arten kognitiver Operationen: *Wissen*, *Verständnis*, *Anwendung*, *Analyse*, *Synthese* und *Evaluation*. Anderson und Krathwohl (2001) haben diese in leicht veränderter Form übernommen und unterscheiden zwischen *Erinnern*, *Verstehen*, *Anwenden*, *Analysieren*, *Bewerten* und *Erzeugen*. In ihrer Taxonomie, die zur Bestimmung kognitiver Lernziele benutzt wird, lassen sich diese sechs Operationen auf vier unterschiedliche Arten von Wissen anwenden. Damit ergeben sich insgesamt 24 verschiedene Arten kognitiver Leistung.

Wissen	kognitiver Prozess					
	1 Erinnern	2 Verstehen	3 Anwenden	4 Analysieren	5 Bewerten	6 Erzeugen
A. Faktenwissen						
B. konzeptionelles Wissen						
C. prozedurales Wissen						
D. metakognitives Wissen						

Tabelle 2: Kognitive Lernziele nach Anderson und Krathwohl (2001)

Es wird davon ausgegangen, dass auch prozedurales und metakognitives Wissen wie konzeptuelles Wissen vom Lerner konstruiert wird und sich der Erwerb mit den auf Piaget zurückgehenden Prozessen der Assimilation und der Akkomodation beschreiben lassen (siehe Abschnitt 3.2.2.1). Diese werden durch Fakten ausgelöst, welche dann Teil der übergeordneten Wissensstruktur werden. Faktenwissen ist damit weniger komplex, das Memorieren jedoch davon abhängig, wie stark diese innerhalb der kognitiven Struktur vernetzt sind. Wie sich bilingualer Unterricht auf die Schulung anspruchsvollerer Denkopoperationen, wie das Anwenden von Wissen oder das Analysieren einer gegebenen Problemstellung auf der Grundlage des erworbenen Wissens, auswirkt, ist unklar. Bekannt ist, dass inhaltspezifisches Vorwissen auch bei der Entwicklung höherer kognitiver Leistungen eine wichtige Rolle spielt. Ergebnisse aus der Experten-Novizen-Forschung zeigen, dass Experten eines Fachgebietes besser als Novizen in der Lage sind, bestehendes Wissen flexibel anzuwenden (Ferber, 2014, S. 8). Erfahrene Schachspieler können z.B. nach einem kurzen Blick auf eine Stellung diese sehr viel besser wiedergeben, analysieren und auch bewerten als Anfänger. Auf der anderen Seite läuft bilingualer Unterricht aber auch Gefahr, aufgrund der zusätzlichen sprachlichen Schwierigkeiten Schüler zu überfordern (*Cognitive Overload*). Das liegt daran, dass komplexere Lernaufgaben in der Regel mehr Informationen enthalten als gleichzeitig im Arbeitsgedächtnis behalten und bearbeitet werden können (siehe Abschnitt 3.2.2.1). Die Kapazität des Arbeitsgedächtnisses hat wiederum einen Einfluss auf die Fähigkeit, höhere Denkopoperationen ausführen zu können. Es konnte etwa gezeigt werden, dass die Kapazität des Arbeitsgedächtnisses stark mit der Fähigkeit zu schlussfolgerndem Denken, wie es im Rahmen von Intelligenztests gemessen wird, korreliert (Krapp & Seidel, 2014, S. 174). Es ist jedoch auch bekannt, dass sich die

Kapazität des Arbeitsgedächtnisses vergrößern lässt, etwa durch Zusammenfassen einzelner Informationen zu sogenannten *chunks*. Dabei handelt es sich um Informationsbündel, in welchen mehrere Informationseinheiten zu nur einer Einheit verknüpft werden. Dadurch wird wieder mehr Kapazität für die kognitive Verarbeitung der Informationen frei. Laut Baddeley (2012) ist dabei eine *zentrale Exekutive* für die Auswahl relevanter Informationen, deren Bearbeitung und Transfer in das Langzeitgedächtnis zuständig. Es gibt nun Hinweise darauf, dass diese zentrale Exekutive bei bilingualen Schülern besonders effektiv arbeitet, und sie - ähnlich wie der erfahrene Schachspieler - besser darin sind, relevante Informationen zu erkennen und zu verarbeiten (Bialystok & Martin: 2004).

3.2.2.3 Der Einfluss der Fremdsprache auf das sachfachliche Lernen

Erwerb von Fachwissen

Die Verwendung der Fremdsprache kann den Konzepterwerb erschweren, wenn die sprachlichen Möglichkeiten der Schüler eine angemessene Konzeptualisierung nicht ermöglichen. In diesen Fällen werden entsprechende Unterstützungsmaßnahmen (*scaffolding*) nötig. Gleichzeitig bietet die Verwendung der Fremdsprache aber, wie im vorangegangenen Abschnitt gesehen, auch lernpsychologische Vorteile. Die nur unvollkommen beherrschte Fremdsprache kann nämlich dazu anhalten, genauer hinzuhören und Phänomene sorgfältiger zu erarbeiten. Das mehrfache Umformulieren auf der Suche nach der bestmöglichen Formulierung schafft dabei weitere Anknüpfungspunkte im Gehirn und könnte eine stärkere neuronale Vernetzung zur Folge haben (Heine, 2010b). Hegerfeldt (2006) sieht in der Verlangsamung des Unterrichtstempos und der durch gezielte Wortschatzarbeit ausgelösten Wiederholungsphasen einen deutlichen Vorteil des bilingualen Unterrichts:

Im bilingualen Sachfachunterricht gehen die Phasen der Wortschatzarbeit stets mit einer inhaltlichen Wiederholung und Festigung, da die sprachliche Umwälzung zwangsläufig an die Inhalte des Sachfachs gekoppelt ist. Das zweimal Hinschauen (Christ 2004, S. 34), das der bilinguale Unterricht erzwingt, bringt eine im Vergleich zum muttersprachlichen Unterricht größere sachfachliche Kompetenz und Sicherheit, da in Letzterem die Festigung und Wiederholung häufig viel zu kurz kommen (Hegerfeldt, 2006, S. 39).

Auch für Heine (2010b) können fremdsprachliche Mängel der Grund für eine tiefere semantische Verarbeitung sein:

Durch die Verwendung der L2 treten Lerner häufig an die Grenzen ihrer Ausdrucksfähigkeit und müssen daher mehr kognitiven Aufwand in die Suche nach einer adäquaten sprachlichen Form investieren (Heine, 2010b, S. 211).

Anhand von Lautdenkprotokollen konnte Heine (2010a) zeigen, dass Lerner, die beim Lösen einer schriftlichen Geografie-Aufgabe in der Fremdsprache auf Vokabellücken stoßen, sich automatisch auf die Suche nach fremdsprachigen Alternativformulierungen machen. Indem sie dazu verschiedene verwandte Begriffe auswählen, diese miteinander vergleichen und sich schließlich für die ihrer Meinung nach treffendste Formulierung entscheiden, befassen sie sich länger mit dem entsprechenden Sachverhalt. Außerdem stellen sie Beziehungen zwischen den leicht voneinander abweichenden Konzepten her, wodurch diese tiefer semantisch verarbeitet werden. Die Suche nach der sprachlichen Form führt somit zu einer intensiveren Rekonstruktion des Wissensgegenstandes, der in Beziehung zu ähnlichen Konzepten abgebildet und damit stärker vernetzt gespeichert wird. Wie die Analyse weiterer Aufgabenlösungsdaten gezeigt hat, kann die durch Wortschatzlücken ausgelöste, tiefere semantische Verarbeitung tatsächlich zu einer besseren Rekonstruierbarkeit der Informationen durch die Schüler des bilingualen Unterrichts führen (vgl. Lamsfuß-Schenk, 2008; Vollmer H. , 2006; Zydati, 2007; Heine, 2010b). Zu der Frage, welche Folgen die Verwendung einer weniger sicher beherrschten Fremdsprache auf das sachfachliche Lernen hat, wurden bisher nur *sekundäre* Konsequenzen betrachtet. Das sind weitere, zusätzliche Sprach- oder Denkhandlungen, die durch die eingeschränkten fremdsprachlichen Möglichkeiten der Lerner ausgelöst werden. Zu den *primären* Konsequenzen zählen die unmittelbar durch die Verwendung der Fremdsprache ausgelösten Denkprozesse in ihrer Verschiedenheit von den Denkprozessen, wie sie durch die Muttersprache ausgelöst werden. Für den schulischen Unterricht geht es dabei insbesondere um den Erwerb wissenschaftlicher Konzepte bzw. die Veränderung von Alltagskonzepten der Schüler hin zu wissenschaftlichen Konzepten (*conceptual change*). Dabei steht die Frage im Vordergrund, wie sich der Konzepterwerb bzw. der Umbau in der Fremdsprache vom Konzepterwerb bzw. Umbau in der Muttersprache unterscheidet. Zur Frage, wie Wissen und sprachliche Realisierung im Gehirn abgespeichert sind und wie sich bi- und monolinguale Sprecher in Bezug auf die Verarbeitung dieser Informationen unterscheiden, gibt es unterschiedliche Meinungen. Es gibt Hinweise darauf, dass die Verwendung der L2 im bilingualen Unterricht den Konzepterwerb erleichtert, wenn das fachwissenschaftliche Konzept in der L1 mit dem Alltagskonzept besetzt ist, in der Fremdsprache jedoch nicht. Dies ist beim Konzept der *Säure* im Deutschen und im Englischen der Fall. Während die Alltagsvorstellung auf den Geschmack etwa im Gegensatz zu *süß* abhebt, definiert die Fachwissenschaft *sauer* im Gegensatz zu *alkalisch* als ein Milieu, mit einem pH-Wert von weniger als 7. Während im Deutschen die Bezeichnung *sauer* sowohl in der Alltags- wie in der Fachsprache verwendet wird, ist dies im Englischen nicht der Fall. Hier unterscheiden sich Alltagssprache und Fachsprache durch die Verwendung der Begriffe *sour*

und *acid*. Mehrere Studien sehen hier einen Vorteil für den Einsatz der Fremdsprache (vgl. z.B. Haagen-Schützenhöfer, Mathelitsch, & Hopf, 2011, S. 234ff. oder Hegerfeldt, 2006, S. 39f.).

Für den bilingualen Unterricht wird außerdem diskutiert, ob die Verwendung der Fremdsprache eine insgesamt gesteigerte kognitive Leistungsfähigkeit des Gehirns zur Folge hat. Diehr (2016, S. 60) weist auf neuste Erkenntnisse in den Neurowissenschaften hin, die zeigen, dass bilinguale Menschen geistig besonders flexibel sind. Rumlich (2016, S. 119) macht jedoch klar, dass in diesem Bereich noch viel Forschungsarbeit nötig ist (vgl. Abschnitt 4.1.3).

3.2.3 Die (fremd-)sprachliche Dimension

3.2.3.1 Spracherwerb im Zeichen von BICS und CALP

Wenn es darum geht, den sprachlichen Erfolg bilingualen Unterrichts lerntheoretisch zu begründen, wird häufig *Krashens Monitor-Theorie* bemüht (Krashen & Terrell, 1983). Der Tradition Chomskys folgend geht er davon aus, dass Spracherwerb automatisch erfolgt, und zwar immer dann, wenn Sprecher für sie bedeutungsvolle und sprachlich verständliche Äußerungen zu verstehen suchen. Quasi automatisch und vom Lerner unbemerkt, überwacht und analysiert ein zentrales Spracherwerbsorgan die sprachliche Form der Äußerung mit dem Ziel, die zugrundeliegende Botschaft zu entschlüsseln. Explizite Hinweise auf die sprachliche Form, also etwa die Bedeutung einzelner Wörter, grammatischer Strukturen oder gar ein separates Einüben sind überflüssig, solange der Input für den Lerner verständlich und für ihn persönlich wichtig ist. Während also eine bewusste, kognitive Auseinandersetzung mit der sprachlichen Form entfällt, sind affektive Faktoren, wie Selbstbewusstsein, Motivation und eine angstfreie Lernumgebung besonders förderlich für den selbstgesteuerten Spracherwerb¹⁰. Die Kommunikation im bilingualen Unterricht ist in erster Linie mitteilungsbezogen, mit einem klaren Fokus auf den inhaltlichen Aspekten des Sachfachs. Damit ähnelt die Spracherwerbssituation im bilingualen Unterricht der eines Kindes, das die Muttersprache lernt (Mehisto, Marsh, & Frigols, 2008, S. 26). Dass aber schulischer Fremdspracherwerb im Gegensatz zum natürlichen Erwerb der Muttersprache ohne die bewusste Hinwendung auch auf die sprachliche Form nicht auskommt, legen die Ergebnisse aus der kanadischen Immersionsforschung nahe. Es zeigte sich, dass die sprachliche Korrektheit von Äußerungen auch solcher Immersionsschüler, die bereits mehrere Jahre an dem entsprechenden Programm teilnahmen, im Vergleich zu gleichaltrigen Muttersprachlern stark eingeschränkt war.

10 Krashen (1983) spricht in diesem Zusammenhang von einem *affective filter*. Diese Metapher beschreibt den Einfluss negativer Emotionen und Einstellungen, die gewissermaßen wie ein Filter den Spracherwerbsprozess beeinträchtigen.

Die Idee, eine Fremdsprache ausschließlich immersiv, also durch Eintauchen in ein fremdsprachliches Sprachbad ohne Hinwendung auf die sprachliche Form zu erwerben, erwies sich in der Folge als wenig effektiv. Dies gilt insbesondere für den BU, da dieser noch weniger Möglichkeiten zum Sprachkontakt bietet als langjährige Immersionsprogramme (Rumlich, 2016, S. 115). Für den bilingualen Unterricht hat sich deshalb die Überzeugung durchgesetzt, dass auch ein bewusster Fokus auf die sprachliche Form und die explizite Analyse sprachlicher Phänomene wichtig sind. Im bilingualen Unterricht nach deutschem Modell geschieht dies vor allem im Fremdsprachenunterricht, der parallel weitergeführt wird. Auch im Rahmen bilingualen Unterrichts können allerdings sprachbezogene Phasen, sofern diese in einem ansonsten mitteilungsbezogenen Unterricht auftauchen, helfen, die sachbezogene Kommunikation zu verbessern (Long, 1991, S. 45ff.). Ob derart bewusst gelerntes sprachliches Wissen jedoch automatisiert werden kann und in unbewusster Form für spontane Sprachproduktion zu Verfügung steht, wird in der Literatur kontrovers diskutiert. Ellis (1991) unterscheidet zwei Arten von *Input*, den bedeutungs- und den form-fokussierten Input, die jeweils zu unterschiedlichen Arten von Sprachwissen führen und an unterschiedlichen Stellen im Gehirn gespeichert werden. Während inhaltsbezogener Input implizites Wissen bewirkt, führt formbezogener Input zu explizitem Wissen. Ellis geht davon aus, dass explizites Wissen dem Lerner dabei hilft, inhaltsbezogenen Input zu verarbeiten und einen höheren Grad an formbezogener Richtigkeit zu erreichen (Ellis, 1991, S. 190ff.). Auch Ellis vermutet - ähnlich wie Krashen - eine Art Schnittstelle, die eine Überführung von explizitem in implizites Sprachwissen ermöglicht. Unabhängig von der Existenz einer solchen Schnittstelle sind jedoch die Erfolge bewusster Hinwendung zu sprachlichen Formen und die Defizite eines ausschließlich inhaltsbezogenen Unterrichts nicht von der Hand zu weisen (vgl. Rumlich, 2016, S. 115). Ein solcher Fokus auf sprachliche Formen ergibt sich jedoch auch auf natürliche Art und Weise, wenn Schüler unbewusst die Reaktionen auf eigene Sprachproduktionen analysieren und daraus Rückschlüsse über die sprachliche Richtigkeit ihrer Äußerung ziehen. Swain (1995) hat deshalb dem Input-Modell Krashens sein Modell der *Output-Orientierung* gegenübergestellt. Es besagt, dass Lerner-Output auch dazu dient, eigene Hypothesen zur sprachlichen Form zu testen und anhand des Feedbacks das eigene metalinguistische Wissen zu erweitern. Wenn Lerner nicht verstanden haben, was gesagt wurde, treten sie in Interaktion mit dem Gegenüber und versuchen durch Rückfragen, Klarheit über den Inhalt sowie über die sprachliche Form einer Äußerung zu erhalten. Long (1980) spricht im Rahmen seiner Interaktionstheorie davon, dass *Bedeutungen ausgehandelt* werden, wenn der Input unverständlich ist. Neben diesen eher zufälligen Momenten, in denen sprachliche Form von einzelnen Schülern reflektiert wird, ist es wichtig - und das hat die kanadische Immersionsforschung gezeigt, die Aufmerksamkeit der Schüler immer wieder auf die

Besonderheiten fremdsprachlicher Formulierung zu lenken, insbesondere, wenn diese sich von der Muttersprache unterscheiden oder wenn sie im gegebenen Kontext von großem Gebrauchswert sind. Explizite Spracharbeit, in der sprachliche Formen benannt, reflektiert und eingeübt werden, sind damit integraler Bestandteil der Didaktik des bilingualen Unterrichts geworden. Damit ist jedoch keinesfalls eine Rückkehr zu den strukturalistischen Unterrichtsformen der audio-lingualen oder audio-visuellen Methode gemeint, denn sprachliche Strukturen werden nicht inhaltlich sinnentleert quasi um ihrer selbst willen betrachtet, sondern immer in Bezug auf den sachfachlichen Kontext und ihrer Funktion in einer inhaltsbezogenen Kommunikationssituation. Damit bleibt bilingualer Unterricht den Maximen eines kommunikativen Fremdsprachenunterrichts verbunden, dessen Unterrichtsgegenstände authentisch und dessen Kommunikationsanlässe in aller erster Linie inhaltsbezogen sind (Breidbach, 2013, S. 14f.). Gleichzeitig sind die Art der Kommunikation und die erworbenen Sprachkompetenzen im bilingualen Unterricht eingeschränkt. Tatsächlich handelt es sich oft um akademische, kognitiv anspruchsvolle und auf abstrakte Zusammenhänge abzielende Formen der Kommunikation, wie sie für die unterrichtliche Behandlung in den Sachfächern üblich ist. Cummings (1978) hat für diese Art des Sprachgebrauchs den Begriff *Cognitive-Academic Language Proficiency* (CALP) geprägt und grenzt diesen von einer eher auf die Bewältigung von Alltagssituationen abzielenden Sprache ab, für die er den Begriff *Basic Interpersonal Communicative Skills* (BICS) verwendet. Cognitive-Academic Language Proficiency, als Sprache von Schule und Wissenschaft verstanden, zeichnet sich durch ein stärker differenziertes und abstrakteres Sprachrepertoire aus. Es handelt sich um ein bildungssprachliches Register, das vor allem durch Strukturen gekennzeichnet ist, mit denen "komplexe und abstrakte Inhalte" ausgedrückt werden können, wobei es Merkmale "konzeptioneller Schriftsprachlichkeit" besitzt (vgl. Thürmann, 2010, S. 139). Nach Thürmann (2010, S. 140) ist dieses Register dadurch gekennzeichnet, dass es "explizit, genau, objektiv, komplex, strukturiert, leidenschaftslos bzw. emotionsfrei, entpersonalisiert, situations- und kontextunabhängig und redundanzarm" ist. Basic Interpersonal Communicative Skills hingegen meint ein Repertoire an kommunikativen Fertigkeiten, wie es zur Bewältigung kognitiv weniger anspruchsvoller, konkreter Alltagssituationen nötig ist, etwa beim Einkaufen, im persönlichen Gespräch mit dem Nachbarn oder beim Schreiben persönlicher E-Mails an Freunde. In der Praxis zeichnet sich die Unterrichtssprache jedoch durch Anteile beider Register aus. Alltagsprache kommt zum Einsatz, weil der Unterricht an den Alltagsvorstellungen der Schüler anknüpft und weil neue Konzepte zunächst in der Sprache der Schüler formuliert werden, um sie begreifbar zu machen. Ting (2011) formuliert dazu:

Good CLIL-teaching thus moves FL-education from Cummins' BICS (Basic Interpersonal Communication Skills) to CALP (Cognitive Academic Language Proficiency) (1981;

1984). CLIL is therefore definitely not about “doing genetics, hydroelectricity and quantum mechanics in a foreign language. The output of CLIL-learning which is filtered through these three Operands enables learners to not only understand the Content well, but also be able to Communicate about content knowledge effectively: that’s CLIL (Ting, 2013).

3.2.3.2 Das mentale Lexikon bilingualer Lerner

Die mentalen Prozesse beim Erwerb von Fachkonzepten unterscheiden sich im bilingualen Unterricht von denen im Regelunterricht. Zur Verdeutlichung entwickelte Diehr (2016) ein Modell, welches in Anlehnung an Pavlenko (2009, S. 132) Konzepte als „eine mentale Repräsentation eines umfassenden und vielschichtigen, aber in sich geschlossenen und strukturierten Wissensbestandteils“ auffasst, wie z.B. WASSERKREISLAUF oder METABOLISMUS“ (Diehr 2016, S. 66). Das *Integrated Dynamic Model* (IDM) von Diehr (2016, S. 70ff) stellt eine Erweiterung des *Modified Hierarchical Model* nach Pavlenko (2009) dar. Dabei berücksichtigt es zusätzlich vier Besonderheiten des Konzepterwerbs im spezifischen Setting bilingualen Unterrichts. Zum einen stellt es die Sprachspeicher der L1 und der L2 sowie den Konzeptspeicher als dynamische Größen dar, deren Größen sich im Verlauf des Lernprozesses verändern. Zwar wird bei bilingualen Lernern nach dem deutschen Modell in der Regel die L1 dominieren, jedoch kann es sein, dass im Verlaufe eines überwiegend monolingual in der L2 geführten Unterrichts der L2-Speicher zumindest für den Bereich der Fachsprache größer ist als der L1-Speicher. Deshalb unterscheidet das IDM zum zweiten innerhalb der L1- und L2-Sprachspeicher zwischen Fach- und Alltagssprache. Diese Unterscheidung geht im Wesentlichen auf Cummings’ Abgrenzung von BICS und CALP zurück. Auch Diehr (2016, S. 72) weist darauf hin, dass beide Registervarietäten im fachlichen Diskurs zum Tragen kommen und miteinander interagieren. Daraus ergibt sich drittens auch, dass nicht mehr grundsätzlich von einer stärkeren L1-Wort-Konzept-Verbindung gesprochen werden kann, weil auch Fachkonzepte in der L2 erworben werden, die in der L1 noch gar nicht vorliegen. Laupenmühlen (2012) beobachtet beispielsweise in einem monolingual-fremdsprachlich geführten Biologieunterricht, dass einigen Schülern nicht klar war, dass die seit Jahren verwendete Bezeichnung *metabolism* die englische Übersetzung für das deutsche Wort *Stoffwechsel* ist und damit auf das gleiche Konzept verweist (Laupenmühlen, 2012, S. 175). Wieder anderen Schülern war unklar, dass sich die zu *proteins* erworbenen Erkenntnisse auch auf *Eiweiße* übertragen lassen. Das bedeutet, dass der Konzeptwechsel von der Alltagsvorstellung hin zum wissenschaftlichen Begriff nur in der L2, also für *protein*, stattgefunden hat und für *Eiweiß* immer noch das alltagsweltliche Konzept vorliegt. Zuletzt unterscheidet das Modell innerhalb des Konzeptspeichers zwischen vier verschiedenen Äquivalenzgraden zwischen Konzepten aus der L1 und der L2. Dabei

wird unterschieden, ob die Äquivalenz vollständig (*precipitation* und *Niederschlag*), partiell (*science* und *Wissenschaft*), scheinbar vorhanden (*tube* und *Tube*) oder fehlend (*half-term* ohne Entsprechung im Deutschen) ist (Diehr, 2016, S. 72f.).

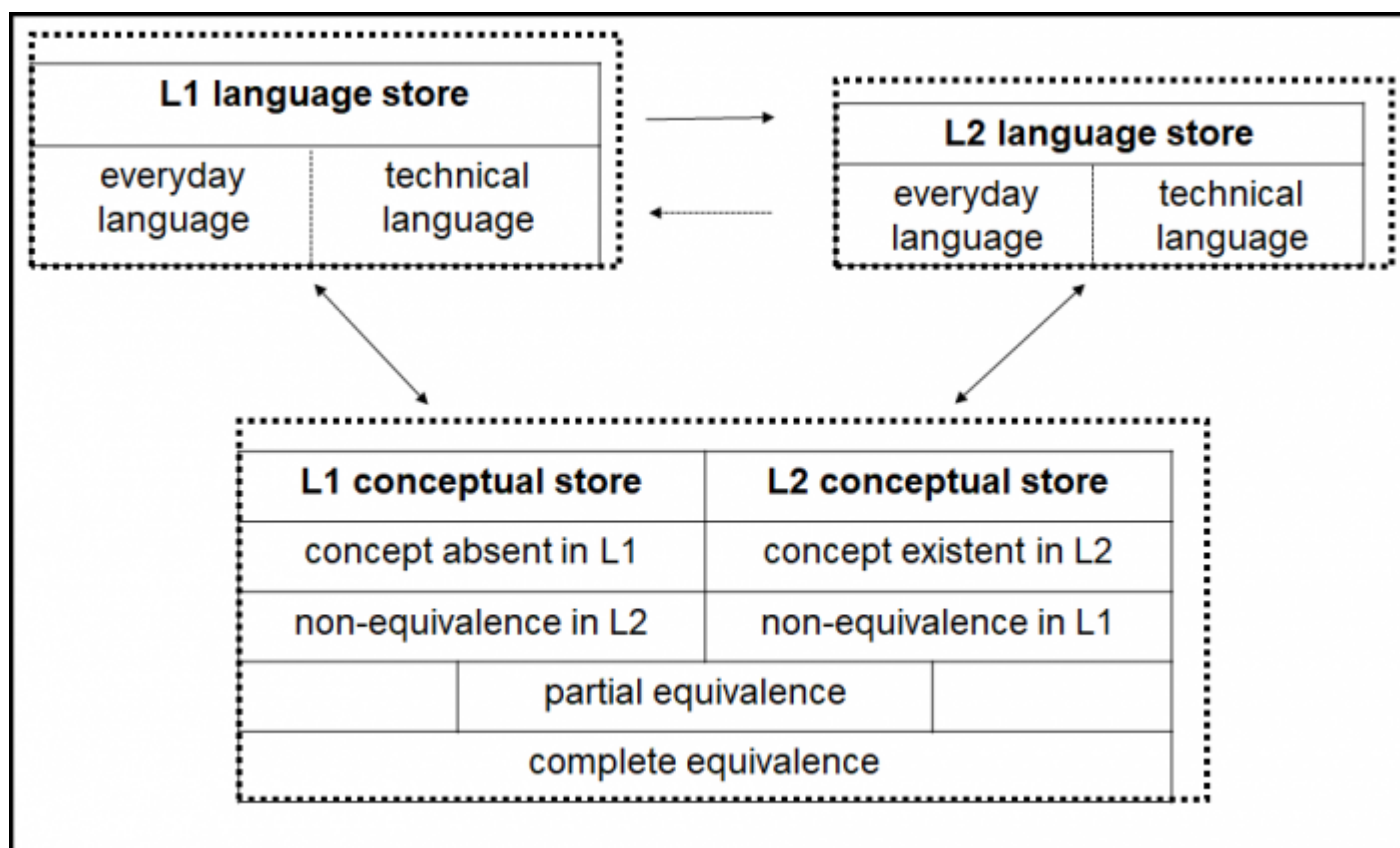


Abbildung 1: Das Integrated Dynamic Model

(entnommen aus Diehr, 2016, S. 71)

Heine (2010b) vermutet - ähnlich wie Pavlenko (2009) und Diehr (2016) - ebenfalls, dass es „zwei getrennte Wissensbereiche“ gibt, „die aber teilweise überlappen und stark untereinander vernetzt sind“ (Heine, 2010b, S. 207). Ausschlaggebend dafür, ob es überhaupt konzeptionelle Unterschiede bei L1- und L2-Begriffen gibt, ist laut Heine jedoch, wann, wo und wie mehrsprachige Lerner die Konzepte erworben haben. Ob Konzepte nämlich in Abhängigkeit von der Sprache unterschiedlich gebildet werden, und sich inhaltlich von den Konzepten der L1 unterscheiden, hängt dabei vor allem von den Begriffen und ihrer kulturellen Bedingtheit ab. So unterscheiden sich das englische *breakfast* sehr wohl vom deutschen *Frühstück*, während sich der englische *hammer* wohl kaum vom deutschen *Hammer* unterscheidet. Der Vermutung, dass darüber hinaus strukturelle oder diskursspezifische Merkmale der Einzelsprachen in der Praxis einen Einfluss darauf haben, wie

Wissen in unterschiedlichen Sprachen aufgebaut wird, steht Heine (2010b, S. 207) skeptisch gegenüber.

Es ist damit eine theoretisch mögliche Option, dass die im BU durch eine L2 aufgebauten Wissensinhalte sich von den im Regelunterricht vermittelten unterscheiden. Allerdings muss hinterfragt werden, welches Ausmaß solche Unterschiede in der Praxis einnehmen können. Die Annahme, bilinguale Lerner erwürben andersartiges fachliches Wissen im BU als Regelschüler, setzt nämlich zum einen voraus, dass lexikalisch, morphosyntaktisch und/oder diskursspezifisch stark voneinander abweichende Sprachen beteiligt sind. Im Fall von typologisch eng verwandten Sprachen wie etwa Deutsch und Englisch ist hier zumindest dahingehend Skepsis geboten, ob die - natürlich vorhandenen - strukturellen Unterschiede tatsächlich so wesentlich in konzeptuelle Strukturen eingreifen, dass man von einem besonderen Merkmal des BU sprechen kann.

Für fachwissenschaftliche Konzepte in den Naturwissenschaften kann dabei überwiegend von vollständiger Äquivalenz ausgegangen werden. Die Verschiedenheit der L1- und L2-Konzepte ist für die Naturwissenschaft schon deshalb als gering zu bewerten, weil naturwissenschaftlichen Konzepten exakten Definitionen unterliegen, die unabhängig von der jeweiligen Kultur des Landes sind. Das englisch *atom* unterscheidet sich eben nicht vom deutschen *Atom*, wenn es im naturwissenschaftlichen Kontext zur Aufklärung chemischer Reaktionen herangezogen wird. Es ist gerade das erklärte Ziel wissenschaftlicher Fachterminologien, universal und nicht kulturspezifisch zu sein. Bonnet (2016, S. 47) weist darauf hin, dass sich fachsprachlich geprägte Register einzelner Sprachen deutlich näher sind als die Allgemeinsprachen. Diese Ähnlichkeiten betreffen auch sprachübergreifende gleiche Strukturmerkmale, wie Nominalisierung, passivische Konstruktionen oder der Gebrauch von Komposita. Damit ist die naturwissenschaftliche Fachsprache deutlich weniger von linguistischer Relativität betroffen als die Allgemeinsprache. Die beschriebene Konzeptgleichheit betrifft, wie gesehen, naturwissenschaftliche Fachkonzepte. Diese werden jedoch im Unterricht in der Regel nicht aus dem Nichts generiert, sondern aus Alltagskonzepten entwickelt (*conceptual change*). Bezüglich der Bekanntheit von Alltags- und Fachkonzepten aus der L1 sind dabei drei Fälle zu unterscheiden:

1. Das fachwissenschaftliche Konzept ist auch aus der L1 nicht bekannt (z.B. Metabolismus).
2. Das fachwissenschaftliche Konzept ist in der L1 mit einem Alltagskonzept besetzt und muss deshalb modifiziert werden (z.B. Licht und Wärme).
3. Das fachwissenschaftliche Konzept ist bereits aus der L1 bekannt. In diesem Fall erfolgt eine Übernahme des Konzepts in die L2 und ggf. eine Modifikation, falls sich die Konzepte in der L1 und der L2 unterscheiden (z.B. "Recycling im englischsprachigen und im deutschen Sprachraum).

Im bilingualen Unterricht müssen in diesen Fällen Maßnahmen ergriffen werden, da sich eine doppelte Fachliteralität nicht von selbst einstellen kann. Es empfehlen sich Sprachvergleiche, um die Unterschiedlichkeit des Alltagssprachlichen und des fachspezifischen Konzepts zu verdeutlichen. Bohrmann-Linde (2012, S. 192) erkennt deshalb im bilingualen Unterricht die Chance zu einem "Mehr an allgemeiner Spracharbeit, gründlichere Fachspracharbeit und eine komparative und kontrastierende Betrachtung der Begriffe aus den Alltags- und den Fachkontexten“.

3.2.4 Die motivationale Dimension

Der Erfolg schulischen Lernens ist von vielen Faktoren abhängig. Dazu zählt unter anderem die Motivation. Motivierte Schüler strengen sich mehr an, sind bereit mehr Zeit in eine Aufgabe zu investieren und sind häufig erfolgreicher als andere (Grünwald, 2006, S. 55). Ein Ziel der Motivationsforschung ist es deshalb, motivationsfördernde Faktoren zu isolieren und Lernumgebungen so zu gestalten, dass Schüler besonders motiviert sind. Für das Setting des bilingualen Unterrichts besteht die Hoffnung dabei darin, einen neuen Weg zu finden, Schüler sowohl für das Erlernen der Fremdsprache als auch für das Erlernen des Sachfachs zu motivieren. Ein Anliegen, das für das Fach Chemie von besonderer Wichtigkeit ist, da es unter Schülern nach wie vor zu den unbeliebtesten Fächern gehört (siehe Abschnitt 4.3.2). Außerdem geht es um die Frage, wie bilingualer Unterricht gestaltet sein muss, dass er nicht genau das Gegenteil bewirkt und Schüler sowohl von der Verwendung der Fremdsprache als auch von den Herausforderungen des Sachfachs abschreckt. Auch diese Frage ist gerade für das Fach Chemie von großer Relevanz, da Schüler häufig Verständnisprobleme artikulieren (Barke & Harsch, 2001, S. 38). Darüber hinaus wird in der Forschung zum bilingualen Unterricht die individuelle motivationale Ausgangslage der Schüler in den Blick genommen, mit dem Ziel, motivationale Faktoren zu isolieren, die solche Lerner charakterisieren, die in besonderem Maße vom bilingualen Unterricht profitieren und solche, die dies nicht tun. Hieraus können sich Empfehlungen für die Durchführung bilingualer Module in Regelklassen ergeben.

3.2.4.1 Grundbegriffe

Der Begriff *Motiv* leitet sich vom lateinischen *movere* - *bewegen* ab, und bezeichnet Alltagssprachlich den Beweggrund, der menschlichem Handeln zugrunde liegt. Solche Motive können Gefühle (z.B. Spaß), Kognitionen (z.B. Wissen), Einstellungen (z.B. Gesellschaft), Bedürfnisse (z.B. Geltung) und Wertüberzeugungen (z.B. Nützlichkeit) sein (Abendroth-Timmer, 2007, S. 30). Eine umfassende wissenschaftliche Definition stammt von Boosch (1983):

Motive sind individuelle, zeitlich relativ stabile, an bestimmte situative Reize gekoppelte, meist erlernte Handlungsdispositionen für innere und äußere Handlungen, die die Auswahl zwischen alternativen Handlungsweisen in Entscheidungssituationen, das Auftreten wiederkehrender bevorzugter oder gemiedener Handlungsweisen im individuellen Lebensverlauf und die Dynamik der Durchführung von Handlungen bestimmen (Boosch, 1983).

Laut Krapp (2014, S. 194) beschreibt *Motivation* im Kontext schulischen Lernens vor allem „die situationsspezifische Bereitschaft des Schülers, neues Wissen oder neue Fähigkeiten zu erwerben“ (Lernmotivation)¹¹. Auch hier spielen sowohl personenbezogene Faktoren wie Bedürfnisse, Motive und Ziele, aber auch situationsbezogene Faktoren, wie Gelegenheiten und Anreize eine entscheidende Rolle. In seinem Handlungsmodell zur Motivation fügt Krapp den drei Phasen der Motivation (finden eines Grundes für die Handlung, bewusste Entscheidung für die Handlung sowie Aufrechterhaltung der Handlung) eine weitere hinzu, nämlich die der Auseinandersetzung mit den Folgen der Handlung (Krapp, 2014, S. 195). Aus der Rückwirkung dieser Auseinandersetzung auf die Motive zukünftiger Handlungen erklärt sich die individuelle Ausgestaltung von Motiven und deren Erlernbarkeit. Damit berücksichtigt das Handlungsmodell neben situationsspezifischen Faktoren sowohl Bedingungsfaktoren der zielgerichteten Handlung (Antezedenzien) als auch deren Konsequenzen. Im Bereich der kognitiven Auseinandersetzung mit den Konsequenzen schulischer Handlungen entwickelten sich eine Vielzahl unterschiedlicher Theorien, die auf dem von Heckhausen (1977) im deutschsprachigen Raum eingeführten *Erwartungs-Wert-Konzept* beruhen. Diesem liegt die Annahme zugrunde, dass die Motivation im Wesentlichen durch zwei Faktoren bestimmt wird, nämlich die Einschätzung des möglichen Nutzens einer Handlung und die subjektive Beurteilung der Erfolgswahrscheinlichkeit des Handelns. Während kognitiv-handlungstheoretische Konzepte vor allem auf der subjektiven Einschätzung eines möglichen Lernerfolgs beruhen, spielen bei den Theorien auf der Basis einer dynamischen Persönlichkeitskonzeption andere Aspekte, wie etwa die persönliche Bedeutung der Lerninhalte und der jeweiligen Gegenstandsbereiche eine zentrale Rolle (vgl. Krapp 2014). Im Mittelpunkt steht dabei die Entwicklung eines bestimmten Persönlichkeitskonzeptes, aus welchem heraus sich Motivation entwickelt. Die Selbstbestimmungstheorie nach Deci und Ryan (2000), die den theoretischen Hintergrund, der für die vorliegende Untersuchung entwickelten Fragebögen zur Motivation liefert, zählt hier dazu (Krapp & Seidel 2014, S. 203).

11 Bei der Leistungsmotivation geht es laut Krapp (2014, S. 194) darum, „die eigene Tüchtigkeit in bestimmten Tätigkeiten zu steigern“.

3.2.4.2 Motivationale Faktoren im bilingualen Unterricht

Apelt (1981, S. 53) benennt zentrale Motive in Bezug auf das Erlernen einer Fremdsprache im Unterricht. Die von Apelt (1981) entwickelte Liste mit Motiven zum Fremdsprachenlernen erscheint auch in Bezug auf den bilingualen Unterricht gewinnbringend, denn die Bereitschaft, eine Fremdsprache im Sachfachunterricht zu benutzen, ist eng mit den Motiven verbunden, welche die Bereitschaft zum Erlernen einer Fremdsprache steuern. Zuerst ist hier dabei das Nützlichkeits- bzw. das Berufsmotiv zu nennen. Aus Sicht von Schülern und Eltern bedeutet bilingualer Unterricht nämlich in erster Linie, die Möglichkeit Fremdsprachenkenntnisse zu vertiefen. Damit verbunden ist die Hoffnung, den Herausforderungen eines europäischen Arbeitsmarktes und eines europäischen Bildungswesens (*Bologna-Prozess*) besser gewachsen zu sein. Auch das Neugier-Motiv spielt bei bilingualen Modulen eine entscheidende Rolle. Schüler wollen wissen, wie der Unterricht in dem ihnen bekannten Sachfach in der Fremdsprache abläuft und ob ihre Fremdsprachenkenntnisse ausreichen, die Lernziele zu erreichen, die bekanntlich jenen im Regelunterricht entsprechen. Damit wird neben dem Wissens- auch das Geltungsmotiv wirksam, da bilingualer Unterricht gemeinhin als schwer gilt und vielfach noch die Vorstellung vorherrscht, dass diese Unterrichtsform insbesondere für leistungsstarke Schüler gedacht ist. Der gemeinhin wahrgenommene hohe kognitive Anspruch bilingualen Unterrichts, der für die Schüler neben den sachfachlichen auch sprachliche Hürden bereithält, kann sich positiv auf die Leistungsmotivation der Schüler auswirken. In seiner Theorie zur Leistungsmotivation geht Atkinson davon aus, dass die Leistungsbereitschaft maßgeblich von zwei gegensätzlichen Motiven, nämlich der *Hoffnung auf Erfolg* und der *Angst vor Misserfolg* beeinflusst wird (Raynor, Brown, & Atkinson, 1974). Beide Motive können nach Heckhausen (1963), der die Theorie im deutschsprachigen Raum geprägt hat, eine erhöhte Anstrengung und eine intensive Auseinandersetzung zu Folge haben. Die Angst vor Misserfolg kann jedoch auch Vermeidungstendenzen zur Folge haben, die dazu führen können, dass der Lerner sich der Lern- sowie der Prüfungssituation entzieht. Das Ausmaß dieser beiden Tendenzen (Hoffnung auf Erfolg und Angst vor Misserfolg) ergibt sich dabei daraus, wie stark das jeweilige Motiv beim Lerner ausgeprägt ist, dem persönlichen Wert, dem er Erfolg bzw. Misserfolg beimisst, sowie der subjektiven Einschätzung der Wahrscheinlichkeit von Erfolg und Misserfolg (Atkinson J. , 1975, S. 393). Entsprechend unterscheidet man zwischen *erfolgsszuversichtlichen und misserfolgsängstlichen Lernern* (Krapp & Seidel, 2014, S. 201). Bei misserfolgsängstlichen Schülern kann bilingualer Unterricht Angst erzeugen, die sich negativ auswirkt, insbesondere, wenn die Schüler über ein negatives *Fähigkeitsselbstkonzept* in der Fremdsprache verfügen. Das Fähigkeitsselbstkonzept beruht auf der subjektiven Einschätzung der eigenen Fähigkeiten. Es ist häufig fachspezifisch ausdifferenziert und enthält auch Informationen über Vorlieben und

Abneigungen. Dieses "mentale Modell der Person von sich selbst" entwickelt sich, wenn die eigenen Leistungen im Vergleich zu den Leistungen anderer beurteilt werden (Krapp & Seidel, 2014, S. 201). Es führt im Laufe der Entwicklung einer Person zu stabilen Annahmen, die das Entstehen von Motivation steuern können. Das Selbstkonzept spielt auch bei der Einteilung in *naturwissenschaftlich* und *fremdsprachlich orientierte Schüler* eine Rolle. Häufig betrachten Schüler sich entweder als naturwissenschaftlich oder als sprachlich begabt und richten ihre fachlichen Interessen entsprechend aus. In der vorliegenden Untersuchung soll die Auswirkung bilingualen Unterrichts auf solche Lerner untersucht werden. Aus der besonderen Rolle, die der Fremdsprache im Fremdsprachenunterricht und in besonderer Weise im bilingualen Unterricht zukommt, ergeben sich weitere Lernmotive. Die Fremdsprache ist nämlich nicht nur Lerngegenstand, sondern gleichzeitig Medium der Kommunikation und aufgrund ihrer sozialen und kulturellen Komponenten eng mit der Identität des Lernalters verbunden (Weskamp, 2001, S. 31). Damit unterscheidet sich das Lernen einer Fremdsprache grundlegend vom Lernen in anderen Fächern. Es geht nämlich nicht nur um das Erlernen bestimmter Fertigkeiten und Regeln, sondern um die Übernahme neuer sozialer und kultureller Verhaltensweisen und die damit verbundene Veränderung und Erweiterung der eigenen Persönlichkeit (Williams & Burden, 1997, S. 115). Die Einstellung des Lernalters zur Sprechergemeinschaft der Fremdsprache erweitert somit den Katalog möglicher Motive. Man spricht in diesem Zusammenhang auch von *integrativen Motiven*. Integrative Motivation liegt vor, wenn eine Sprache gelernt wird, um Kontakt zu den Sprechern dieser Sprachgemeinschaft aufzunehmen und Teil dieser Gemeinschaft zu werden. Dabei stellt insbesondere das Erlernen des Englischen bzw. des amerikanischen Englisch als Verkehrssprache einer globalen Jugendkultur einen besonderen Reiz für Jugendliche dar, der dazu beitragen kann, Schüler auch zur Kommunikation über chemische Inhalte zu motivieren.

3.2.4.3 Motivation in den Naturwissenschaften: Gendereffekte

In Bezug auf die Naturwissenschaften, die sich eher selten (mit Ausnahme des Fachs Biologie) in den bilingualen Angeboten der Schulen wiederfinden, herrscht bei Verantwortlichen sowie bei Schülern und Eltern oft Zweifel, ob die oft als „schwer“ betrachteten Inhalte sich unter Verwendung der Fremdsprache überhaupt angemessen vermitteln lassen (siehe Abschnitt 2.4). Zudem finden sich unter den naturwissenschaftlich-mathematisch interessierten Schülern häufig auch solche, die sich für sprachlich weniger begabt halten und sich entsprechend weniger für Sprachen interessieren. Es gibt Hinweise darauf, dass sich unter diesen naturwissenschaftlich-mathematisch orientierten Schülern verhältnismäßig viele Jungen befinden und es gibt deshalb Grund zur Sorge, dass diese durch einen überwiegend fremdsprachlich geführten Unterricht demotiviert werden (Verriere, 2014,

S. 242f.). Auf der anderen Seite ist bilingualer Unterricht in den Naturwissenschaften aber auch mit der Hoffnung verbunden, die fremdsprachlich interessierten Schüler, die sich nur wenig für die mathematisch-naturwissenschaftlichen Fächer interessieren, durch die Möglichkeit zum Gebrauch der Fremdsprache für diese Sachfächer zu motivieren. Hier gibt es Hinweise, dass zu dieser Gruppe verhältnismäßig viele Mädchen gehören (Verriere, 2014, S. 242f.). Zahlreiche Studien belegen etwa, dass Mädchen über ein niedrigeres Selbstkonzept und weniger Interesse in den Fächern Mathematik, Chemie und Physik verfügen (Jansen, Schroeders, & Stanat, 2013, S. 357f.). Die Hoffnung, Mädchen durch bilingualen Unterricht für die Naturwissenschaften stärker zu begeistern, wurde in der Literatur deshalb vielfach geäußert (vgl. z.B. Bosenius, 2009, Maxis-Gehrke & Bonnet, 2001). Wie es zur Ausbildung dieser gegensätzlichen Interessen im Bereich der sprachlich und mathematisch-naturwissenschaftlichen Fächern kommt, erklärt das *Internal/External-Frame-of-Reference-Modell* von Marsh (1986). Dabei geht man davon aus, dass intraindividuelle Vergleiche der Lerner in Bezug auf ihre Leistungen in den Fächern des sprachlichen bzw. des mathematisch-naturwissenschaftlichen Aufgabenfeldes, Auswirkungen auf ihr Selbstkonzept in diesen Fächern haben. Lerner mit guten Leistungen in mathematisch-naturwissenschaftlichen Fächern werten ihr Selbstkonzept in diesen Fächern auf, gleichzeitig jedoch ihr Selbstkonzept in den Sprachen ab. Entsprechend entwickelt sich bei Lernern mit guten Leistungen in den Sprachen in diesen Fächern ein positives Selbstkonzept, während für die mathematisch-naturwissenschaftlichen Fächer das Gegenteil der Fall ist. Solche Kontrasteffekte betreffen allerdings nicht nur für das Kompetenzerleben, sondern auch affektive Faktoren, wie Freude, Angst oder Interesse. Sie sind umso größer, je weiter die Fächer im Fächerkanon auseinander liegen, sie spiegeln das stereotype Bild vom sprachlich bzw. mathematisch begabtem Schüler wider (Piesche, 2016, S. 34ff.). Zu diesem Bild gehört auch, dass sprachliche Begabung eher ein weibliches, und naturwissenschaftliche Begabung eher ein männliches Attribut ist. Zahlreiche Untersuchungen zum bilingualen Unterricht in den Naturwissenschaften widmen sich deshalb gezielt solchen Gendereffekten (z.B. Piesche 2016, Verriere 2014, Rodenhauser 2016).

3.3 Gestaltung bilingualen Chemieunterrichts

Ausgehend von den in Abschnitt 3.2 geschilderten Theorien zum Lernen im bilingualen Unterricht werden im folgenden Abschnitt zentrale Elemente unterrichtsmethodischer Praxis beschrieben, die dazu beitragen, die oben geschilderten Lernprozesse zu initiieren. Wie bilingualer Chemieunterricht gestaltet werden sollte, der einen konstruktivistisch angelegten Lernprozess anregt, wird in Abschnitt 3.3.1 erläutert. Dabei treten neben die Erkenntnisse aus dem Bereich der Didaktik des bilingualen Unterrichtes insbesondere wichtige Impulse aus der Didaktik des Chemieunterrichts.

Der darauffolgende Abschnitt beschreibt Diskursfunktionen, deren Schulung zur Förderung höherer Denkopoperationen beiträgt (Abschnitt 3.3.2). Die sich anschließenden Abschnitte sind dem (fremd-)sprachlichen Lernerfolg und der Frage gewidmet, wie doppelte Sachfachliteralität erreicht werden kann und welche Rollen dabei der Mutter- und der Fremdsprache zukommen (Abschnitt 3.3.3). Im letzten Abschnitt (Abschnitt 3.3.4) wird schließlich der Wortschatz in den Blick genommen, der im Mittelpunkt der Erfassung des (fremd-)sprachlichen Lernerfolges steht.

3.3.1 Der Konstruktivismus und seine Konsequenzen für die Unterrichtsgestaltung

[...] im Bereich der Lernmethoden sind für den bilingualen Chemieunterricht keine gänzlich neuen Methoden nötig, sondern vielmehr ein Schulterschluss der Lehrmethoden aus der Fremdsprachenvermittlung und der Chemiedidaktik bzw. eine Inkorporation der entsprechenden Methoden aus dem Bereich der (Fremd-)Sprachdidaktik mit in den (bilingualen) Sachfachunterricht hinein (Bohrmann-Linde, 2012, S. 196).

Sowohl die Chemie- als auch die Fremdsprachendidaktik halten zentrale Elemente für eine Didaktik des bilingualen Chemieunterrichts bereit. Diese müssen jedoch sorgfältig ausgewählt und aufeinander abgestimmt werden.

Beiträge aus der Chemiedidaktik

Konstruktivistische Lerntheorien haben auch die Chemiedidaktik geprägt. Laut Ferber (2014, S. 4) gilt auch für den Chemieunterricht, "dass Lern- und Kompetenzentwicklungen begünstigt werden, wenn Lernumgebung und -methode aktives; konstruktives Lernen in sozialen Kontexten, in konkreten Situationen und unter Berücksichtigung des historischen und kulturellen Kontextes ermöglichen". Riemeier (2007, S. 70f.) nennt fünf Merkmale einer konstruktivistischen Sichtweise des Lernens. Lernprozesse sind demnach konstruktiv, selbstdeterminiert, individuell, sozial und situiert. Damit wird deutlich, dass Wissen auf der Basis von vorhandenem Vorwissen aktiv und selbstbestimmt, d.h. von außen allenfalls angeregt, nicht jedoch gesteuert abläuft. Das Lernen gestaltet sich gemäß den individuellen Voraussetzungen der Lerner unterschiedlich, wobei das neu erworbene Wissen an die Situation geknüpft ist, in der es gelernt wurde (vgl. auch Piesche 2016, S. 48). Für die Naturwissenschaft bedeutet konzeptuelles Lernen, dass vorhandene Alltagskonzepte durch Fachkonzepte ersetzt werden. Die Bereitschaft zum Konzeptwechsel setzt voraus, dass Lerner mit ihren Alltagskonzepten unzufrieden sind und diese durch verständliche, plausible und für die weitere Arbeit nützliche Konzepte ersetzt werden können (Posner, Strike, Hewson, & Gertzog, 1982).

Im Unterricht geht es deshalb immer wieder darum, kognitive Konflikte zu erzeugen, die einen Konzeptwechsel initiieren (Barke & Harsch, 2001, S. 38ff.). Der Nutzen neuer Konzepte kann durch konsequenten Alltagsbezug spürbar werden. Der naturwissenschaftliche Unterricht ist deshalb oft problemorientiert angelegt und fordert die Schüler dazu heraus, fachliches Wissen zielgerichtet zur Lösung alltagsrelevanter Probleme einzusetzen. Besonders ausgeprägt ist dieser Lebensweltbezug im Ansatz *Chemie im Kontext* (Demuth, Gräsel, Parchmann, & Ralle, 2008). Ausgangspunkt einer Unterrichtseinheit basierend auf dieser Unterrichtskonzeption ist ein lebensweltlicher Kontext, anhand dessen die Bedeutung der Chemie für das tägliche Leben deutlich wird. Dabei werden chemische Fachinhalte entwickelt und auf gemeinsame Basiskonzepte zurückgeführt, an welchen sich die Konstruktion chemiespezifischen Fachwissens orientiert (Demuth, Gräsel, Parchmann, & Ralle, 2008). Forschendes Lernen steht für einen Ansatz, bei dem Schüler aus Wissenselementen, die in experimentellen Phasen gewonnen werden, ihr Wissen aktiv konstruieren. Schüler lernen dabei, Experimente planvoll einzusetzen und die Prinzipien chemischer Forschung zu verstehen. Im Sinne eines wissenschaftspropädeutischen Unterrichts kommt dem Prozess der naturwissenschaftlichen Erkenntnisgewinnung dabei eine zentrale Bedeutung zu (vgl. auch Rodenhauser & Preisfeld, 2016, S. 222f.). Im weitverbreiteten Modell nach Popper werden Hypothesen auf der Grundlage experimenteller Befunde generiert (induktives Vorgehen) und anschließend durch zusätzliche Experimente verifiziert oder falsifiziert (deduktives Vorgehen). Die Erkenntnisse führen zu neuen bzw. modifizierten Hypothesen, die anschließend ebenfalls experimentell überprüft und weiter präzisiert werden (*scientific method*). Das hypothetisch-deduktive Vorgehen ist typisch für das entdeckende Lernen. Schüler erwerben dadurch nicht nur grundlegende Kenntnisse im Bereich des prozeduralen Wissens, sondern erkennen auch, wie fachliches Wissen in den Naturwissenschaften gewonnen wird. Damit sind sie in der Lage, den eigenen Wissenserwerb zunehmend selbstständig zu steuern und dabei auch den zugrunde liegenden Erkenntnisweg zu hinterfragen (Pfeifer, Häuseler, & Lutz, 1992, S. 138).

Beiträge aus der Fremdsprachendidaktik

Das Primat des Sachfachs besagt, dass sachfachliche Ziele und Inhalte im Vordergrund bilingualen Unterrichts stehen. Damit greifen für den bilingualen Unterricht in besonderem Maße die Wirkungsmechanismen eines auf echter Kommunikation aufbauenden Spracherwerbs. Anders als im Fremdsprachenunterricht entwickeln die Lerner ein *echtes* Verstehens- und Mitteilungsbedürfnis und erwerben so, ähnlich wie im natürlichen Spracherwerb, dann sprachliches Wissen, wenn der Input verständlich ist (vgl. Abschnitt 3.2).

Insbesondere authentische Texte tragen dazu bei, die Mitteilungsorientierung der unterrichtlichen Kommunikation bewusst werden zu lassen und gleichzeitig auch kulturelle Besonderheiten zu vermitteln. Im Chemieunterricht bieten sich dazu etwa Auszüge aus englischen Lehrwerken oder kurze Videosequenzen an. Hier zeigt sich jedoch ein vielfach benanntes Kernproblem bilingualen Unterrichts, nämlich die Diskrepanz zwischen den inhaltlich-kognitiven Ansprüchen des Sachfachs und dem eingeschränkten sprachlichen Leistungsvermögen der Schüler (Wildhage, 2000, S. 218). Die selbstständige Erarbeitung von Unterrichtsinhalten in der Fremdsprache setzt deshalb eine sorgfältige Auswahl und Aufbereitung der eingesetzten Medien voraus. Input sollte im Sinne der Vygotskyschen *zone of proximal development* an vorhandene Wissensbestände anknüpfen, reichhaltig sein, sprachlich und sachfachlich jedoch nur so viel an neuer Information enthalten, wie vom Lerner selbstständig verarbeitbar ist. Da authentische Texte je nach Sprachstand der Lerner zu schwer sind, müssen diese in der Regel überarbeitet werden. Coyle et al. (2012, S. 92ff.) machen dazu eine Reihe von Vorschlägen, welche die Bereiche Lexik, Struktur und Textart betreffen. So schlägt er vor, komplexe Ausdrücke, die nicht zentral für die zu erlernende Fach- und Arbeitssprache sind, durch weniger komplexe zu ersetzen. Außerdem kann die Umwandlung der Textart, beispielsweise von einem anspruchsvollen Fließtext in eine Tabelle, welche relevante Informationen sprachlich verkürzt darstellt, zur sprachlichen Entlastung und zur inhaltlichen Fokussierung beitragen (Coyle, Hood, & Marsh, 2012, S. 96f.)¹². Gezieltes *scaffolding* und die Bereitstellung geeigneter Methoden zur Unterstützung rezeptiver Fertigkeiten helfen dabei, die fremdsprachliche Hürde immer selbstständiger zu überwinden. Thürmann (2013, S. 239) empfiehlt deshalb eine insgesamt verlangsamte Sprechweise, das Gewähren von *sufficient waiting time*, eine klare Artikulation sowie den Verweis auf außersprachliche Informationsquellen, wie Objekte oder Bilder. Für Leisen (2013, S. 157f.) stellt der Wechsel der Darstellungsform den "didaktischen Schlüssel zum fachlichen Verstehen" dar, da sich die Ebenen wechselseitig ergänzen und "was sie [Schüler, F.H.] in der einen Darstellungsform nicht verstehen, erschließt sich ihnen vielleicht in einer anderen"¹³.

12 Thürman (2013, S. 230) warnt jedoch vor solchen defensiven Strategien, bei denen die Anforderungen zeitweise oder dauerhaft gesenkt werden, weil die sprachliche Komplexität im rezeptiven und im produktiven Bereich gesenkt wird. Vielmehr empfiehlt er offensive Strategien, welche sich dadurch auszeichnen, dass die fachinhaltlichen Anforderungen beibehalten werden, die Schüler aber mit dem nötigen funktional-sprachlichen Rüstzeug ausgestattet werden, diesen Anforderungen zunehmend selbstständig gerecht zu werden.

13 Vorschläge für einen planvollen Wechsel der Repräsentationsform in den Naturwissenschaften finden sich bei Stäudel (2008, S. 167).

Schüler lernen so, Informationen aus dem außersprachlichen Kontext zunehmend selbstständig zu erschließen und zur Unterstützung des eigenen Verstehensprozesses zu verwenden. In schriftlichen Unterrichtsphasen, in denen kontinuierliche, wie diskontinuierliche Texte zum Einsatz kommen, kommt der Schulung methodischer Fertigkeiten große Bedeutung zu. Dazu zählen z.B. die Arbeit mit Wörterbüchern, der gezielte Rückgriff auf unterschiedliche Lesetechniken bzw. Lesestrategien, die Informationsentnahme und ihre Aufbereitung durch Textmarkierungen sowie das Anfertigen strukturierter Notizen in Form von *mind maps* oder Tabellen (vgl. Thürmann, 2013, S. 241). Kennzeichnend für den bilingualen Unterricht dabei ist, dass das *scaffolding* im Rahmen curricularer Zielsetzungen planvoll erfolgt. Es ist Bestandteil zielorientierter Unterrichtsvorbereitung und erfolgt keinesfalls zufällig. Gesteuerte Spracharbeit im bilingualen Unterricht bedeutet für Thürmann (2013, S. 239) curriculares-systemisches *scaffolding*, während Immersion auf anlassbezogene, spontane Unterstützung setzt¹⁴. Natürlich spielt auch das anlassbezogene *scaffolding* im bilingualen Unterricht eine wichtige Rolle, da der Unterricht offen und schülerorientiert angelegt ist. Der Unterrichtsverlauf und die damit nötigen sprachlichen Unterstützungsmaßnahmen sind also nicht immer vollständig planbar. Die Konstruktion von Wissen erfolgt in schulischen Lernsituationen jedoch nicht ausschließlich im Rückgriff auf Materialien, sondern vor allem im Austausch mit den Mitschülern und dem Lehrer. Diesem sozial-konstruktivistischem Ansatz kommt im bilingualen Unterricht eine besondere Bedeutung zu, weil sich sowohl sachfachliches als auch sprachliches Wissen im bilingualen Unterricht über das Aushandeln von Bedeutungen und Form vollzieht. Lerner können so neue Konzepte sowohl im Hinblick auf ihre Bedeutung als auch im Hinblick auf ihre sprachliche Form an bereits vorhandenes und individuell unterschiedlich ausgeprägtes Vorwissen anknüpfen. Mit dem Aushandeln sprachlicher Bedeutung geht damit auch ein Aushandeln sprachlicher Form einher. Damit solche Aushandlungsprozesse immer wieder im Unterricht stattfinden können, müssen Lehrer entsprechende Voraussetzungen schaffen. Dies betrifft die Wahl der Sozialform aber auch die Bereitstellung entsprechender Redemittel. Der Erfolg solcher Prozesse zeigt sich im Output der Lerner, die daher immer wieder in schriftlichen und mündlichen Phasen zur Sprachproduktion angehalten (*Outputorientierung*) werden. Entsprechende Unterstützungsangebote (*output scaffolding*) richten sich dabei nach den sprachlichen, kognitiven und kommunikativen Anforderungen der jeweiligen Lernaufgabe. Coyle et al. (2012, S. 105ff.) greifen hier auf

14 Für die Unterscheidung zwischen curricular-systemischem und anlassbezogenem *scaffolding* finden sich in der Literatur zahlreiche andere Bezeichnungen, wie *designed-in scaffolding* vs *point-of-need scaffolding* (Sharpe, 2001, S. 32) oder *hard scaffolds* vs *soft scaffolds* (Brush & Saye, 2002, S. 2).

Erkenntnisse aus dem klassischen Fremdsprachenunterricht zurück und nennen u.a. *questions*, *vocabulary headers*, *gap sentences*, *sentence starters*, *spider diagrams* und *substitution tables*. Für den Bereich der mündlichen Sprachproduktion verweist Thürmann (2013, S. 230) auf die Bedeutung des Lehrerfeedbacks z.B. durch Formulierungshilfen, die der Lehrer im Rahmen von *bridging* oder *prompting* bereithält, wobei er sowohl die inhaltliche als auch die sprachliche Seite der Schüleräußerung erfasst. Auch durch eine *laute Denksprache*, bei der der Lehrer sein Denken und Handeln beschreibt und kommentiert, kann ein zusätzliches Angebot an sprachlichen Mitteln bereitgestellt werden. Sicherungsphasen kommt im bilingualen Unterricht eine besondere Bedeutung zu, da es aufgrund der Verwendung einer nicht vollständig beherrschten Fremdsprache sehr leicht passieren kann, dass Schüler dem Unterricht nicht mehr folgen können. Die Gründe sind vielfältig und können individuelle Vokabellücken, mangelnde Aufmerksamkeit oder ein insgesamt zu hohes Sprechtempo des Lehrers sein. Sicherungsphasen sollten deshalb häufiger als im regulären Unterricht eingeplant werden (Richter & Zimmermann, 2003, S. 120). Dabei tragen insbesondere schriftliche Phasen zu einer Entschleunigung des Unterrichtstempos bei und entlasten das Hörverstehen durch die parallele Bereitstellung schriftlicher Informationen. Hier eignen sich klassische Methoden wie Tafelanschrieb oder der Einsatz von Arbeitsblättern aber auch kreative oder spielerische Methoden, wie Taboo, Rätsel oder Fantasiegeschichten (vgl. Bohn und Doff, 2010, S. 82). An weiteren Auflistungen mit Merkmalen guten bilingualen Unterrichts mangelt es in der Literatur nicht. Holm (2013, S. 174f.) nennt etwa die stärkere Visualisierung, den umfassenderen Einsatz von Medien, den häufigeren Methodenwechsel, die intensivere Erklärung und Umwälzung und eine insgesamt höhere Fokussierung und Stringenz der Inhalte (vgl. auch Piesche 2016, S. 23). Abschließend sei hier auf die recht umfangreiche Darstellung von Mehisto et al. (2008, S. 29) hingewiesen, die insgesamt 30 *core features* einer Methodik bilingualen Unterrichts nennen.

3.3.2 Diskursfunktionen einüben

Chemiespezifische Diskursfunktionen sind im Rahmen der vorliegenden Untersuchung insofern von Bedeutung, als sie eng mit bestimmten Denkopoperationen verknüpft sind. Die Möglichkeit zur Schulung höherer Denkopoperationen durch den Einsatz der Fremdsprache wird im Rahmen der vorliegenden Studie untersucht. Das aus den Bildungsstandards entwickelte Kompetenzmodell und die sich aus diesem Modell ergebenden Aufgaben zur Lernerfolgsüberprüfung, wie sie in den vorliegenden Sachfachtests zum Einsatz kommen, greifen auf die bekannten Taxonomien kognitiver Leistungen zurück. Diese sind eng mit dem von Zydatiś (2002) entwickelten Konzept der Diskursfunktionen verbunden. Diskursfunktionen beschreiben auf Makroebene textgebundene

Denkoperationen, die typischerweise bei der Bewältigung fachspezifischer Aufgaben zum Tragen kommen. Sie umfassen, anders als die Mikrofunktionen, längere Diskursabschnitte wobei auch rhetorische und textlinguistische Aspekte zum Tragen kommen (Dalton-Puffer C. , 2013, S. 140)¹⁵. Fachspezifische Diskursfunktionen lassen sich beispielsweise aus den in den Curricula der Länder formulierten Kompetenzerwartungen isolieren. Diese Vorgehen wählen auch Vollmer (2011) und Zydati (2013). Die von Vollmer (2011) aufgestellte Liste bietet sich fr die vorliegende Untersuchung zum bilingualen Chemieunterricht an, weil sie anhand der Curricula der Fcher Mathematik, Biologie und Geschichte entwickelt wurde.

1. Benennen, Definieren
2. Beschreiben, Darstellen
3. Berichten, Erzhlen
4. Erklren, Erlutern
5. Bewerten, Beurteilen
6. Argumentieren, Stellung beziehen

Tabelle 3: Diskursfunktionen nach Vollmer (2011)

Vergleicht man Vollmers (2011) Liste mit den Operatoren zur Erstellung von Abituraufgaben im Fach Chemie in Nordrhein-Westfalen auf der Grundlage der einheitlichen Bildungsstandards fr die Naturwissenschaften, so ergeben sich viele Parallelen (MSW-NRW, 2015). Die folgende Tabelle zeigt hierzu eine entsprechende Auswahl.

¹⁵ In der Literatur finden sich dazu unterschiedliche Klassifizierungen. Gem ihrer Funktionalitt unterscheidet Thrman (2010, S. 147) sechs Diskursfunktionen: *naming/defining, describing, reporting, explaining, evaluating und arguing/taking a stance*. Dalton-Puffer (2004, S. 26) hingegen identifiziert insgesamt vierzehn und nennt diese *academic language functions*.

Diskursfunktionen laut Vollmer (2011)	Operatoren für das Zentralabitur im Fach Chemie in Nordrhein-Westfalen laut Kernlehrplan
benennen, definieren	nennen, angeben
beschreiben, darstellen	beschreiben, darstellen, dokumentieren, skizzieren
berichten, erzählen	
erklären, erläutern	erklären, erläutern, begründen, deuten, interpretieren,
bewerten, beurteilen	beurteilen, bewerten
argumentieren, Stellung beziehen	diskutieren, Stellung nehmen

Tabelle 4: Diskursfunktionen und Operatoren für das Fach Chemie

Die hier beschriebenen Makrofunktionen stellen komplexe kognitiv-sprachliche Einheiten dar, die, wenn sie von Schülern eingelöst werden, längere mündlich oder schriftlich artikuliert Beiträge einfordern. Im Unterricht der Sekundarstufe I treten sie bei längeren Äußerungen im Rahmen von Diskussionen oder bei Vorträgen in Erscheinung. Offene, schriftlich zu bearbeitende Aufgaben können dazu dienen, sowohl die inhaltlich-kognitive als auch die sprachliche Seite solcher Diskursfunktionen zu schulen. Für den Unterricht in der Sekundarstufe I sind dabei neben den dargestellten Makrofunktionen auch die Mikrofunktionen von Bedeutung. Diese sind kurz und oft an bestimmte Satzstrukturen gebunden oder durch entsprechende Ergänzungen markiert. Zydati (2002, S. 48) nennt 23 solcher Diskursfunktionen, die mit bestimmten kognitiven Prozessen verbunden sind. Teilweise ordnet er diesen dabei auch englischsprachige Realisierungsmöglichkeiten zu (sprachliche Exponenten). Dalton-Puffer (2013) stellt in ihrer Untersuchung zur Praxis bilingualen Unterrichts fest, dass viele Diskursfunktionen im bilingualen Unterricht zu selten realisiert werden und fordert deshalb, wie auch Lose (2007), an dieser Stelle explizite Spracharbeit. Lose (2007, S. 104ff.) schlägt beispielsweise aus dem Fremdsprachenunterricht bekannte Satzverbindungsübungen vor. Dabei müssen die Schüler zu verschiedenen Satzanfängen die passende Fortführung auswählen und aus einer Liste eine Konjunktion ergänzen (z.B. *because, due to, since, as*), um den logischen Charakter des

Begründungszusammenhangs zu markieren. Lose weist jedoch darauf hin, dass die Sätze inhaltlich eng auf die unterrichtliche Thematik bezogen sein müssen, um nicht den Status einer reinen Sprachübung zu bekommen. Auch in der vorliegenden Unterrichtsreihe kommen verschiedene schriftliche Übungen zu Diskursfunktionen zum Einsatz (vgl. Abschnitt 5.5.1).

3.3.3 Rollen von Fremd- und Muttersprache

Lernerfolge können nur dann sinnvoll miteinander verglichen werden, wenn klar ist, welchen Anteil die Fremd- und Muttersprache im Unterricht hatten. Die Frage muss also in Bezug auf das hier vorliegende Modul präzisiert werden. Über die Rolle, welche die Mutter- bzw. Schulsprache im bilingualen Unterricht spielt, besteht jedoch nach wie vor wenig Einigkeit. Dies betrifft nicht nur den zeitlichen Umfang, sondern insbesondere die Funktion, die ihr im Rahmen eines integrierten Inhalts- und Sprachlernens zukommt. So räumen Coyle et al. (2012, S. 16) zwar ein, dass der Gebrauch der Muttersprache im Sinne eines *translanguaging* sinnvoll sei, führen jedoch nicht aus, wie dieser erfolgen kann. Laupenmühlen (2012, S. 167f.) registriert in den ministeriellen Vorgaben einzelner Bundesländer eine enorme Bandbreite, vom "monolingual-fremdsprachigen bis zum zweisprachigen Unterricht mit dem Ziel der doppelten Sachfachliteraltät".

Während bilingualer Unterricht lange Zeit als monolingual in der Fremdsprache geführter Sachfachunterricht praktiziert wurde, hat sich mittlerweile die Erkenntnis durchgesetzt, dass auch die Schulsprache einer gezielten Förderung bedarf, nicht zuletzt, weil die geforderte doppelte Sachfachliteraltät (vgl. Vollmer, 2005) eben auch eine angemessene Ausdrucksfähigkeit in dieser Sprache bedeutet. Diehr (2012, S. 17ff.) unterscheidet in Bezug auf die Rolle der Muttersprache drei Typen bilingualen Unterrichts. Dabei verwendet sie den Begriff BU synonym mit CLIL und grenzt beide vom Konzept der Immersion ab, bei der die Fremdsprache den Status einer Zweitsprache hat, die auch außerhalb der Schule von den Schülern benutzt wird. Beim Typ A nach Diehr dient die Fremdsprache als das bestimmende Medium des Lernens, welches im Idealfall ohne Rückgriffe auf die Muttersprache auskommt. Sinnvoll erscheint der Gebrauch der Muttersprache bei sprachlichen Schwierigkeiten im Anfangsunterricht bzw. zur parallelen Einführung der muttersprachlichen Fachterminologie. Während beim bilingualen Unterricht Typ A keine nennenswerten Vorteile im Einbezug der Muttersprache gesehen wird, geht man beim Typ B davon aus, dass der systematische und planvolle Einbezug der Muttersprache sich vorteilhaft auf das Lernen auswirkt. Die Fremdsprache wird als Leitsprache gesehen und ist die dominante Sprache. Die Muttersprache leiste jedoch wichtige "Zubringerdienste" und kann funktional zum Erreichen der Ziele bilingualen Unterrichts eingesetzt werden (Diehr, 2012, S. 25). Ein unreflektierter ständiger Sprachwechsel (*language hopping*) wird jedoch abgelehnt. Butzkamm (2000, S. 105f.)

schlägt hierzu "in den Unterricht integrierte, kurzfristige muttersprachliche Hilfestellungen" im Sinne einer *Pendelstrategie* vor. Der Sprachwechsel erfolgt dabei mit dem Ziel, die fremdsprachige Kommunikation aufrechtzuerhalten. Butzkamm sieht insgesamt "sieben Formen der planvollen Mitbenutzung der Muttersprache" vor. Diese betreffen die Möglichkeit der Schüler auf Deutsch zu antworten, ein deutsches Schulbuch zur häuslichen Arbeit zu benutzen, in eingeschobenen Unterrichtsphasen das Verständnis im muttersprachlich geführten Gruppengespräch zu klären und die Nutzung nachgereichter, zweisprachiger Wort- und Phrasenlisten. Beim Typ C werden Schul- und Fremdsprache als gleichwertig betrachtet. Dem Erwerb der schulsprachlichen Diskursfähigkeit wird große Bedeutung beigemessen, wobei davon ausgegangen wird, dass Fachbegriffe auch auf Deutsch konstruiert werden müssen (Diehr, 2012, S. 27).

3.3.4 Wortschatzarbeit

Die Einteilung des Wortschatzes orientiert sich im bilingualen Unterricht vor allem an der Bedeutung, die ihm bei der Erarbeitung sachfachlicher Inhalte zukommt. Eine aus der angelsächsischen Literatur bekannte Unterscheidung betrifft die Sprache, die für die inhaltliche Arbeit unentbehrlich ist (*content-obligatory language*) und jene Sprache, die zwar hilfreich, jedoch nicht absolut notwendig ist (*content-compatible language*). *Content-obligatory language* umfasst auf lexikalischer Ebene vor allem Fachbegriffe, spezielle Wendungen und Wörter mit mehreren Bedeutungen. Auch Formulierungshilfen zur Verbalisierung zentraler kognitiver Operationen (Diskursfunktionen) oder zur Erarbeitung konzeptuellen Wissens im unterrichtlichen Diskurs (Diskursvokabular) zählen dazu. *Content-compatible language* betrifft die Sprache, die Schüler zur Versprachlichung ihrer Beobachtungen und Erkenntnisse benötigen. Sie sind eher allgemeinsprachlicher Natur und bilden die Grundlage, auf der Konzepterwerb und fachsprachliche Darstellungen beruhen (Mehisto, Marsh, & Frigols, 2008, S. 104f.). Der beiläufige Wortschatzerwerb macht, wie verschiedene Studien zeigen, nur einen geringen Teil des Vokabulars aus und ist „störungsanfällig“ (Kersten, 2012, S. 197). Deshalb darf der Erwerb des für die eigene Sprachproduktion zur Verfügung stehenden Wortschatzes nicht dem Zufall überlassen werden, sondern muss systematisch und planvoll angelegt sein. Nach Nation (2001) gelingt dies dann, wenn die Aufmerksamkeit des Lerner auf den neuen Wortschatz gelenkt wird (*noticing*), wenn Lerner diesem wiederholt und in vertrauten Kontexten begegnen (*retrieval*) und wenn Möglichkeiten zum kreativen Umgang mit den neuen Begriffen geschaffen werden (*creative use*).

Zur Semantisierung kann dabei auf die aus dem Fremdsprachenunterricht bekannten Verfahren zurückgegriffen werden. Die Aktivierung unterschiedlicher Sinneskanäle durch multisensorische Darbietung erweist sich auch im bilingualen Unterricht als vorteilhaft. Dazu zählt neben der

Visualisierung durch Bilder und Objekte auch der gezielte Einsatz von Gestik und Mimik oder die direkte Verknüpfung von Gesagtem mit Gesprochenem nach dem Prinzip der *Total Physical Response*. Diese Möglichkeit ergibt sich z.B., wenn Schüler sprachliche Stimuli, etwa im Rahmen eines naturwissenschaftlichen Experimentes, unmittelbar in konkrete Handlungen übersetzen (Hartmannsgruber, 2014, S. 25). Kersten (2012, S. 198ff.) weist außerdem darauf hin, dass die Verknüpfung eines Wortes mit dem bereits bekannten Wortschatz eine tiefere semantische Verarbeitung bewirkt und sich genauso positiv auf die Behaltensleistung auswirkt wie das Erkennen des persönlichen Nutzens eines Wortes. Wortschatzübungen sollten deshalb auch im bilingualen Unterricht regelmäßig durchgeführt werden (Kersten, 2012, S. 202). Laut Kersten (2012, S. 200ff.) sind drei Prinzipien bei der Durchführung von Übungen, die zu einer tiefen Verarbeitung des Wortschatzes und zu nachhaltigen Behaltensleistungen führen, zu beachten: Erstens sollte der Wortschatz thematisch strukturiert sein, was im bilingualen Unterricht durch die enge Verknüpfung mit sachfachlichen Inhalten oft gegeben ist. Zweitens muss die Wiederholung regelmäßig und in unterschiedlichen Verwendungskontexten stattfinden. Drittens sollten zahlreiche unterschiedliche Übungsformate eingesetzt werden, die sowohl die syntagmatische wie auch die paradigmatischen Beziehungen zwischen den Wörtern verdeutlichen. Für Kersten (2012, S. 200ff.) eignen sich deshalb aus dem Fremdsprachenunterricht bekannte Techniken wie das Anlegen von Wortfeldern, mind maps oder semantischen Netzen, auch Übungen, die Kollokationen oder die Bildung von Komposita betreffen. Die Vermittlung muttersprachlicher Fachbegriffe orientiert sich dabei an dem von Diehr (2016, S. 71) entwickelten Integrated Dynamic Model (IDM) des mentalen Lexikons, das zwischen vier Graden konzeptueller Äquivalenz unterscheidet (siehe Abschnitt 3.2.3.2). Bei vollständiger Konzeptäquivalenz kann der einfache Verweis auf das entsprechende Konzept in der Muttersprache erfolgen, während bei nur partieller Konzeptäquivalenz solche Unterrichtsphasen auch eine ganze Stunde dauern können (Frisch, 2016, S. 96).

3.3.5 Planung bilingualen Unterrichts

Die Planung der Unterrichtsreihe erfolgte nach dem *4C-Modell* von Coyle et al. (2012).

Für den bilingualen Unterricht schlagen Coyle et al. (2012, S. 41ff.) dabei die Bereiche *Content*, *Communication*, *Cognition* und *Culture* als zentrale Planungsbereiche vor, die auch im Zentrum

anderer Planungsmodelle stehen¹⁶. Diese sind eingebettet in einen *context*, der integriertes Sprach- und Inhaltslernen (CLIL) ermöglichen soll.

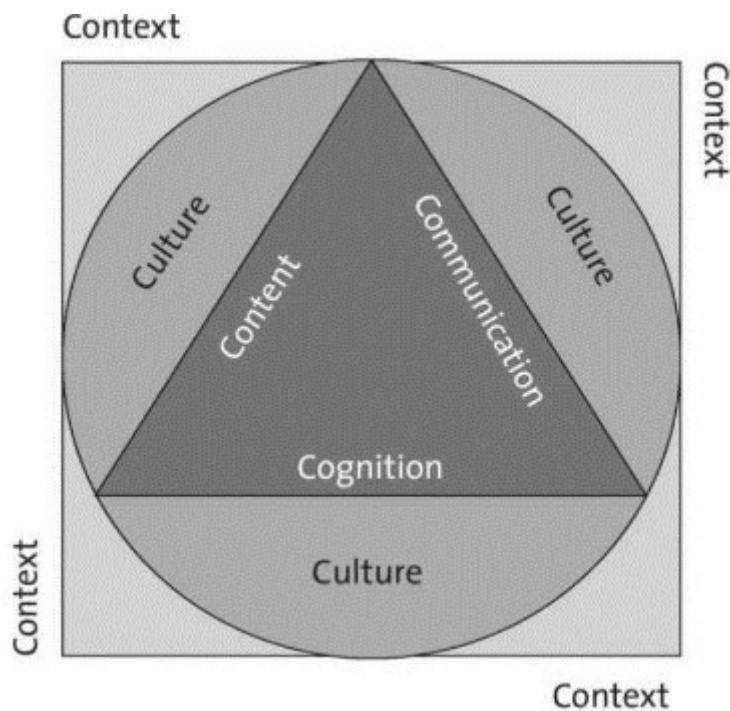


Abbildung 2: Das 4C-Modell

(entnommen aus Coyle et al., 2012, S. 41)

Im bilingualen Unterricht besteht die zentrale Herausforderung darin, eine Brücke zwischen dem geringen Sprachvermögen der Lerner und den hohen kognitiven Anforderungen des Sachfachs zu schlagen. Das von Coyle et al. (2012) vorgeschlagene *Tool Kit* zur Unterrichtsplanung umfasst deshalb noch zwei weitere Werkzeuge, nämlich das *Language Triptych* und die *CLIL-Matrix* (Coyle et al, 2012, S. 49). Das Language Triptych dient dazu, den Spracherwerb systematisch zu planen, indem zwischen *language of / for* und *through learning* differenziert wird. Dazu greift es auf die bereits weiter oben dargestellte Unterscheidung zwischen content-obligatory und content-compatible language zurück (siehe Abschnitt 3.3.4)¹⁷. Die auf Cummings (1984) zurückgehende CLIL-Matrix soll hingegen helfen, sprachliche und kognitive Schwierigkeiten aufeinander

¹⁶ Auch das von Meyer (2010, S. 24) vorgeschlagene *Pyramid-Modell* oder das „heuristische“ Planungsmodell von Otten und Wildhagen (2003, S. 34f.) greifen auf die vier von Coyle et al. formulierten vier Cs als Planungsgrundlage zurück.

¹⁷ Eine detaillierte Darstellung des language triptych findet sich bei Doyle et al. (2012, S. 36ff.).

abzustimmen, um eine angemessene Progression zu ermöglichen, die eine Überforderung der Schüler verhindert¹⁸.

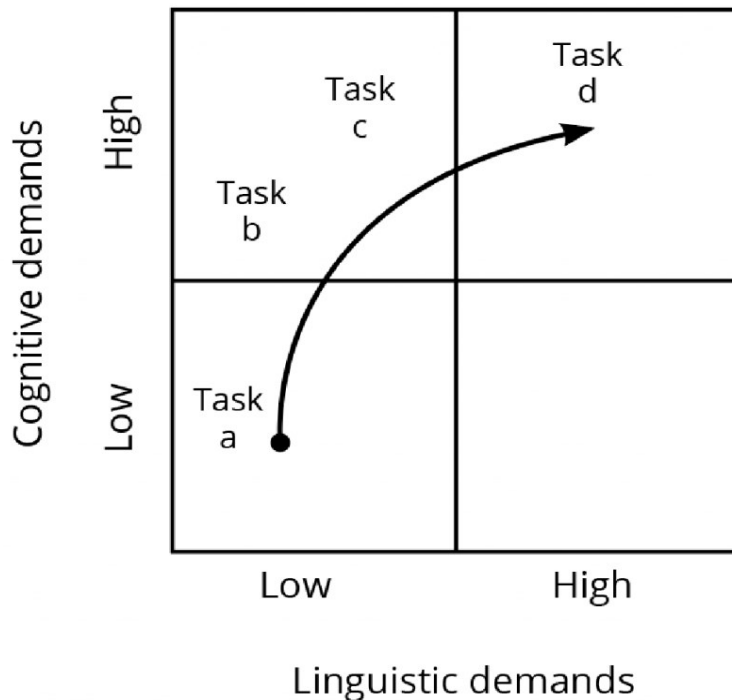


Abbildung 3: Die CLIL-Matrix

(entnommen aus Coyle et al. (2012, S. 68))

Dabei vollzieht sich das Lernen im bilingualen Unterricht idealerweise innerhalb von drei Quadranten der Matrix, die sich in ihrem kognitiven und ihrem sprachlichen Anspruch unterscheiden (vgl. Abbildung oben). Die Aufgaben zu Beginn einer bilingualen Unterrichtsreihe zeichnen sich durch ein geringes sprachliches und kognitives Niveau aus, weil beispielsweise Vorwissen aktiviert und interessante Fragestellungen gesammelt werden (Quadrant 1). Im Rahmen der inhaltlich ausgerichteten Arbeit steigt dann zunächst nur das kognitive Anspruchsniveau (Quadrant 2) und erst im letzten Schritt, etwa in Sicherungs- und Übungsphasen, steigt das sprachliche Niveau an (Quadrant 3). Um den kognitiven Anspruch einzelner Lernaufgaben zu bestimmen und sicherzustellen, dass auch anspruchsvollere Denkopoperationen einbezogen werden, empfehlen auch Coyle et al. (2012, S. 58) auf die Taxonomie von Bloom (1956) bzw. ihre

18 Auch Ting (2013) empfiehlt ein Gleichgewicht zwischen kognitiven und sprachlichen Anforderungen und begründet dies anhand der begrenzten Kapazität des Arbeitsgedächtnisses, das nur einen begrenzten Umfang an neuen Informationen für die parallele Verarbeitung zulässt (vgl. Abschnitt 3.2.2.1).

Erweiterung durch Anderson & Krathwohl (2001) zurückzugreifen und besonderes Augenmerk auf die Formulierung von Lehrerfragen zu legen, durch welche solche Denkprozesse initiiert werden. Coyle (2012, S. 56ff.) rät für die Unterrichtsplanung zu folgendem Vorgehen:

1. Festlegen der Inhalte (content).
2. Zuordnen konkreter Denkoperationen zu den Inhalten (cognition).
3. Analyse der erforderlichen sprachlichen Grundlagen und ihres Erwerbs (communication).
4. Ausgestaltung der Möglichkeiten zu interkulturellem Lernen (culture).

Der Vorteil des 4C-Modells liegt darin, dass es zentrale Planungsfelder abbildet und Wechselbeziehungen erkennbar werden lässt. Es fehlen jedoch konkrete Hinweise zur methodischen Umsetzung. Für letztere schlagen die Autoren vor, eine Theorie aus der Praxis der Unterrichtenden zu entwickeln (*theory of practice*). Lehrer sollen ihren Unterricht deshalb einer ständigen, gemeinsamen Analyse unterziehen (*inquiry-based approach*) und methodische Varianten reflektieren¹⁹. Während sich die dargestellten Modelle auf die Planung von Unterrichtsreihen beziehen, finden sich nur wenige Vorschläge, die Bezug auf die Planung einzelner Stunden nehmen²⁰. Die meisten Autoren schlagen vor, sich an den gängigen Artikulationsschemata des jeweiligen Sachfachs zu orientieren (Bohrmann-Linde, 2012). Auch Coyle et al. (2012) verweisen auf die Erkenntnisse der Didaktik des jeweiligen Sachfachs bei der Unterrichtsplanung:

[...] CLIL teachers should stay true to their instincts as content teachers in terms of having a guiding methodology which is appropriate for that subject. [...] science teachers use the experience gained through conducting experiments to raise questions and indicate methods of finding solutions. CLIL teachers must allow the subject to emerge in the same way as it usually would despite the role of the other language (Coyle et al., 2012, S. 92).

19 Coyle et al. (2012, S. 70ff.) empfehlen dazu eine Technik, bei der Videoaufnahmen einzelner Unterrichtssequenzen gemeinsam von Lehrern (und ggf. auch Schülern) im Hinblick auf solche Situationen untersucht werden, in denen effektives Lernen stattgefunden hat (*LOCIT-Technik: Lesson Observation and Critical Incident Technique*).

20 Hoffmann (2022, S. 7ff.) etwa schlägt für die Grundstruktur einer bilingualen Unterrichtsstunde ein Modell bestehend aus vier Phasen vor: Sprachliche Vorentlastung (Semantisierung), Erarbeitung, Festigung / Sicherung und Transfer. In den vom MSB-NRW (2012, S. 10) herausgegebenen Handreichungen für den bilingualen Unterricht im Fach Biologie wird ein kompetenzorientiertes Planungsraster zu Reihenplanung vorgeschlagen. Dabei werden die Kompetenzbereiche des Sachfachs durch die Bereiche *Bilinguale Sprachkompetenz* und *Interkulturelle Kompetenz* ergänzt. Zusätzlich werden Möglichkeiten zum fächerübergreifenden Arbeiten und zur kriteriengeleiteten Kompetenzdiagnose sowie Leistungsfeststellung als weitere Planungsfelder ausgewiesen.

4 Stand der Forschung: Der bilinguale Unterricht in den Naturwissenschaften

In diesem Kapitel wird der aktuelle Forschungsstand zur Wirksamkeit bilingualen Unterrichts in den Naturwissenschaften und insbesondere im Fach Chemie dargestellt. Die vorliegende Untersuchung erfährt damit eine Einordnung und Präzisierung hinsichtlich ihres Beitrags zur Erforschung bilingualen Unterrichts. Der dreifachen Zielsetzung der Studie folgend, werden wichtige Arbeiten zum sachfachlich-kognitiven Lernerfolg (Abschnitt 4.1), zum (fremd-) sprachlichen Lernerfolg (Abschnitt 4.2) sowie zu Auswirkungen auf die Motivation der Schüler zitiert (Abschnitt 4.3).

4.1 Der sachfachlich-kognitive Lernerfolg

Weil Forschungsergebnisse zum bilingualen Unterricht nur dann verglichen werden können, wenn klar ist, unter welchen Bedingungen und in welcher Form bilingualer Unterricht erteilt wurde, werden im Folgenden wichtige empirische Ergebnisse differenziert nach den ihnen jeweils zugrundeliegenden Organisationsformen des bilingualen Unterrichts vorgestellt. Dabei werden zunächst Ergebnisse aus Lang- (Abschnitt 4.1.1) und Kurzformen (Abschnitt 4.1.2) betrachtet, wobei zwischen qualitativen Studien mit geringer Probandenzahl (Abschnitt 4.1.2.1) und quantitativen Vergleichsstudien (Abschnitt 4.1.2.2) mit größerer Probandenzahl unterschieden wird. Den Erkenntnissen zu den kognitiven Vorteilen durch den bilingualen Unterricht ist ein separater Abschnitt gewidmet (Abschnitt 4.1.3), bevor im letzten Abschnitt Ergebnisse zu den Faktoren vorgestellt werden, die sachfachlichen Lernerfolg im bilingualen Unterricht bedingen (Abschnitt 4.1.4).

4.1.1 Erkenntnisse aus Immersionsprogrammen und bilingualen Zweigen

Vorbehalte, bilingualer Unterricht sei zu schwierig, weil die Fremdsprache nicht in gleichem Maße beherrscht werde wie die Muttersprache, erscheinen nicht nur plausibel, sondern werden auch durch Ergebnisse aus der Immersionsforschung gestützt. So beobachtete Cummings (1976) im Rahmen seiner Studien zum kanadischen Immersionsprogramm, dass Schüler, deren Fremdsprachenkompetenz einen gewissen Schwellenwert unterschreitet, große Schwierigkeiten haben, dem Unterricht in der Fremdsprache zu folgen (*threshold hypothesis*). Weder sprachlich noch sachfachlich konnten diese Schüler vom Immersionsunterricht profitieren. Auf der anderen Seite konnte jedoch gezeigt werden, dass bei Schülern, die über ausreichend Fremdsprachenkenntnisse verfügen, nach drei bis vier Jahren Immersionsunterricht keine

nennenswerten Wissensdefizite gegenüber den monolingual unterrichteten Kindern beobachtet werden. In Kanada ist deshalb, so Vollmer (2000), niemand mehr besorgt, ob durch den Immersionsansatz genügend Fachwissen vermittelt werden kann (ebd., 2000, S. 54). Problematisch stellt sich laut Marsh et al. (2000) auch die Situation des Immersionsunterrichts in Hong Kong dar. Ihre umfassende Untersuchung (N=12784) unter High School Schülern offenbarte erhebliche fachliche Defizite bei den Immersionsschülern in den Naturwissenschaften. Lo und Lo (2014) konnten diese Ergebnisse in ihrer Meta-Analyse von 24 weiteren Studien zum Hong Konger Immersionsansatz bestätigen (vgl. Piesche, 2016, S. 24). Vögedings (1995) Beobachtungen am Istanbul Lisesi, einer türkischen Regierungsschule mit deutscher Abteilung, verstärken die Zweifel am Lernerfolg sprachlich überforderter Kinder. Seine Untersuchungen zum deutschsprachigen Chemieunterricht türkischer Schüler offenbarten die Last und Qual, die solche Kinder erleben (zitiert nach Butzkamm, 2000, S. 100f.). So fordert auch Leisen (1991) aufgrund seiner Erfahrungen aus dem Physikunterricht an deutschen Auslandsschulen eine stärkere Verbindung von Sprach- und Sacharbeit. Auf der anderen Seite wird zu Recht darauf hingewiesen, dass die Ergebnisse des Immersionsunterrichts nicht ohne weiteres auf den bilingualen Unterricht übertragen werden können, insbesondere weil diese Schüler sowohl im Vorfeld als auch parallel zum bilingualen Unterricht zusätzlich Unterricht in der Fremdsprache erhalten. In seiner Studie zum bilingualen Chemieunterricht in Polen konnte Gregorczyk (2012) beispielsweise keine sachfachlichen Defizite bilingualer Schüler feststellen. Gregorczyk (2012) untersuchte insgesamt 265 Schüler aus zehn polnischen Mittelschulen, wobei 233 Schüler über fast zwei Jahre Chemieunterricht auf Polnisch und 32 Schüler Chemieunterricht auf Englisch erhielten. Die Schüler wurden von drei Lehrern unterrichtet und absolvierten einen Abschlusstest auf Polnisch, bestehend aus 22 Items.

4.1.2 Erkenntnisse aus Modulen und anderen diskontinuierlichen Formen

4.1.2.1 Qualitative Studien

Bonnet (2004) untersuchte das Wissen von zwei zehnten Klassen eines bilingualen Zweigs an einem Hamburger Gymnasium im Anschluss an eine bilinguale Unterrichtsreihe zum Thema *Redoxchemie*. Er videografierte die auf die Schülerexperimente folgenden Gruppendiskussionen und wertete diese qualitativ aus. Dabei ermittelt er für unterschiedliche chemische Konzepte die jeweilige Kompetenzstufe, auf der sich die Schüler präsentierten. Seine Ergebnisse verglich er mit den Ergebnissen von Spörlein (2003) und Todtenhaupt (1995) zum gleichen Thema bei Schülern aus Regelklassen. Bonnet (2004, S. 273) kommt zu dem Ergebnis, dass die konzeptuelle Kompetenz der bilingual unterrichteten Schüler bezogen auf die chemischen Kernkonzepte *Reaktion* und *Stoff/Teilchen* denen der monolingual-deutsch unterrichteten Schüler mindestens

gleichwertig ist. Koch (2005) verglich die im Rahmen einer naturwissenschaftlichen Arbeitsgemeinschaft in Klasse 5 an einem Gymnasium in Schleswig-Holstein erworbenen Konzepte zu den Themen *Luft* und *Wasser* miteinander. Dazu ließ sie die Schüler der bilingualen sowie der deutschsprachigen Gruppe (N = 39, N = 24) im Anschluss an die einjährige Arbeitsgemeinschaft *concept maps* erstellen und wertete diese qualitativ und quantitativ aus. In ihrer Analyse konnte sie keine Vorteile im konzeptuellen Wissen der monolingual-deutsch unterrichteten Schüler gegenüber den bilingual-englisch unterrichteten Schülern nachweisen (Koch, 2005, S. 96). Die bilingual unterrichteten Schüler, die bereits auf Vorerfahrung im bilingualen Unterricht aus der Grundschule verfügten, schnitten sogar etwas besser ab als die auf Deutsch unterrichtete Kontrollgruppe (Koch, 2005, S. 103). Kondring und Ewig (2006) untersuchen in ihrer Vergleichsstudie mit zwei neunten Klassen (N = 26, N = 20) eines nordrhein-westfälischen Gymnasiums die sachfachlichen Leistungen im Fach Biologie. Dazu führten sie im Anschluss an das sechsstündige Modul zum Thema *Gehör* einen 20-minütigen Wissenstest durch, der aus geschlossenen und wenigen offenen Items bestand. Diesen bewerteten sie anhand von Schulnoten und erkannten, dass keine signifikanten Unterschiede zwischen den fachlichen Leistungen der bilingual und der auf Deutsch unterrichteten Schüler zu beobachten waren. Hegerfeldt (2006) analysierte das Wissen von Schülern einer neunten Klasse eines Hamburger Gymnasium mit bilinguaalem Zweig im Anschluss an eine Unterrichtsreihe im Anfangsunterricht Chemie zu den Themen *Aggregatzustände* und *Gemische*. Anhand von Fallbeispielen analysiert sie Sequenzen des fragend-entwickelnden Unterrichtsgesprächs sowie schriftlich bearbeitete Aufgaben und resümiert, dass der bilinguale Chemieunterricht bei entsprechender sprachlicher Vorbereitung nicht zu sachfachlichen Defiziten führt (Hegerfeldt, 2006, S. 40). Rittersbacher (2006) untersuchte zwei neunte Klassen einer baden-württembergischen Realschule im Anschluss an ein Modul zur Erforschung der Kerze. In den drei Doppelstunden arbeiteten die Schüler weitgehend selbstständig in Gruppen, überlegten gemeinsam Fragestellungen, planten Experimente, werteten diese aus und präsentierten ihre Ergebnisse im Plenum. Dabei ging es auch darum, zu untersuchen, inwieweit bilingualer Unterricht gelingen kann, wenn die Schüler selbstständig als Forschende arbeiteten, mit minimaler Unterstützung durch den Lehrer. Sie verglich die Lernzuwächse der bilingualen und der deutschsprachigen Klasse, in dem sie die Arbeit in den Gruppen, die erzielten mündlich oder schriftlich dargestellten Ergebnisse und die Präsentationen verglich. Rittersbacher (2006) kommt zu dem Schluss, dass die Schüler aus beiden Klassen aus sachfachlicher Sicht vergleichbare Ergebnisse erzielten und folgert, dass bilingualer Unterricht auch bei "echtem naturwissenschaftlichen Arbeiten, wenn minimale Steuerung durch die Lehrkraft stattfindet", funktioniert (ebd., 2006, S. 34). Zu ähnlich positiven Ergebnissen kommt Luprichova (2013). Sie unterrichtete zwei Grundschulklassen (N = 21, N = 21) im Fach

Naturwissenschaften auf Englisch (bilingual) bzw. auf Tschechisch. In den anschließenden sachfachlichen Posttests stellte sie signifikant bessere Ergebnisse bei den bilingual unterrichteten Kindern fest (vgl. Pokrivcakova, 2013, S. 32). Die Ergebnisse sind jedoch mit Vorsicht zu betrachten, da Selektionseffekte nicht berücksichtigt wurden.

4.1.2.2 Quantitative Vergleichsstudien

Bei Untersuchungen zum sachfachlichen Lernerfolg handelt es sich bis in die 2010er Jahre hinein überwiegend um Falluntersuchungen und qualitative Studien. Bohn und Doff (2010, S. 75) beklagen deshalb den Mangel „empirisch abgesicherter Forschungsergebnisse auf breiter Basis“. Dies ändert sich in den folgenden Jahren mit dem Erscheinen größer angelegter empirischer Vergleichsstudien. In diesen wird immer häufiger auch untersucht, inwieweit Wissensunterschiede tatsächlich auf die Verwendung der Fremdsprache und nicht auf andere Effekte zurückzuführen sind. Damit ändert sich auch der Fokus der Untersuchungen, weil stärker als bislang die Bedingungen, unter denen gelernt wird, in den Blick genommen werden. Alter, Schulform und Vorerfahrung der Probanden rückten jetzt noch mehr in den Mittelpunkt des Forschungsinteresses. Hier sind zunächst die Studien von Hartmannsgruber (2014) und Piesche (2016) zu nennen. Sie betreffen junge Schüler der Sekundarstufe I mit vergleichsweise geringen Kenntnissen in der Fremdsprache (Klassen 6 und 7 an Realschulen). Hartmannsgruber (2014) untersuchte insgesamt 634 Schüler in 30 verschiedenen siebten Klassen an zehn unterschiedlichen Realschulen in Baden-Württemberg. Dabei wurde an jeder Schule über vier Wochen ein Modul im Umfang von zehn Stunden zum Thema *Wasser* jeweils auf Deutsch und - von derselben Lehrerin in einer anderen Klasse - bilingual unterrichtet. Die Erfassung der Wissenszuwächse erfolgte anhand von Tests im Pre-, Post-, Follow-Up-Design. Zusätzlich wurden Tests in einer weiteren Klasse pro Schule durchgeführt, die keinen Unterricht zum Thema erhalten hatte. Diese Gruppe diente als Kontrollgruppe. Hartmannsgruber konnte zeigen, dass auch Realschüler ohne Vorerfahrungen im bilingualen Unterricht und bei vergleichsweise geringer Fremdsprachenkompetenz signifikante Wissenszuwächse verzeichnen konnten. Diese waren jedoch bedeutend geringer als in den auf Deutsch unterrichteten Gruppen, wobei der Unterschied zum Follow-Up-Test schrumpfte und dann nicht mehr signifikant war. Hartmannsgruber (2014, S. 94) schließt deshalb, dass trotz der deutlichen Wissensunterschiede der beiden Testgruppen die signifikanten Lernzuwächse den Erfolg des bilingualen Unterrichts im sachfachlichen Bereich bestätigen. Piesche (2016) untersuchte ebenfalls eine große Anzahl (N = 722) an Realschülern aus Baden-Württemberg, die über keinerlei Vorerfahrung im bilingualen Unterricht verfügten. Dabei erhielten 30 Klassen der Jahrgangsstufe 6 im Fach Physik über zehn Stunden Unterricht in einem Modul zum Thema *Schwimmen und Sinken*.

Während die Testgruppe bilingual unterrichtet wurde, fand der Unterricht in der Kontrollgruppe auf Deutsch statt. Das Wissen wurde wieder durch Pre-, Post-, Follow-Up-Tests erfasst, die 35 geschlossene und eine offene Aufgabe beinhalteten und auf bekannte Items aus der TIMS-Studie zurückgriffen. Wie schon bei Hartmannsgruber (2014) in seiner Untersuchung zum bilingualen Biologieunterricht, verzeichnet Piesche (2016) auch für die bilingual unterrichteten Schüler signifikante Lernzuwächse. Diese schneiden jedoch im Vergleich zu den Schülern der Regelklasse erneut signifikant schlechter ab, wobei jedoch der Effekt mit etwa einem Viertel der Standardabweichung gering ist. Anders als Hartmannsgruber, bei dem die Differenz in den sachfachlichen Lernzuwächsen vom Post- zum Follow-Up-Test sank, kann Piesche derartige Veränderung nicht beobachten. Piesche (2016) findet damit keinen Hinweis auf eine bessere Behaltensleistung der bilingual unterrichteten Schüler (Piesche, 2016, S. 89). Die Forschungsergebnisse zu älteren Schülern mit höheren Fremdsprachenkompetenzen (in der Regel B1 oder B2) fallen im Hinblick auf den sachfachlichen Lernerfolg der bilingual unterrichteten Schüler noch positiver aus. Weder Haagen-Schützenhöfer (2011) noch Rodenhauser (2016) können signifikante Unterschiede in den Wissenstest im Anschluss an die Module feststellen, was bedeutet, dass das sachfachliche Lernen in den bilingualen Gruppen genauso gut gelingt wie in den Gruppen des Regelunterrichts. Die ausgewählten Untersuchungen decken dabei unterschiedlich lange Zeiträume ab, in denen bilingualer Unterricht erteilt wurde. Diese reichen von einem Tag (Labortag in einem biologischen Labor bei Rodenhauser) bis zu einigen Wochen (Physikmodul bei Haagen-Schützenhöfer). Haagen-Schützenhöfer (2011) untersuchte das Wissen von insgesamt 205 Schülern aus elf elften Klassen an sechs verschiedenen Gymnasien in Österreich im Anschluss an ein Modul zum Thema „Magnetismus“. Dabei führte sie Pre- und Post-Tests durch, bei denen geschlossene, halb-offene und offene Items zum Einsatz kamen. Die Schüler verfügten über ein gutes Fremdsprachenniveau (B1) hatten aber keinerlei Vorerfahrung im bilingualen Unterricht. Die Auswertung der Testergebnisse zeigte Leistungszuwächse in beiden Lerngruppen, wobei jedoch kein signifikanter Unterschied festgestellt werden konnte. Haagen-Schützenhöfer (2011) hält deshalb fest, dass ihre Ergebnisse zumindest andeuten, dass bilingualer Unterricht nicht unweigerlich schlechtere fachliche Leistungen verursacht. Rodenhauser (2016) testete Schüler der Oberstufe verschiedener weiterführender Schulformen in Nordrhein-Westfalen (N=490) im Anschluss an einen Labortag zum Thema „Genetic Fingerprint“. Die quasi-experimentelle Vergleichsstudie wurde in mehreren Grund- und Leistungskursen im Fach Biologie durchgeführt, wobei die Laborsprache jeweils Englisch bzw. Deutsch war. Außerdem testete Rodenhauser eine Kontrollgruppe, die nicht an einem entsprechenden Labortag teilgenommen hatte. Die Wissenstests wurden im Pre-, Post-, Follow-Up-Test-Design durchgeführt, wobei überwiegend *Multiple-Choice-*

Items zum Einsatz kamen. Anders als bei Haagen-Schützenhöfer verfügte die Mehrheit der untersuchten Schüler bereits auf Vorerfahrung im bilingualen Unterricht, da sie Schulen mit bilinguaalem Zweig angehörten. Obwohl die auf Deutsch unterrichteten Schüler etwas besser im Post-Test abschnitten als die bilingual unterrichteten Kinder, waren die Unterschiede nicht signifikant und nahmen zum Follow-Up-Test hin weiter ab.

4.1.3 Kognitive Vorteile durch bilingualen Unterricht

Während in den meisten der zitierten Untersuchungen sachfachlicher Lernerfolg über das erworbene Wissen erfasst wird, sind Untersuchungen eher selten, bei denen auch höhere Denkopoperationen wie das Anwenden oder der Transfer des Wissens erfasst werden. Eine Ausnahme ist die Studie von Osterhage (2007), der eine Überlegenheit der bilingualen Schüler im Kompetenzbereich *konvergentes Denken* feststellte. Er untersuchte dabei die sachfachlichen Kompetenzen von 129 bilingual und 107 auf Deutsch unterrichteter Schüler aus neun Klassen an drei unterschiedlichen Bremer Gymnasien im Fach Biologie. Anhand von 18 Items aus der PISA-Untersuchung von 2004 analysierte er sechs der insgesamt sieben im Rahmen der PISA-Studie betrachteten Teilkompetenzen der *scientific literacy*. Für die drei untersuchten Inhaltsfelder (*Fortpflanzung und Sexualität, Atmung und Fotosynthese* sowie *Biochemie und Ernährung*) konnte er signifikant bessere Leistungen der Schüler des bilingualen Zweiges in den drei Teilkompetenzen *konvergentes Denken, Umgang mit Graphen* und *Umgang mit Zahlen* und vergleichbare Leistungen in den anderen drei Kompetenzbereichen nachweisen. Damit kann er zwar zeigen, dass bilinguale unterrichtete Schüler im Fach Biologie vergleichbar gute und teilweise sogar bessere Leistungen erbringen können als die Schüler im Regelunterricht, er stellt jedoch klar:

Im Lichte dieser Ergebnisse nun zu behaupten, bilingualer Unterricht sei der bessere Unterrichtstyp in Bezug auf das Sachfachlernen im Fach Biologie, wäre fatal und unüberlegt, da eine solche Aussage zahlreiche Faktoren, wie die Zusammensetzung der Schülerschaft und mögliche Vorselektionen für den bilingualen Unterricht, außer Acht lässt (Osterhage, 2007, S. 48).

Zu ähnlichen Ergebnissen kommt Zydati (2012, S. 24), der die sachfachbezogenen Diskurskompetenzen von Regelschlern und Schlern bilingualer Zweige an drei Berliner Gymnasien am Beispiel einer Unterrichtseinheit zum Thema *Bevlkerungsexplosion im 19. Jahrhundert* vergleicht (N=133). Dabei stellt er fest:

[...] CLIL learners excel at the cognitively more complex tasks involving interpretation, drawing conclusions and/or text-bound writing. Since the testees of the two groups could not be matched with regard to their cognitive dispositions, this result could well be a consequence of the selective nature of CLIL learners' admittance to these classes (which demands further investigation with matched or randomised samples. (Zydati, 2012, S. 24).

Bonnet (2004), der für das Fach Chemie ein eigenes Kompetenzmodell entwickelt hat, ordnet die Schüler seiner Studie anhand von Unterrichtsbeobachtung den einzelnen Stufen des Modells zu. Auch in seiner Untersuchung deutet sich eine Überlegenheit der bilingualen Schüler an. Interessante Hinweise auf kognitive Effekte bilingualen Unterrichts liefern auch Studien an Alzheimerpatienten. Diese deuten an, dass der Gebrauch einer Fremdsprache im Alter die Leistungsfähigkeit des Gehirns erhalten kann. Mittlerweile konnte gezeigt werden, dass diese kognitive Fähigkeitsreserve (*cognitive reserve*) nicht nur im Falle eines natürlichen Bilingualismus zu beobachten ist, sondern auch bei verspätet einsetzendem Fremdsprachenlernen (vgl. Diehr, 2016, S. 60). So erhofft man sich auch für den bilingualen Unterricht Trainingseffekte aufgrund der permanenten Doppelbeanspruchung des Gehirns, das neben inhaltlichen auch komplexe sprachliche Informationen verarbeiten muss und zwischen Fremd- und Schulsprache vermittelt. Piesche (2016, S. 22f.) stellt mit Blick auf die Arbeiten von Bialystok (1999) und Kollegen (Bialystok & Martin, 2004) fest, dass „bilinguale Lernende besser darin sind, irrelevante Informationen zu unterdrücken, was das Arbeitsgedächtnis vor Überlastung bewahrt und sich so positiv auf kognitive Prozesse auswirkt“. Vollmer (2006) beobachtet außerdem ein höheres Maß an prozeduralem Wissen bei bilingualen Schülern und vermutet, dass sprachliche Probleme zu intensiveren mentalen Konstruktionsprozessen führen, welche eine tiefere semantische Verarbeitung der Fachinhalte bzw. einen höheren Grad an Elaboriertheit der Inhalte zur Folge haben (vgl. Wolff, 1997; Badertscher & Bieri, 2009; Heine, 2010a; Lamsfuß-Schenk, 2008). Das Modell der Verarbeitungstiefe von Craik und Lockhart (1972) besagt, dass Informationen auf unterschiedlichen Niveaus verarbeitet werden. Zu diesen gehören die perzeptuelle und die semantische Verarbeitung. Nach Craik und Lockhart führt die semantische Verarbeitung zu besseren Behaltensleistungen als die oberflächliche Wahrnehmung von z.B. phonemischen Merkmalen. Dies liegt daran, dass bei der semantischen Verarbeitung neue Informationen mit bereits vorhandenem Hintergrundwissen verknüpft werden (Heine, 2010b, S. 209f.). Vokabellücken in der Fremdsprache veranlassen Schüler dazu- so die Theorie, alternative Formulierungen auszuprobieren und dabei verwandte Konzepte miteinander in Beziehung zu setzen. Auch Rodenhauser (2016, S. 150) deutet ihre guten Ergebnisse zum sachfachlichen Lernerfolg gedächtnispsychologisch und stellt fest, dass der *Extraneous Load*, der durch die Verwendung der Fremdsprache hervorgerufen wird, nicht so hoch ausfällt, wie

erwartet und folgert, dass immer noch genügend Kapazitäten des Arbeitsgedächtnisses zur Bearbeitung der biologischen Inhalte zur Verfügung stehen. Offensichtlich tragen eine größere Elaboriertheit und eine tiefere semantische Verarbeitung dazu bei, die zusätzliche Belastung des Arbeitsgedächtnisses zu kompensieren. Eine umfangreiche Studie zur kognitiven Wirksamkeit bilingualen Unterrichts im Primarbereich liefert Böttger (2020). In einer an 21 Grundschulen durchgeführten Längsstudie über vier Jahre, an der sich über 900 Schüler und 42 Lehrer beteiligten, stellte Böttger fest, dass auch die Leistungen in den Fächern Mathematik und Deutsch "gleich gut bis signifikant besser" waren als bei den Regelschülern. Böttger begründet das mit dem Einfluss, den die Verwendung der Fremdsprache auf die Entwicklung des Gehirns hat und resümiert:

Damit bestätigt sich die Hypothese, dass zweisprachig Unterrichtete bzw. Aufwachsende durch die Herausbildung von neuen Netzwerken im Gehirn kognitive Vorteile entwickeln, die weit über das reine Englischlernen hinausreichen (Böttger H. , 2019).

Auch van der Craen et al. (2007b, S. 74ff.) stellen fest, dass bilingual unterrichtete Schüler weniger Aufwand zur Lösung mathematischer Probleme investieren müssen und führen dies auf eine größere Plastizität des Gehirns zurück (van de Craen, Mondt, Allain, & Gao, 2007b).

Wenngleich die Schülerinnen und Schüler in Böttgers (2020) Studie deutlich jünger waren als in der vorliegenden Arbeit, sind die Ergebnisse doch insofern interessant, als bilingualer Unterricht in einer dem Modulunterricht durchaus vergleichbaren Organisationsform angeboten wurde. Er erfolgte nämlich im Rahmen des normalen Unterrichts phasenweise oder über mehrere Unterrichtsstunden hinweg in den Fächern Kunst, Musik, Sport sowie weiteren Fächern. Der Unterricht orientierte sich dabei an den gültigen Lehrplänen und den dort formulierten Kompetenzerwartungen. Die vorherige Aufnahme in eine bilinguale Klasse erfolgte ohne gezielte Zuteilung der Kinder auf Antrag der Eltern.

4.1.4 Faktoren, die den sachfachlichen Lernerfolg beeinflussen

Motivation

Der Einfluss der Orientierung auf den sachfachlichen Lernerfolg ist im bilingualen Unterricht im Bereich der Naturwissenschaften bisher wenig erforscht. Piesche (2016) untersucht in ihrer empirischen Studie zum bilingualen Physikunterricht an Realschulen den Einfluss von Selbstkonzept und Interesse in Bezug auf Englisch und die Naturwissenschaften bei Schülern der Jahrgangsstufe 6. Ihre Hypothese, dass „die Lerner mit hoher Motivation in NWA [Naturwissenschaftliches Arbeiten: Fächerverbund aus den Fächern Physik, Chemie und Biologie, F.H.] im englischsprachigen NWA-Unterricht schlechtere Leistungen erzielen als im

deutschsprachigen Pendant“ konnte anhand der empirischen Daten in dieser Allgemeinheit nicht bestätigt werden (Piesche, 2016, S. 109). Jedoch kann sie für die Schüler des bilingualen Unterrichts einen statistisch signifikanten Zusammenhang zwischen den Ergebnissen im Wissenstest und dem Selbstkonzept im Fach Englisch nachweisen. Schüler mit einem geringen bzw. durchschnittlichen Selbstkonzept in Englisch erzielten in ihrer Untersuchung im bilingualen Unterricht statistisch signifikant schlechtere Ergebnisse als vergleichbare Schüler im Regelunterricht (Piesche, 2016, S. 107). Ein ähnlicher Zusammenhang zwischen dem Selbstkonzept in den Naturwissenschaften und dem sachfachlichen Lernerfolg ergab sich aus ihren Daten nicht. Während auch das Interesse an Englisch keinen signifikanten Einfluss in ihrer Untersuchung hatte, war dies beim Interesse an den Naturwissenschaften sehr wohl der Fall. Das Interesse in NWA wirkte sich nämlich im monolingualen Unterricht „bei hohem thematischem Vorwissen bzw. bei hohem Vorwissen in Physik signifikant auf die Testleistung aus“, während der Einfluss im bilingualen Unterricht „tendenziell negativ war“ (Piesche, 2016, S. 111).²¹ Ähnliches gilt bei niedrigem bzw. durchschnittlichem Interesse an Englisch. Auch hier wirkt sich das Interesse in NWA signifikant positiv auf den sachfachlichen Lernerfolg im Regelunterricht aus, während bilinguale Schüler dann Vorteile haben, wenn ihr Interesse an beiden Fächern gleichermaßen hoch ist (Piesche, 2016, S. 113f.). Dies bedeutet, dass Schüler mit einem geringen bzw. mäßigen Selbstkonzept in Englisch sowie Schüler, die in Physik leistungsstark sind bzw. ein geringes bzw. durchschnittliches Interesse an Englisch haben, dann im Regelunterricht bessere Leistungen erzielen als im bilingualen Unterricht, wenn sie gleichzeitig ein großes Interesse an den Naturwissenschaften haben.

Fremdsprachenkompetenz

Ausgehend von den Untersuchungen Cummings zu Immersionsprogrammen in Kanada (Cummings 1978), wird auch für den bilingualen Sachfachunterricht in Deutschland diskutiert, inwieweit geringe fremdsprachliche Kenntnisse die Leistungsfähigkeit in dieser Unterrichtsform beeinträchtigen. Cummings hatte zwei Schwellen der Sprachkompetenz bilingualer Lerner identifiziert. Unterschreitet das sprachliche Niveau in Erst- und Zweitsprache ein bestimmtes Niveau, so leidet der Lernerfolg und Schüler können keine vergleichbaren kognitiven Leistungen erbringen wie Schüler der Regelklassen. Oberhalb eines zweiten, deutlich höheren Niveaus

21 Dies deckt sich mit den Ergebnissen einer vom Institut für die Pädagogik der Naturwissenschaften (IPN) in Kiel durchgeführten Untersuchung. Hier korrelierte das Fachinteresse an Physik mit der entsprechenden Schulnote mit $r = 0.38$. (Hoffmann, Häussler, & Lehrke, 1998).

hingegen übertreffen die Leistungen der bilingualen Schüler sogar die der Regelschüler (*double threshold hypothesis*). Zydatiś (2012), der sachfachliche und fremdsprachliche Leistungen von Regelschülern und Schülern bilingualer Zweige miteinander verglich (DEZIBEL-Studie), konnte die beiden von Cummings postulierten Grenzen anhand sprachlicher Merkmale für die Fremdsprache Englisch bestimmen. Es zeigte sich bei den auf Englisch getesteten bilingualen Schülern, dass fachbezogene Diskurskompetenzen in der Fremdsprache nur dann ausgebildet werden können, wenn die Schüler über ein hohes Maß an lexiko-grammatischen Kompetenzen und eine hohe allgemeine Sprachfähigkeit verfügen. Abendroth-Timmer (2007, S. 274) hingegen konstatiert in ihrer Untersuchung zum bilingualen Modulunterricht in unterschiedlichen Sachfächern, dass "bilingualer und mehrsprachiger Unterricht auf allen Stufen der fremdsprachlichen Kompetenz möglich ist" und plädiert deshalb für die stärkere Einbeziehung weiterer Fremdsprachen wie Französisch und Spanisch. Farrel (2011), der die Ergebnisse maltesischer Schüler in den Abschlussprüfungen untersuchte, stellt jedoch fest, dass bilingual unterrichtete Schüler die besten Leistungen in den Fächern Mathematik und Physik erzielten, wenn sie sowohl in Maltesisch als auch in Englisch besonders gut waren. Marsh et al. (2000) fanden ebenfalls Hinweise auf den positiven Einfluss fremdsprachlicher Fähigkeiten in Bezug auf den Wissenszuwachs bei chinesischen Immersionsschülern. Die Schüler zeigten zum Teil erhebliche Defizite in den Leistungen in den Naturwissenschaften, die jedoch bei den Schülern, die in Englisch besonders leistungsstark waren, geringer ausfielen. Piesche (2016) ermittelt keinen statistisch signifikanten Zusammenhang zwischen der Englischkompetenz und den Ergebnissen im Wissenstest. Es waren nicht die Schüler mit den besseren Leistungen in der Fremdsprache, die gegenüber den monolingual unterrichteten Schülern vergleichbare Leistungen erzielten, sondern die, die ihre Leistungsfähigkeit in Englisch höher einschätzten und über ein besseres Selbstkonzept verfügten (Piesche, 2016, S. 153).²² Hegerfeldt (2006) stellt im bilingualen Chemieunterricht für die Probleme bei der Darstellung von *Lösungsvorgängen* und *Aggregatzuständen* auf Teilchenebene fest, dass ähnliche Schwierigkeiten auch in einem rein muttersprachlichen Unterricht aufgetreten wären²³. An anderer Stelle räumt sie jedoch in Bezug auf die Sprachkompetenz ihrer Schüler ein:

22 Anders als Zydatiś (2012) führte Piesche jedoch die Wissenstest in der Muttersprache Deutsch durch und bevorzugte geschlossene bzw. halb-offene Items. Fremdsprachenkenntnisse sind für die Bearbeitung von Piesches Tests nicht nötig. Die Tests der DEZIBEL-Studie enthalten hingegen überwiegend offene Aufgaben, die längere Sequenzen der Sprachproduktion in der Fremdsprache erforderten.

23 Bonnets (2004) Analysen zum bilingualen Chemieunterricht in verschiedenen zehnten Klassen eines Hamburger Gymnasiums zeigen ebenfalls, dass die auftretenden Verständnisprobleme nicht auf den Gebrauch der Fremdsprache, sondern vielmehr auf die Abstraktheit der unterrichteten Konzepte zurückzuführen sind. So zeigt er

Hätte meine Lerngruppe hingegen geringe Sicherheit in der Fremdsprache gehabt, wäre sie bilingualen Unterricht nicht gewohnt gewesen oder hätte diesen sogar als „zu schwer“ abgelehnt, so hätte die Fremdsprache eine enorme Hürde dargestellt, welche sich zweifelsohne auf das Unterrichtsniveau und den Stoffumfang ausgewirkt hätte. Skeptikern des bilingualen Unterrichts ist also insofern Recht zu geben, als dass der Boden sorgfältig bereitet werden muss, um qualitative Einbußen im Sachfach zu verhindern (Hegerfeldt A. , 2006, S. 37).

4.2 Der (fremd-)sprachliche Lernerfolg im Bereich der Lexik

Anders als im Bereich des sachfachlichen Lernerfolges liegen im Bereich der fremdsprachlichen Lernzuwächse umfangreiche empirische Untersuchungen vor, welche die fremdsprachlichen Kompetenzvorteile bilingual unterrichteter Schüler gegenüber Regelschülern belegen. Eine Übersicht über die neun größten empirischen Untersuchungen zum fremdsprachlichen Mehrwert des bilingualen Unterrichts an bilingualen Zweigen in Deutschland bietet Rumlich (2016, S. 204f.). Studien zeigen, dass die Vorteile der bilingual unterrichteten Schüler insbesondere die rezeptiven Fertigkeiten, wie Hör- und Leseverstehen, sowie den Wortschatz betreffen. Diese Beobachtungen decken sich mit den Ergebnissen der kanadischen Immersionsforschung, die zeigen, dass Immersionsschüler im rezeptiven Bereich Kompetenzen eines Muttersprachlers erreichen können (Day & Shapson, 1996). Auf der anderen Seite scheinen bestimmte Bereiche vom bilingualen Unterricht nur wenig zu profitieren. Dalton-Puffer (2008, S. 5) unterscheidet zwischen solchen Fremdsprachenkompetenzen, die bevorzugt durch den bilingualen Unterricht beeinflusst sind und solchen, die wenig beeinflusst sind.

Favourably affected	Unaffected or Indefinite
Receptive skills	Syntax
Vocabulary	Writing
Morphology	Informal/non-technical language
Creativity, Risk-taking, fluency, quantity	Pronunciation
Emotive/affective outcomes	Pragmatics

Tabelle 5: Auswirkungen bilingualen Unterrichts auf unterschiedliche Sprachkompetenzen

(entnommen aus Dalton-Puffer C., 2008, S. 5)

zu den beobachteten Schwierigkeiten im Bereich der *Redoxchemie*, dass ein Wechsel in die Muttersprache nicht zur inhaltlichen Klärung beiträgt.

Während die fremdsprachlichen Erfolge bilingualen Unterrichts in bilingualen Zügen unbestritten sind, ist völlig unklar, inwieweit diese auf das Format bilingualen Unterrichts selbst oder auf andere Faktoren zurückzuführen sind. Rumlich (2016) konnte empirisch nachweisen, dass Selektionseffekte, Effekte der Klassenzusammensetzung sowie Effekte, bedingt durch die gezielte, zusätzliche fremdsprachliche Vorbereitung maßgeblich für die am Ende der Klasse 8 gemessenen Vorteile bilingualer Schüler verantwortlich sind. Dabei untersuchte er in einer Longitudinalstudie 38 Gymnasialklassen (N=953) an Regelschulen und Schulen mit bilingualem Zweig in Nordrhein-Westfalen. Anhand von C-Tests bestimmte er die Entwicklung der allgemeinen Fremdsprachenkompetenz von der sechsten bis zur achten Klasse. Auch Dallinger (2015) konnte deutliche Selektionseffekte nachweisen. Die Vergrößerung des Wortschatzes, insbesondere durch Fachbegriffe und Ausdrücke aus dem akademischen Register, wird, wie gesehen, in vielen Untersuchungen zur fremdsprachlichen Wirksamkeit bilingualen Unterrichts neben den rezeptiven Fertigkeiten meist an erster Stelle genannt (vgl. außerdem Wode et al., 1996; Zydati, 2007; Kampmeier, 2013). Auch Dalton-Puffer (2008, S. 6) sieht mit Blick auf die strukturellen Elemente von Sprache die Lexik als den Bereich, der am stärksten vom bilingualen Unterricht profitiert: The greatest gain in terms of the language system, however, is undoubtedly produced in the lexicon: through studying content subjects in the foreign language CLIL learners process larger vocabularies of technical and semi-technical terms and possibly also of general academic language which gives them a clear advantage over their EFL-peers (Dalton-Puffer, 2008, S. 6). Zahlreiche empirische Untersuchungen haben in der Vergangenheit gezeigt, dass bilinguale Lerner einen größeren und differenzierteren Wortschatz haben als Regelschüler (z.B. Baker, 2002; Jexenflicker & Dalton-Puffer, 2010; Zydati, 2007). Auch Burmeister und Daniel (2002) stellen in ihrer Übersicht zu den Ergebnissen des Kieler Immersionsprojekts an Gymnasien abschließend fest, dass die größten Vorteile der CLIL-Schüler im Bereich des L2-Wortschatzes und der Diskurskompetenz bestehen. In ihrer Longitudinalstudie zur Entwicklung der Sprachfähigkeiten von Immersions- und Regelschülern in den Klassen 7 (N=543) und 10 (N=517) wurden die Leistungen von IM-Schülern, Nicht-IM-Schülern und von Schülern aus Regelklassen untersucht. Dabei erhielten die IM-Schüler bilingualen Unterricht nach dem Nordrhein-Westfalen-Modell für bilinguale Zweige. Dieses beinhaltet einen fremdsprachlichen Vorlauf in den Klassen 5 und 6 und BSFU ab Klasse 7, bei erhöhter Stundenzahl in den Sachfächern. Basierend auf einem Simulationsspiel (*A difficult decision*) wurde ein kommunikativer Test entwickelt, der unter anderem Aussagen zum produktiven Wortschatz der Schüler erlauben soll. Dabei bearbeiten die Schüler drei Aufgaben mündlich und eine schriftlich. Die mündlichen Daten wurden transkribiert und es wurden alphabetische Wortlisten erstellt, die Aufschluss über den Umfang, die Korrektheit, den Ursprung und die Differenziertheit des produktiven Wortschatzes der Schüler geben. Es zeigte

sich, dass der Wortschatz der IM-Schüler differenzierter und erheblich größer war als der der Regelschüler, da sie mehr lexikalische Varianten sowie Vokabeln benutzen, die weder im Test noch im Vokabular des Lehrwerks zum regulären Englischunterricht vorkamen (Burmeister & Daniel, 2002, S. 510ff.). Interessanterweise waren keine Einbußen bei der lexikalischen Korrektheit zu beobachten. Diese werden häufig für den bilingualen Unterricht befürchtet, weil ja explizite Bezüge auf die sprachliche Form im Unterricht meist fehlen. Zu ähnlich guten Ergebnissen in Bezug auf die lexikalischen Kompetenzen der bilingualen Schüler kommt Zydatiś (2007) in der DEZIBEL-Studie (Deutsch-Englische-Züge in Berlin), in der er die Sprachleistungen bzw. Sprachfähigkeiten von Schülern der 10. Klasse an drei verschiedenen Berliner Gymnasien untersuchte. Die Schüler des bilingualen Zweigs, die in Klasse 9 und 10 bilingualen Unterricht in den Fächern Erdkunde, Geschichte und/oder Biologie erhalten hatten (N=106), schnitten dabei auch in allen anderen Bereichen (Lese- und Hörverstehen, textsortengebundenes Schreiben, kontextualisierter Gebrauch lexikalischer und grammatischer Phänomene (*Use of English*) und mündliche Sprachkompetenz) signifikant besser ab als die Regelschüler (N=85). Beim Wortschatz jedoch gab es den größten Unterschied zwischen den Gruppen. Auch die Texte der bilingualen Schüler zeichneten sich durch eine deutlich größere Differenziertheit in der Wortwahl aus²⁴. Auch in den *Zentralen Abschlussprüfungen nach Klasse 10* in Nordrhein-Westfalen wurde die Überlegenheit der bilingualen Schüler insbesondere im Bereich Wortschatz deutlich (Kampmeier, 2013). Getestet wurden Leistungen in den Bereichen Hör-, Leseverstehen und Schreiben an 46 Gymnasien in Nordrhein-Westfalen (N=3763), wobei die Leistungen von Regelschülern und Schülern bilingualer Zweige (N=346) verglichen wurden. Da Selektionseffekte bei diesem Vergleich nicht berücksichtigt wurden, kann jedoch auch hier aus den Ergebnissen nicht abgeleitet werden, inwieweit die Unterrichtsform des bilingualen Unterrichts diesen Vorsprung bewirkt. Weitere Studien gingen in der Folge der Frage nach, ob die lexikalischen Vorsprünge der bilingualen Schüler den Fach- bzw. den Allgemeinwortschatz betreffen. Baker (2002) konnte in ihrer Untersuchung zur akademischen Sprachfähigkeit bilingualer Schüler im Fach Biologie zeigen, dass diese über einen insgesamt größeren Wortschatz verfügen als Regelschüler. Baker (2002) untersuchte zwei bilinguale und zwei Regelklassen eines Berliner Gymnasiums, die ab Klasse 10 in den Fächern Erdkunde, Politik und Biologie bilingual unterrichtet wurden. Außerdem absolvierten die bilingualen Schüler in der Mittelstufe eine Phase mit verstärktem Englischunterricht. Die Tests wurden in den Klassen 10 und

24 Rumlich (2016) macht deutlich, dass Zydatiś zwar eine Reihe von Kontrollvariablen untersucht, diese aber nicht in ein umfassendes statistisches Modell einbezieht, so dass der Anteil des lexikalischen Vorsprungs, der auf die Unterrichtsform des bilingualen Unterrichts zurückzuführen ist, nicht bestimmt werden kann.

11 durchgeführt. Den Wortschatz erfasste sie dabei mit Hilfe von C-Tests, bei denen ausschließlich Inhalts- und keine Funktionswörter getilgt worden waren. Dabei wurden sowohl biologische Fachtexte als auch belletristische Texte eingesetzt, um eine Differenzierung zwischen Fach- und Allgemeinwortschatz zu ermöglichen. Die bilingualen Schüler erzielten in allen Texten im Mittel bessere Ergebnisse als die Regelschüler, bei denen eine größere Variation in den Testergebnissen vorlag. Die Unterschiede waren besonders deutlich in Texten, die viele unbekannte Vokabeln enthielten. Laut Baker (2002, S. 142) haben die bilingualen Schüler nicht nur einen erheblich größeren Wortschatz, sondern verfügen auch über effektivere Strategien, Vokabellücken zu schließen. Dieser Vorsprung wurde im Bereich der Fachsprache besonders deutlich (Baker, 2002, S. 142). Inwieweit diese Vorteile auf den bilingualen Unterricht als Unterrichtsform zurückzuführen sind, kann jedoch auch hier abschließend nicht eindeutig gesagt werden, da eine Reihe von Variablen (z.B. allgemeine Englischkompetenz, akademisches Selbstkonzept und Interesse im Fach Englisch, Kontakt mit der L2 in der Freizeit, sozioökonomischer Status, kognitive und allgemeinsprachliche Fähigkeiten) nicht erfasst wurden. Die bilingualen Schüler waren schon deshalb im Vorteil, da sie in der Mittelstufe zusätzlichen Englischunterricht erhalten hatten. In ihrer Vergleichsstudie mit Regelschülern und bilingualen Schülern an weiterführenden Schulen in Finnland, konnten auch Merikivi und Pietilä (2014) deutlich bessere Leistungen der bilingualen Schüler im Wortschatz bestätigen. Dabei wurde die *Breite des Wortschatzes* der Schüler im sechsten (N=149) sowie im neunten Schuljahr (N=181) getestet. Zu beiden Messzeitpunkten wurden dabei signifikant bessere Leistungen der bilingualen Schüler im rezeptiven wie im produktiven Wortschatz festgestellt. Die Forscher untersuchten jedoch nicht nur Fachbegriffe, sondern Begriffe unterschiedlicher Häufigkeiten und fanden auch für besonders häufige Begriffe der Allgemeinsprache signifikante Unterschiede zugunsten der bilingualen Schüler. Mit abnehmender Häufigkeit wurden diese jedoch deutlich größer. Stets erwies sich dabei der rezeptive Wortschatz, der mit dem produktiven korrelierte, als deutlich größer als der produktive. Sylven (2004) kann die Überlegenheit der bilingualen Schüler für den Bereich der Fachsprache bestätigen, findet jedoch im Bereich der Allgemeinsprache keine oder nur geringfügige Vorteile der bilingualen Schüler an schwedischen Schulen (vgl. Dalton-Puffer, 2008, S. 6). Außerdem stellt sie fest, dass das außerschulische Leseverhalten ein besserer Prädiktor für das Abschneiden in Wortschatztests ist als die Teilnahme am bilingualen Unterricht (vgl. Dalton-Puffer C. , 2008). Auch Austad (2013) konstatiert in seiner Untersuchung in vier zehnten Klassen in Norwegen (N=71) keine Vorteile der Schüler in den beiden bilingualen Klassen. Im Gegenteil: Hier schneiden die Regelschüler in den Tests zur Wortschatzbreite sogar besser ab. Diese widersprüchlichen Ergebnisse aus Skandinavien werden verständlich, wenn man den Einsatz der Fremdsprache im bilingualen Unterricht der

norwegischen Schüler betrachtet. Laut Austad (2015) betrug der Anteil der Fremdsprache im bilingualen Unterricht nämlich maximal 30 % und auch die verwendeten Lehrwerke waren nicht in der Fremd- sondern in der Muttersprache verfasst. Zudem hatten die Schüler der zehnten Klasse, die untersucht wurden, erst seit Beginn des Schuljahres, d.h. nur wenige Monate bilingualen Unterricht. In der finnischen Studie von Merikivi und Pietilä (2014) hingegen werden bilinguale Lerner verglichen, die bereits in der Grundschule, d.h. ab der ersten Klasse bilingual unterrichtet wurden. Der Umfang, in dem bilingualer Unterricht erteilt wurde, betrug anfänglich 80 % und sinkt im Verlaufe der Schulzeit. Er beträgt aber auch am Ende der Sekundarstufe noch 40 %. Außerdem wird bilingualer Unterricht hier überwiegend ausschließlich in der Fremdsprache erteilt und auch fremdsprachliche Unterrichtsmaterialien werden verwendet. Das Ausmaß der Zuwächse im Bereich Lexik hängt aber nicht nur vom Umfang der Fremdsprache im bilingualen Unterricht ab, sondern auch von der Unterrichtsmethodik. Die Frage etwa, ob der Wortschatzerwerb durch explizit Maßnahmen unterstützt bzw. ob Wortschatz auch gezielt geübt wird, spielt hier eine Rolle. Dalton-Puffer (2008, S. 6) bemerkt dazu:

In general the study of vocabulary learning in instructed settings has shown that gains are particularly great if vocabulary is worked on explicitly: interestingly, vocabulary is usually the only linguistic aspect which is explicitly treated in CLIL lessons (Dalton-Puffer C. , 2008, S. 6).

Diese Beobachtung wird durch die Untersuchung Gablasovas (2015) unterstützt, die die Leistungen tschechischer Schüler eines bilingualen Zweigs (N=35) und mit denen von Regelschülern (N=36) im Hinblick auf ihr Verständnis von Fachbegriffen verglich. Die Schüler mussten dazu die Bedeutung von 12 in einem Text über Neuseeland vorkommenden Fachbegriffe angeben. Dabei sollten sie sechs der 12 Begriffe in der L1 und die restlichen sechs in der L2 definieren, unabhängig davon, ob der Text in der L2 (Testgruppe) oder in der L1 (Kontrollgruppe) vorgelegt worden war. Die Texte wurden zuvor gelesen und gehört. Die Definitionen erfolgten unmittelbar im Anschluss jeweils mündlich und wurden im Hinblick auf Korrektheit und Vollständigkeit analysiert. Es zeigte sich, dass die bilingualen Schüler den Regelschülern im Hinblick auf das Verständnis der dargebotenen Fachbegriffe unterlegen waren. Ihre Definitionen waren deutlich häufiger unvollständig und fehlerhaft als die der Regelschüler. Während 70 % der Definitionen der L1-Schüler von hoher Qualität waren, war dies nur bei etwa 20 % der L2-Schüler der Fall. Hier wiesen über 40 % der Definitionen erhebliche semantische Mängel auf (Gablasova, 2015, S. 67). Dabei ist zu berücksichtigen, dass das L2-Niveau der bilingualen Schüler nach 3.5 Jahren Teilnahme am bilingualen Unterricht bei B2/C1 lag und als durchaus hoch eingestuft werden kann. Gablasova

(2015, S. 72) empfiehlt deshalb weitere pädagogische Maßnahmen im Bereich des Erwerbs von Fachbegriffen im bilingualen Unterricht, da der einmalige Kontakt selbst in erklärenden Kontexten und bei einem hohen L2-Niveau der Lerner offenbar nicht ausreichen, um ein angemessenes Verständnis zu gewährleisten.

4.3 Auswirkung auf die Motivation

Im Folgenden soll ein Überblick über den Stand der Forschung zur Motivation von Schülern im bilingualen Unterricht gegeben werden. Dazu werden zunächst Studien zu allgemeinen, sachfachübergreifenden Effekten des neuen Lernsettings beschrieben (Abschnitt 4.3.1), bevor Untersuchungen zu den mathematisch-naturwissenschaftlichen Sachfächern (Abschnitt 4.3.2) sowie zum Fach Chemie (Abschnitt 4.3.3) vorgestellt werden, welche die besondere motivationale Ausgangslage in diesen Fächern berücksichtigen. Die Motivation stellt im Kontext des bilingualen Unterrichts ein recht junges Forschungsfeld dar und wurde lange vernachlässigt. Mittlerweile kommen unterschiedliche forschungsmethodische Verfahren zur Anwendung. In qualitativen Studien werden dabei meist auf der Grundlage von Unterrichtsbeobachtungen und Interviews Fallbeispiele untersucht (z.B. Abendroth-Timmer, 2007). Sehr viel größere Fallzahlen werden in den quantitativen Studien betrachtet, bei denen überwiegend Fragebögen für Lerner wie Lehrer eingesetzt werden. Die zu untersuchenden motivationalen Konstrukte werden dabei meist nach einem Ranking-Verfahren erfasst, bei dem mehrere Items zu Einsatz kommen (z.B. Fehling, 2005; Dallinger, 2015). Häufig werden solche quantitativen Untersuchungen auch durch qualitative Studien, im Sinne einer methodischen Triangulation, ergänzt. In diesem Fall kommen zusätzlich meist offene Fragen und leitfadengestützte Interviews zum Einsatz (z.B. Verriere, 2014; Lipski-Buchholz, 2019).

4.3.1 Motivation für den bilingualen Unterricht

Die positiven Auswirkungen bilingualen Unterrichts auf die Motivation von Schülern und Lehrern sind heute weitgehend unbestritten und zählen für nicht wenige Autoren zu den bedeutendsten Effekten dieser Unterrichtsform (vgl. Coyle D. , 2006, S. 11). Insbesondere bei Eltern und Schülern ist das Interesse am bilingualen Unterricht groß. Die Zunahme und Ausdifferenzierung bilingualer Unterrichtsangebote in den letzten Jahren sind ein Beleg dafür (vgl. Abschnitt 2.2). Bilingualer Unterricht wird von den Schülerinnen und Schülern oft als sinnvoll und nützlich bewertet (Ohlberger & Wegner, 2018, S. 80).

Auch Lehrer erfahren zusätzliche Motivation, weil sie den Diskurs mit Kollegen aus dem Sachfach bzw. aus der Fremdsprache als gewinnbringend für ihre pädagogische Arbeit erleben. Coyle (2006, S. 12) stellt dazu fest, dass es zu sehr fruchtbaren Wechselwirkungen kommen kann, weil sich die gesteigerte Motivation der Lehrer auf die Schüler überträgt und umgekehrt. In verschiedenen Untersuchungen konnte die positive Einstellung von Lehrern und Schülern nachgewiesen werden (z.B. Dalton-Puffer, 2007; Viehbrock, 2007). Laut Heine (2013, S. 219) geben Schüler in der Regel an, dass sie den bilingualen Unterricht als motivationssteigernd und interessanter erleben und sich durch die L2 nicht behindert fühlen. Darüber hinaus zeigt sich, dass der bilinguale Unterricht den Schülern Spaß macht. Meyer (2002) untersuchte Schüler aus vier achten Klassen, sechs neunten und drei zehnten Klassen (N=292) an vier Gymnasien mit bilingualem Zweig in Rheinland-Pfalz. Dabei zeigte sich, dass die Mehrheit der untersuchten Schüler angab, im bilingualen Unterricht Spaß zu haben, auch wenn dieser teilweise weniger gut verständlich und anstrengender war als der Regelunterricht. Je älter die Schüler waren, umso verständlicher und weniger anstrengend war der bilinguale Unterricht für sie. Vor allem die älteren Schüler fanden den bilingualen Erdkundeunterricht besser als den Regelunterricht. Dabei stellt Meyer (2002) jedoch durchaus klassen- bzw. schulspezifische Unterschiede fest. Auch Abendroth-Timmer (2007), die Schüler zwischen 14 und 18 Jahren, die an unterschiedlichen bilingualen Modulen teilgenommen hatten, befragte, weist auf die positive Wirkung des bilingualen Modulunterrichts auf die Motivation der Schüler hin. Die Ergebnisse ihrer qualitativen Studie, die sie als Hypothesen verstanden wissen will, deuten an, dass sachfach- und fremdsprachenorientierte Schüler gleichermaßen für den bilingualen Unterricht motiviert werden können, da diese Unterrichtsform beide Schwerpunktinteressen bedient. Die von ihr beobachteten Schüler lassen sich bereitwillig auf die neue Lernform ein, wenn sie das Bewusstsein erlangen, dem Unterricht sprachlich gewachsen zu sein. Ältere Schüler können darüber hinaus über die Sachfachinhalte oder durch schulexterne Ziele motiviert werden, die der Berufs- oder Studienvorbereitung dienen. Auch Held (2017) findet ein hohes Maß an instrumenteller Motivation bei italienischen Schülern. Bei der von ihm untersuchten Variante bilingualen Unterrichts werden über einen Zeitraum von bis zu fünf Jahren zwischen zwei und vier Fächer in der Fremdsprache Englisch unterrichtet. Er stellt aber auch fest, dass insbesondere Schüler, die sich weniger für Fremdsprachen interessieren oder über ein geringes akademisches Selbstkonzept verfügen, dem bilingualen Unterricht durchaus kritisch begegnen und befürchten, überfordert zu werden (Held 2017). Seine Untersuchung an vier weiterführenden Schulen zeigte zudem, dass die Angst vor dem bilingualen Unterricht dort weit verbreitet ist. Fast 90 % der Schüler gaben an, wenigstens eine Form von Angst erlebt zu haben. Dabei waren „Testangst“, „Angst vor der fremdsprachlichen Kommunikation“ und „Angst vor negativer Bewertung“ die häufigsten

Formen; diese traten bei 20 – 30 % der befragten Schüler (N=404) auf. Smit (2008) meint sogar, dass Schüler geradezu von der Aufgabe überwältigt sein können, sachfachliche Inhalte in der Fremdsprache zu erarbeiten. Otwinowska (2013) stellt dazu jedoch fest, dass negative Auswirkungen auf die Lernbereitschaft der Schüler vor allem dann zu erwarten sind, wenn der bilinguale Unterricht zu früh eingeführt wird und seine methodische Umsetzung nicht klar durchdacht ist.

4.3.2 Motivation für die mathematisch-naturwissenschaftlichen Sachfächer

Die Naturwissenschaften und insbesondere Chemie und Physik gehören nach wie vor zu den unbeliebtesten Fächern unter Schülern. Sie gelten als „kalt“, abstrakt und haben nur wenig Bezug zur Lebenswirklichkeit der Schüler. Schon 2005 rangierte Chemie laut Sasol-Studie zusammen mit Physik auf dem letzten Platz in der Rangfolge der beliebtesten Unterrichtsfächer (Becker, Hildebrandt, & Kühlmann, 2012, S. 353). PISA 2015 berichtet sogar von einem Rückgang des Interesses an den Naturwissenschaften seit 2006 sowie weniger positiv ausgeprägten Selbstbildern (Schiepe-Tiska, Simm, & Schmidtner, 2016). Die instrumentelle Motivation liegt 2015 dabei unter dem OECD-Durchschnitt. Nur die Hälfte der Jugendlichen ist der Meinung, dass die Naturwissenschaften in ihrem späteren Leben wichtig sein werden (Schiepe-Tiska, Simm, & Schmidtner, 2016). Im internationalen Vergleich fällt Deutschland dabei nach wie vor durch ausgeprägte Geschlechterdifferenzen zugunsten der Jungen auf. Die Auswirkungen bilingualen Unterrichts auf die Motivation in den naturwissenschaftlichen Sachfächern sind auch deshalb in den letzten Jahren vermehrt zum Gegenstand didaktischer Forschung geworden. Ergebnisse liegen jedoch erst vereinzelt vor und sind teilweise widersprüchlich. Für die mathematisch-naturwissenschaftlichen Fächer gibt es empirische Studien zu den Fächern Physik (Piesche, 2016), Biologie (Hartmannsgruber, 2014; Rodenhauser, 2016; Ohlberger, 2019) und Mathematik (Prüfer, 2012 bzw. Verriere, 2014 und Lipski-Buchholz, 2019). Vergleichbare Untersuchungen zu den Auswirkungen bilingualen Unterrichts auf die Motivation im Fach Chemie fehlen bislang. Hartmannsgruber (2014) kommt in Bezug auf die Motivation bilingualer Schüler im Biologieunterricht zu einem positiven Fazit, denn er stellt fest, dass auch Schüler, die zufällig ausgewählt wurden und sich nicht für die Teilnahme am bilingualen Unterricht selbst entschieden haben, durch den bilingualen Unterricht nicht demotiviert werden (Hartmannsgruber, 2014, S. 94). In seiner Untersuchung zu einem Modul zum Thema *Wasser* an der Realschule (vgl. Abschnitt 4.2) stellte er keine Nachteile in der Motivationsentwicklung der Siebtklässler fest. Mithilfe der *Kurzskala intrinsischer Motivation (KIM)* (Wilde, Bätz, Kovaleva & Urhahne, 2009) erhob Hartmannsgruber (2014) das *Interesse*, die *wahrgenommene Kompetenz*, die *wahrgenommene*

Wahlfreiheit und den *empfundene Druck* der Schüler. Zu keinem der drei Messzeitpunkte ermittelte er dabei signifikante Unterschiede zwischen den mono- und bilingual unterrichteten Schülern. Es zeigten sich lediglich tendenzielle Vorteile zugunsten der monolingualen Versuchsgruppe in den Bereichen Interesse, wahrgenommene Kompetenz sowie Druck. Scheersoi (2008) kam in ihrer Untersuchung zum bilingualen Biologieunterricht in einer neunten Klasse eines Gymnasiums zu ähnlichen Ergebnissen. Dabei setzte sie eine Mischung aus quantitativen und qualitativen Methoden wie Fragebögen, Lerntagebuch, Motivationskurve und Notizen der Lehrkraft ein. Die Auswertung ergab für 20 Lernende keine Veränderung der Motivation, während sieben Lernende eine positive Veränderung beschreiben und nur drei Lernende von Motivationsverlusten berichten. Rodenhauser (2016) konnte die positiven Ergebnisse auch für Oberstufenschüler (N=490) bestätigen. Sie befragte in ihrer quasi-experimentellen Interventionsstudie Schüler unterschiedlicher Biologiekurse, die an einem Labortag zum Thema *genetischer Fingerabdruck* teilgenommen hatten und verglich die Antworten mit denen von Schülern, die an einem Labortag auf Deutsch, sowie mit Schülern, die an keinem Labortag teilgenommen hatten. Wie Hartmannsgruber (2014) greift sie dabei auf die Kurzskala zur intrinsischen Motivation (KIM) (KIM) zurück und stellt bei insgesamt hohen Werten keine statistisch signifikanten Unterschiede zwischen den bilingual und den monolingual unterrichteten Schülern fest. Die Verwendung der Fremdsprache wirkt sich demnach nicht negativ auf die Motivation der Schüler aus. Auch für das aktuelle Biologieinteresse, welches Rodenhauser (2016) sechs Wochen im Anschluss an den Laborbesuch ermittelt, finden sich keine statistisch signifikanten Unterschiede. Anders als Hartmannsgruber (2014) und Rodenhauser (2016) ermittelt Piesche (2016) für das Fach Physik durchaus motivationale Nachteile für bilingualen Schüler. Bei ihrer Untersuchung zu einem Modul zum Thema *Schwimmen und Sinken* an der Realschule konstatiert sie insgesamt sowohl für die mono- als auch für die bilingual unterrichteten Sechstklässler überdurchschnittlich hohe Werte für die motivationalen Konstrukte *Kompetenzerleben*, *Angst*, *intrinsische Motivation* und *kognitive Mitarbeit*. Jedoch ermittelte sie im Vergleich signifikant schlechtere Werte bei den bilingual unterrichteten Schülern im Bereich *Kompetenzerleben* und *Angst*. Auch die Werte für die intrinsische Motivation und die kognitive Mitarbeit wiesen auf tendenzielle Nachteile der bilingualen Schüler hin, wenngleich diese Unterschiede statistisch nicht signifikant waren (Piesche, 2016, S. 153). Otwinoska und Forsys (2015) verzeichnen in ihrer Untersuchung zu polnischen Grundschulern (N=140) ebenfalls eine zunehmende Angst und eine intellektuelle Hilflosigkeit im bilingualen Unterricht in den Naturwissenschaften und im Fach Mathematik. Dabei zeigt sich, dass gerade das problemlösende Denken unter den negativen Emotionen leidet. Otwinowska (2013) führt dies vor allem auf Mängel

in der Unterrichtsplanung und die Verwendung sprachlich unangemessener Arbeitsmaterialien zurück. Insbesondere leistungsschwächere Schüler mit schlechten Noten in den Naturwissenschaften bzw. Mathematik sind dann leicht überfordert. Ohlberger und Wegner (2019, S. 8) erfassen in ihrer Untersuchung unter Zehntklässlern, die an zwei bilingualen Modulen im Fach Biologie teilgenommen hatten, in Bezug auf das Fach auch Biologie motivationale Nachteile, insbesondere bei Schülern mit einer hohen Angst vor Englisch bzw. mit einer geringen Selbstwirksamkeitserwartung. Für Schüler mit einer großen Angst vor dem Gebrauch der englischen Sprache ermitteln sie zwar eine Abnahme dieser Angst sowie eine tendenziell bessere Selbstwirksamkeitserwartung im Anschluss an den Modulunterricht, die Affinität für das Fach Biologie nimmt jedoch bei diesen Schülern nach der Teilnahme am zweiten Modul statistisch signifikant ab. Bei diesen Schülern handelt es sich laut Ohlberger und Wegner überwiegend um jene, die Biologie in besonderem Maße mögen und die Autoren vermuten, dass sie den Einbezug der Fremdsprache vermutlich deshalb ablehnen, weil er die Aufmerksamkeit von biologischen Themen ablenkt (Ohlberger & Wegner, 2019, S. 10). Auch für Schüler mit einer geringen Selbstwirksamkeitserwartung finden Ohlberger und Wegner zwar ein verbessertes Erleben dieser Selbstwirksamkeit im Anschluss an das Modul, gleichzeitig verlieren auch diese Schüler nach dem zweiten Modul tendenziell ihre Affinität für das Fach Biologie. Damit weisen Ohlberger und Wegner auf zwei Gruppen von Schülern hin, deren Biologiemotivation unter den Bedingungen bilingualen Unterrichts leiden kann, nämlich solche mit einer großen Sprachangst und solche mit einer geringen Selbstwirksamkeitserwartung für den bilingualen Unterricht. Gleichzeitig stellen sie aber fest, dass für die Schüler der anderen Gruppen keine nennenswerten motivationalen Einbußen zu verzeichnen sind. So bleiben für Schüler mit einer geringen Angst vor Englisch die Affinität für Biologie sowie die Selbstwirksamkeitserwartung auch nach den beiden Modulen nahezu identisch. Ähnliches gilt für Schüler mit einer hohen Selbstwirksamkeitserwartung. Hier nehmen Biologieaffinität und Selbstwirksamkeitserwartung im Anschluss an den bilingualen Unterricht nur geringfügig ab und liegen auf einem hohen Niveau (Ohlberger und Wegner, 2019). In ihrer Pilotstudie zum bilingualen Modulunterricht im Fach Mathematik untersucht Prüfer (2012) die *Lernbereitschaft* von Schülern der siebten Klasse (N=29), die über sechs Wochen an einem Modul zum Thema *antiproportionale Zuordnung* teilgenommen hatten. Den Begriff der Lernbereitschaft führt Prüfer (2012) auf die Grundbedürfnistheorie von Deci und Ryan (2000) zurück. Demnach tritt intrinsische Motivation dann auf, wenn die drei Grundbedürfnisse *Kompetenzerleben*, *soziale Eingebundenheit* und *Autonomieerleben* erfüllt sind. Prüfer (2012) ergänzt für ihr Konzept der Lernbereitschaft den Aspekt der *Zufriedenheit*, da laut Krapp (2000) die intrinsische Motivation mit einem Gefühl der Zufriedenheit einhergeht. Die Schüler wurden während des Moduls dreimal

anhand von Fragebögen zu ihrem Verständnis (Kompetenzerleben), Meldeverhalten (soziale Eingebundenheit), Autonomieerleben (empfundene Schwierigkeit) und ihrer Freude am Mitmachen (Zufriedenheit) befragt. Prüfer (2012) beobachtet, dass die Lernbereitschaft signifikant mit der Beliebtheit des Faches korreliert. Schüler, die angegeben hatten, ein bestimmtes Fach, in diesem Fall Mathematik oder Englisch besonders zu mögen, erwiesen sich auch in diesem Fach als lernbereiter. Im Rahmen ihrer Hauptuntersuchung (Verriere, 2014) konnten diese Ergebnisse jedoch nicht bestätigt werden und es ergab sich ein differenzierteres Bild. Verriere (2014) untersuchte sechs Klassen der Jahrgangsstufen 7 bis Q1 an unterschiedlichen Schulen (N=141). Es handeltet sich dabei überwiegend um Gymnasien, wobei zwei Klassen aus bilingualen Zweigen stammten. Der Modulunterricht wurde vom Fachlehrer oder von einem anderen Lehrer für ca. vier Wochen kurz vor den Sommerferien durchgeführt. Dabei wurden sehr unterschiedliche Themen unterrichtet. Wie in der Pilotstudie, ermittelte Verriere (2014) ihre Daten zur Lernbereitschaft anhand von Schülerfragebögen und Lehrerinterviews. Verriere (2014) wertet die Daten der sechs Klassen dabei mithilfe von quantitativen und qualitativen Methoden im Rahmen von Fallstudienanalysen aus und fasst ihre Ergebnisse abschließend zusammen. Für die betroffenen Schüler hatte Englisch einen sehr hohen Stellenwert, was auch darauf zurückzuführen ist, dass ein Teil der Schüler aus bilingualen Zweigen stammte. Dem Fach Mathematik hingegen standen diese Schüler mit sehr gemischten Gefühlen gegenüber. Betrachtet man die Ergebnisse ganzer Klassen, gibt es solche, in denen die Lernbereitschaft weitgehend unverändert bleibt oder abnimmt aber auch jene, in denen sie steigt. Letzteres war in drei der sechs betrachteten Klassen der Fall (Verriere, 2014, S. 250f.). Auch bei der Freude am Modul finden sich Klassen, in denen diese Freude im Mittel steigt bzw. sinkt. Dabei ist keine Korrelation zwischen der Freude am Modul und der Lernbereitschaft zu verzeichnen, so dass die These, mehr Freude am Modul führe auch zu einer größeren Lernbereitschaft, nicht belegt werden konnte. Den Grund für den Rückgang der Lernbereitschaft sieht Verriere (2014, S. 251) aber nicht zwangsläufig im Einsatz der Fremdsprache. Vielmehr scheint die Lage des Moduls am Ende des Schuljahres hier eine Rolle zu spielen. Auch in anderen Fächern geht die Lernbereitschaft kurz vor den Sommerferien in der Regel deutlich zurück. Hinzu kommt, dass auch die Teilnahme am Unterricht sehr unregelmäßig war, was, wie die Lehrer in den Interviews bestätigten, auf die bevorstehenden Ferien zurückzuführen ist (Verriere, 2014, S. 249). Somit kann auch die Abwesenheit im Unterricht dafür verantwortlich sein, dass Schüler sich weniger kompetent und autonom erlebten und ihre Lernbereitschaft deswegen sank. Zudem war der tatsächliche Anteil der Fremdsprache in den sechs Modulen laut Auskunft der Schüler stark unterschiedlich. Dieser war insbesondere in der Oberstufe bei Themen mit hohem mathematischem Abstraktionsgrad und einem damit einhergehenden stärkeren Rückgriff auf mathematisch-abstrakte Darstellungsformen deutlich

geringer als in der Mittelstufe. Einen Zusammenhang zwischen der Freude am Modul bzw. der Lernbereitschaft und dem Einsatz der Fremdsprache herzustellen, ist daher aus den von Verriere (2014) ermittelten Daten nicht möglich.

Auswirkungen von Interessensschwerpunkt und Geschlecht

Rodenhauser (2016) untersuchte den Einfluss des Fachinteresses auf die Motivation im bilingualen Biologieunterricht. Für die unterschiedlichen Schülertypen („Naturwissenschaftler“, „Fremdsprachler“ und „Allrounder“), die Rodenhauser (2016) mittels Clusteranalyse bestimmt, kann sie dabei höchst signifikante Unterschiede hinsichtlich der intrinsischen Motivation feststellen. So findet sie bei den „Naturwissenschaftler“ die größte Freude an den biologischen Kursanteilen, während die „Fremdsprachler“ im Labor den größten Druck hinsichtlich der biologischen Kursanteile und die „Naturwissenschaftler“ den größten Druck hinsichtlich der fremdsprachlichen Kursanteile empfinden. Für bilingual unterrichtete Schüler mit einer fremdsprachlichen Disposition („Fremdsprachler“) bemerkt Rodenhauser (2016, S. 199) außerdem eine statistisch signifikante Steigerung des biologischen Fähigkeitsselbstkonzeptes. Sie führt dies darauf zurück, dass sprachlich begabtere Schüler besser mit den Versuchsanleitungen und den Hintergrundtexten zurechtkommen und entsprechend erfolgreicher experimentieren (vgl. Rodenhauser & Preisfeld, 2016, S. 106f.). Da Fähigkeitsselbstkonzept und dispositionales Interesse häufig korrelieren, vermutet Rodenhauser (2016), dass sich ein längerer Kontakt mit der Fremdsprache im Fach Biologie durchaus positiv auf das Biologieinteresse auswirken könnte. Für Scheersoi (2008) ergibt sich in ihrer Studie zum bilingualen Biologieunterricht ein uneinheitliches Bild. So sind unter den sieben Schülern, deren Motivation sich verbessert hat, vier sprachorientierte Lernende, ein sachfachorientierter Lernender und zwei Lernende ohne Vorlieben. Unter den drei Schülern mit Motivationsverlusten befindet sich auf der anderen Seite je ein sachfachorientierter Lernender, ein sprachorientierter Lernender und ein Lernender ohne Vorliebe. Verriere (2014) unterschied in ihrer Studie vier Interessensgruppen: „mathematikab-“ bzw. „mathematikzugewandt“ (mas und mzs) sowie „englischab-“ bzw. „englischzugewandt“ (eas und ezs) und wertete ihre Ergebnisse im Hinblick auf die Zugehörigkeit der Schüler zu einer dieser Gruppen aus. Die mathematikzu- bzw. mathematikabgewandten Schüler reagieren insgesamt positiv auf das Modul. Erstere werden, entgegen den Befürchtungen, nicht durch den Einsatz der Fremdsprache abgeschreckt und ihre Lernbereitschaft sinkt nicht. Mathematikabgewandte Schüler hingegen profitieren sogar in verschiedenen Kategorien vom Modul, insbesondere in ihrer Zufriedenheit (Verriere, 2014, S. 253). Auf der anderen Seite profitieren englischzugewandte Schüler nicht zwangsläufig vom Modul. Ihre Lernbereitschaft hat im Verlauf des Moduls in mehreren Klassen abgenommen. Auch die (Vor-)Freude auf das Modul

war bei diesen Schülern sowie bei den mathematikabgewandten Schülern nicht größer als in den anderen Gruppen. Sie erklärt dies so, dass "die Integration einer als positiv empfundenen Fremdsprache in den negativ empfundenen Mathematikunterricht eher auf Ablehnung als auf Zuneigung stoßen kann" (Verriere, 2014, S. 253). Anders als Verriere (2014) findet Lipski-Buchholz (2017) in ihrer Untersuchung zu Siebt- und Neuntklässlern, die an einem bilingualen Modul im Fach Mathematik teilgenommen hatten, dass die motivationale Disposition mathematisch-naturwissenschaftlich orientierter Lerner im bilingualen Mathematikunterricht erkennbar leidet, und im regulären Unterricht eindeutig besser ist. Für viele der von ihr untersuchten Konstrukte, wie extrinsische Motivation, Selbstkonzept, Selbstwirksamkeit, Zufriedenheit, Einstellung zum Unterricht bzw. zum Mathematiklernen, Kompetenzerleben und Aktivität im Unterricht konstatiert sie deutlich schlechtere Werte im bilingualen als im Regelunterricht (Lipski-Buchholz, 2019, S. 211). Die unterschiedliche Vorgehensweise bei der Interessensgruppenbildung könnte dabei laut Lipski-Buchholz (2019) die Ursache für die gegenläufigen Ergebnisse sein. Bei Verriere (2014) sind die Interessensgruppen nämlich nicht überschneidungsfrei, da sie innerhalb der mathematikzugewandten Lernergruppe nicht nach Schülern differenziert, die englischzu- oder englischabgewandt sind. Dabei ist es laut Lipski-Buchholz (2019) durchaus denkbar, dass die Einstellung zum Englischunterricht einen Einfluss auf die Lernbereitschaft im bilingualen Mathematikunterricht ausübt. Auf der anderen Seite stellt sie für fremdsprachlich interessierte Schülerinnen und Schüler eine positivere motivationale Disposition als im regulären Mathematikunterricht fest. In der Hinzunahme der Fremdsprache im Mathematikunterricht sieht sie deshalb einen motivationalen Zugewinn. Dabei bewerten fremdsprachlich interessierte Schüler die Mathematik im bilingualen Modul als "interessant". Insgesamt äußern sich diese Schüler jedoch nicht erkennbar positiv über das mathematische Lernen als solches. Auch in Bezug auf das sachfachliche Lernen stellt Lipski-Buchholz (2019) deshalb fest, dass eine Motivierung für diesen unbeliebten Lernbereich durch das Modul nicht unmittelbar erzielbar war. Gleiches gelte auch für das fremdsprachliche Lernen und sie führt die "emotionale Distanz" der Fächer Mathematik und Englisch als mögliche Ursache an. Lipski-Buchholz (2019, S. 264) resümiert daher:

Insgesamt zeigt an dieser Stelle der bilinguale Mathematikunterricht sein motivationales Potenzial für die fremdsprachlich interessierten Schülerinnen und Schüler. Hier gilt es aber zukünftig noch zu untersuchen, ob sie bei der unterrichtlichen Beteiligung auf dem Niveau stehen bleiben, dass sie sich nur bei sprachenbezogenen Fragen einbringen, oder ob sie auch langfristig für eine aktivere Auseinandersetzung mit den mathematischen Inhalten gewonnen werden können (Lipski-Buchholz, 2019, S. 264).

Zur Frage, welche Rolle das Geschlecht bei der Motivation für den bilingualen Unterricht spielt, liegen unterschiedliche Ergebnisse vor. Zum einen zeigen Studien, dass deutlich mehr Mädchen als Jungen am bilingualen Unterricht teilnehmen (vgl. z.B. Zydati, 2007, S. 73), zum anderen finden Schlemminger und Buchmann (2013) in ihrer Erhebung zur Einstellung gegenüber dem bilingualen Sachfachunterricht an Realschulen, dass sich mehr Jungen für den bilingualen Unterricht interessieren als Mädchen. Für die mathematisch-naturwissenschaftlichen Sachfächer sind die Daten derzeit noch dürftig. Prüfer (2012) etwa untersucht die Wirkung des Geschlechts auf die Motivation im bilingualen Mathematikunterricht. Sie ermittelte schon zu Beginn des Moduls geschlechtsspezifische Unterschiede zu Gunsten der Mädchen, die signifikant häufiger angegeben hatten, Englisch mehr zu mögen Mathematik. Es zeigte sich außerdem, dass die „Freude am Mitmachen“ bei Mädchen im Vergleich zum Regelunterricht signifikant größer war. Jungen wie Mädchen gaben zwar an, im bilingualen Unterricht Mathematik genauso viel verstanden zu haben wie im Regelunterricht (Prüfer, 2012, S. 158f.), es zeigte sich aber auch, dass während das „Verständnis“ und die „Freude am Mitmachen“ im Laufe des Moduls bei den Mädchen anstieg, bei den Jungen in beiden Kategorien ein Abwärtstrend zu beobachten war. Im Rahmen ihrer Hauptstudie konnte Verriere (2014) diese Ergebnisse jedoch nicht bestätigen und stellt fest, dass Mädchen weder mehr (Vor-)Freude auf das bilinguale Modul erleben noch in ihrer Lernbereitschaft positiver reagieren als Jungen. Andererseits ergeben sich für die Mädchen tendenziell bessere Werte für die Zufriedenheit und die empfundene Schwierigkeit. Der Vermutung, dass Mädchen eher vom bilingualen Mathematikunterricht profitieren als Jungen, kann sie deshalb nur teilweise zustimmen. Auch Piesche (2016, S. 142) findet in ihrer Untersuchung zum bilingualen Physikunterricht keine Hinweise darauf, dass Mädchen vom bilingualen Unterricht stärker profitieren als Jungen. Die von ihr erhobenen Daten bezogen sich auf die Variablen Kompetenzerleben, Angst, intrinsische Motivation und kognitive Mitarbeit im bilingualen Unterricht (Piesche, 2016, S. 132). Sie weist jedoch darauf hin, dass die von ihr untersuchten Sechstklässler möglicherweise zu „jung waren, und geschlechtsspezifische Identitäten, die sich v.a. in der Pubertät entwickeln, noch nicht genug ausgereift waren“ (Piesche, 2016, S. 142).

4.3.3 Erste Ergebnisse zur Motivation im bilingualen Chemieunterricht

Zur Motivation im bilingualen Chemieunterricht liegen bisher kaum Ergebnisse vor. Es fehlen insbesondere Studien, bei denen die Motivation systematisch und auf breiter Datenbasis erforscht wird. Die vorliegenden Erkenntnisse beziehen sich vor allem auf Unterrichtsbeobachtungen von Lehrenden oder auf Auskünfte von Schülern aus Interviews oder Fragebögen (Klingauf, 2002). Die Ergebnisse wurden dabei überwiegend an kleinen Stichproben gewonnen. Klingauf (2002, S. 60)

befragte im Anschluss an einige Wochen bilingualen Chemieunterrichts Achtklässler eines gymnasialen Zweigs einer Gesamtschule (N=17) zu ihrem Unterrichtserleben. Die Schüler gaben mehrheitlich an, neugierig auf den bilingualen Unterricht in Chemie zu sein und „Lust darauf“ gehabt zu haben. Den Spaß am bilingualen Unterricht führten die Schüler jedoch überwiegend nicht explizit auf das Englische im Chemieunterricht zurück. Kemper und Becker (2015) beobachten jedoch in ihrer Fallstudie zum bilingualen Chemie- und Physikunterricht in Frankreich, dass die fremdsprachenorientierten Schüler sich insbesondere für die fremdsprachlichen Anteile bilingualen Unterrichts interessieren. Obwohl die Schüler in Klingaufs (2002) Studie das Englische im bilingualen Unterricht nicht als übermäßig anstrengend empfanden und die englischsprachigen Anteile im Unterricht gut verstehen konnten, waren sie nicht sicher, ob sie ihre Leistungen im Fach Englisch durch den bilingualen Unterricht verbessern können. Nur sechs der 17 Schülern fanden, dass dies so sei, weitere sechs waren unentschieden und die restlichen fünf waren der Meinung, dass dies nicht so sei. Einen späteren Nutzen aufgrund der Verwendung des Englischen im Chemieunterricht sahen die meisten Schüler sehr wohl in Klingaufs (2002) Studie, und die Mehrheit gab an, gerne ein weiteres Chemiemodul machen zu wollen. Auch Rittersbacher (2006) beobachtet in ihrer Untersuchung zu einem bilingualen Chemiemodul zum Thema *Investigating the candle* in Klasse 9 einer Realschule eine erhöhte Motivation in der bilingualen Lerngruppe. Sowohl in der bilingual unterrichteten Gruppe als auch in der auf Deutsch unterrichteten Vergleichsgruppe führen die Schüler selbstgeplante Experimente durch und werten diese anhand bestimmter Leitfragen aus. Rittersbacher (2006, S. 35) weist auch auf das in der Untersuchungsgruppe veränderte Arbeitsverhalten hin, in der die klassische Rollenverteilung beim Experimentieren aufgebrochen wurde und sich neue Kooperations- und Kommunikationsformen ergaben. Sie führt dies auch auf die Einführung der aus dem Fremdsprachenunterricht bekannten Rolle des *language inspectors*, zurück, die eine noch intensivere Zusammenarbeit der Schüler zur Folge hatte (Rittersbacher, 2006, S. 35). Es kann vermutet werden, dass sich Schüler dadurch autonomer erleben, weil auch die zusätzlichen sprachlichen Herausforderungen gemeinsam und ohne weitere Steuerung von außen gemeistert werden. Hülten (2014) untersuchte den fremdsprachlichen und den sachfachlichen Lernerfolg von bi- und monolingual unterrichteten Siebtklässlern eines Gymnasiums im Anschluss an ein vierwöchiges Modul zum Thema *Oxidation and combustion*. Er verzeichnet dabei ebenfalls Hinweise auf ein hohes Kompetenzerleben der bilingualen Schüler. Diese gaben an, den Unterricht nicht als zu schwer zu empfinden und nicht den Eindruck haben, weniger gelernt zu haben als im Regelunterricht (Hülten, 2014, S. 120). Diese Einschätzung belegen auch die Ergebnisse der im Anschluss an das Modul durchgeführten Chemietests. Hülten (2014) verzeichnete außerdem Hinweise darauf, dass chemieabgewandte Schüler durch den bilingualen Unterricht zusätzlich

motiviert werden. Rund ein Fünftel der Schüler, die den bilingualen Unterricht nach eigener Auskunft interessant fanden, hatte nämlich im Vorfeld angegeben, Chemie eigentlich nicht zu mögen. Die große Mehrheit der bilingualen Schüler äußerte darüber hinaus den Wunsch, häufiger bilinguale Module im Fach Chemie durchzuführen (vgl. Hülten, 2014, S. 120ff.). Während Klingauf, Hülten, Bonnet und Rittersbacher den bilingualen Chemieunterricht in den Klassen 7, 8, 9 und 10 untersuchten, nehmen Fehling und Bardl im Schuljahr 2004/5 laut Brunnert et al. (2009) den Chemieunterricht an einer höheren Berufsfachschule in den Blick. Die Gruppe, bestehend aus Chemisch-Technischen Assistenten im zweiten Ausbildungsjahr, wurde dabei über sieben Stunden in einem bilingualen Modul zum Thema *Atomic absorption spectroscopy* unterrichtet (Brunnert, Harsch, & Heimann, 2009, S. 57ff.). Obwohl die Klasse fremdsprachlich sehr heterogen zusammengesetzt war, konnten die Schüler vergleichbar gute Leistungen erzielen wie monolingual unterrichtete Schüler, sie waren aber wesentlich motivierter als im regulären Englischunterricht. Offensichtlich, so die Autoren, hatten viele „englischmüde“ Schüler die Bedeutung des Englischen im internationalen naturwissenschaftlichen Kontext erkannt und sahen in dem Modul die Möglichkeit, diese in Bezug auf chemische Inhalte anzuwenden (vgl. Brunnert, Harsch, & Heimann, 2009, S. 59). Damit zeigt sich das bereits oben angesprochene hohe Maß an instrumenteller Motivation für das Englischlernen im bilingualen naturwissenschaftlichen Unterricht. Während die Mehrheit der Schüler positiv auf das Modul reagierte, es als "anspruchsvoller" aber gleichsam auch "einprägsamer" erlebten und sich insgesamt "herausgefordert fühlten", gab es auch kritische Stimmen insbesondere von sprachlich schwächeren Schülern, die sich "stellenweise überfordert" fühlten. Auch die Bitte, keinen prüfungsrelevanten Stoff im Modul zu unterrichten, deutet auf ein teilweise gesunkenes Kompetenzerleben im Fach Chemie hin (Brunnert, Harsch, & Heimann, 2009, S. 59). Pavisic (2011) berichtet über ähnlich positive Effekte des bilingualen Chemieunterrichts bei Oberstufenschülern am *Istituto Tecnico per Attività Sociali* "Galileo Galilei" in Jesi (AN), einer weiterführenden Schule mit technischem Schwerpunkt in der Provinz Ancona, Italien. Die siebzehnjährigen Schüler, die an einem Modul in organischer Chemie teilgenommen hatten, arbeiteten intensiver zusammen und zeigten eine insgesamt größere Mitarbeit. Die Schüler waren motiviert, die Fremdsprache anzuwenden, weil sie jetzt einen Sinn darin sahen, die Sprache zu beherrschen. Sowohl ihr Selbstvertrauen in Bezug auf die Fremdsprache als auch in Bezug auf Chemie stieg (Pavisic 2011).

5 Design der Untersuchung: Forschungsfragen und Methodologie

Der bilinguale Unterricht ist bis jetzt nicht adäquat unter gewöhnlichen Schulbedingungen mit repräsentativen Schülerpopulationen untersucht worden. Nicht nur Mikrostudien zu Prozessen der Konzeptbildung in einer Fremdsprache, Methodenuntersuchungen, Lehrer-Schüler-Interaktionen Lehrer-Schüler-Einstellungen wie Sozialisation sind nötig, sondern auch größere Meta-Studien, die feststellen, in welchem Umfang, mit welchen Methoden, ab welchem Sprachkompetenzniveau und mit welcher Lehrervorbereitung CLIL allen SuS zur Verfügung gestellt werden soll, um das derzeitige Niveau naturwissenschaftlicher Kompetenzen zu erhalten oder gar zu steigern (Aristov, 2013, S. 171)

Der vorliegenden Untersuchung liegt die Überzeugung zugrunde, dass eine Studie zum Erfolg von bilinguaem Unterricht berücksichtigen muss, welche Voraussetzungen die Schüler mitbringen und wie der Unterricht gestaltet wird. Hier bedarf es, so die Überzeugung des Autors, weiterer Forschungsbemühungen, welche die Erfolge bilingualen Unterrichts auch in Abhängigkeit von der Inszenierungsform setzen. Die vorliegende Arbeit soll hierzu einen Beitrag leisten. In den vorangegangenen Abschnitten 3.3 und 5.5 wurden deshalb die didaktisch-methodischen Prinzipien, nach denen das vorliegende Modul gestaltet wurde, beschrieben. Im Folgenden gilt es auszuführen, welche Voraussetzungen der Schüler mitbringen sollte und wie diese im Rahmen der Untersuchung erfasst werden (Abschnitt 5.4). Der vorliegende Abschnitt beginnt jedoch mit einer Beschreibung des Forschungsgegenstandes und der Rahmenbedingungen unter denen geforscht wird (Abschnitt 5.1). Nach einer Charakterisierung des methodologischen Ansatzes (Abschnitt 5.2) werden die gewählten Untersuchungsmethoden im Hinblick auf die drei Forschungsschwerpunkte der Studie konkretisiert (Abschnitt 5.3). Zuletzt wird das Modul, das der Analyse zugrunde liegt, vorgestellt, und das zu erwerbende Fachwissen und der fremdsprachliche Wortschatz skizziert (Abschnitt 5.5).

5.1 Forschungsgegenstand und Rahmenbedingungen

Im Rahmen der Untersuchung soll der Modulunterricht im Fach Chemie im Hinblick auf seine sachfachliche, (fremd-)sprachliche und motivationale Wirksamkeit hin überprüft werden. Aufgrund der aktuellen Forschungslage zum bilingualen Chemieunterricht (vgl. Kapitel 4) ergeben sich insgesamt drei Fragekomplexe:

Erster Fragekomplex: Leistungen im Sachfach	
Können im bilingualen Chemieunterricht im Sachfach vergleichbare Leistungen erzielt werden wie im Regelunterricht oder ist Chemie als Sachfach zu „schwer“ für den bilingualen Unterricht? Gibt es sogar eine Überlegenheit bilingualer Schüler, weil Inhalte tiefer semantisch verarbeitet werden und deshalb besser memoriert werden? Führt die größere kognitive Beanspruchung sogar dazu, dass auch kognitiv anspruchsvollere Aufgaben besser gelöst werden können? Welche Möglichkeiten und Grenzen ergeben sich auf der anderen Seite aufgrund der eingeschränkten Fremdsprachenkompetenz und der unterschiedlichen Motivation der Schüler? Welche weiteren Faktoren bzw. Faktorenkombinationen unterstützen den Lernerfolg im bilingualen Unterricht und welche behindern ihn? Inwieweit unterscheiden sich diese von den Prädiktoren, die den Lernerfolg im Regelunterricht bestimmen?	
Zweiter Fragekomplex: Sprachlicher Lernerfolg im Bereich Wortschatz	
Welche Lernzuwächse sind im Bereich des muttersprachlichen und des fremdsprachlichen Fachwortschatzes zu verzeichnen? Lernen bilinguale unterrichtete Schüler die deutschen Fachtermini genauso gut wie die Schüler des Regelunterrichts? Werden darüber hinaus auch die englischen Fachbegriffe von den bilingual unterrichteten Schülern gelernt? Welche Zugewinne sind im Allgemeinwortschatz feststellbar und wie nachhaltig sind diese? Stehen die erlernten Begriffe auch sechs Wochen nach Ende des Moduls noch zur Verfügung? Welche Unterschiede sind in Bezug auf die rezeptive bzw. die produktive Beherrschung der Lexik zu beobachten?	
Dritter Fragekomplex: Motivation im BU	
Wie unterscheidet sich die Motivation im bilingualen Modul von der Motivation im regulären Chemieunterricht? Gibt es Unterschiede auf Klassenebene bzw. je nach Orientierung oder Geschlecht der Schüler? Mit welchen Annahmen begegnen die Schüler bilinguaem Unterricht und wie ändern sich diese Annahmen im Anschluss an das Modul?	

Abbildung 4: Die drei zentralen Fragekomplexe der Untersuchung

An der Hauptuntersuchung nahmen insgesamt sechs achte Klassen eines nordrhein-westfälischen Gymnasiums teil (N=149). Drei Klassen wurden bilingual unterrichtet (Untersuchungsgruppe) und weitere drei Klassen auf Deutsch (Kontrollgruppe). Die Hauptuntersuchung fand in den Schuljahren 2018/19 sowie 2020/21 statt²⁵. Im Schuljahr 2018/19 wurden je zwei Klassen bilingual deutsch-englisch und eine Klasse auf Deutsch unterrichtet. Im Schuljahr 2020/21 erhielt eine Klasse bilingualen Unterricht und zwei Klassen wurden auf Deutsch unterrichtet. Die Stichproben waren mit N= 74 (Untersuchungsgruppe) und N= 75 (Kontrollgruppe) etwa gleich groß. Die Pilotierung erfolgte im Schuljahr 2017/18. An der Pilotuntersuchung nahmen zwei achten Klassen desselben Gymnasiums teils (N= 52). Die Untersuchungsgruppe wurde bilingual und die Kontrollgruppe auf Deutsch unterrichtet. Insgesamt nahmen damit 201 Schüler an der Untersuchung teil. In der Unterrichtsreihe geht es um die Elementfamilien der *Alkali- und Erdalkalimetalle*. Die Gegenstände des zehnstündigen Moduls gehören damit zu dem in Klasse 8 obligatorisch zu unterrichtenden Inhaltsfeld *Elementfamilien, Atombau und Periodensystem* des Kernlehrplans Chemie für die Sekundarstufe I in Nordrhein-Westfalen. In allen achten Klassen der Hauptuntersuchung wurde das Modul vom selben Lehrer unterrichtet, der gleichzeitig Chemielehrer der Klasse war. Der Modulunterricht erfolgte am Ende des ersten Schulhalbjahrs. Alle Klassen wurden in der gleichen Weise auf das Modul vorbereitet.

5.2 Methodologischer Ansatz

Im folgenden Abschnitt wird der methodologische Ansatz der vorliegenden Untersuchung vorgestellt. Ziel ist es, die Möglichkeiten und Grenzen des gewählten Forschungsansatzes im Hinblick auf die gewonnenen Erkenntnisse und ihre Aussagekraft aufzuzeigen.

Für die Studie wurde ein quasi-experimentelles Forschungssetting gewählt, bei dem die Schüler im Klassenverband unterrichtet und getestet werden. Anhand der so gewonnenen Daten werden im Vorfeld formulierte Hypothesen überprüft und anschließend bestätigt oder verworfen (hypothesentestendes Verfahren). In Abschnitt 5.2.1 wird nun zunächst das Forschungssetting genauer beschrieben. Die Verfahren zur Generierung und Testung von Hypothesen (Abschnitt 5.2.2) sowie zur Erfassung und Auswertung der Daten mithilfe statistischer Verfahren (Abschnitt 5.2.3) stehen im Mittelpunkt der beiden folgenden Abschnitte. Weil die gewonnenen Daten immer wieder nach Interessensgruppen differenziert ausgewertet werden, wird abschließend auch das Verfahren zur Bildung dieser Interessensgruppen vorgestellt (Abschnitt 5.2.4).

25 Aufgrund der coronabedingten Schulschließungen im Schuljahr 2019/20 konnte der Unterricht in diesem Schuljahr nicht wie geplant stattfinden.

5.2.1 Grundsätzliches

Der Frage nach der sachfachlichen, (fremd-)sprachlichen und motivationalen Wirksamkeit des bilingualen Unterrichts näherten sich Forscher in der Vergangenheit aus unterschiedlichen Richtungen. Dies hat dazu geführt, dass in den Studien zum bilingualen Unterricht heute eine Vielzahl unterschiedlicher Forschungsmethoden zum Einsatz kommen, wobei verstärkt auch auf Verfahrensweisen der empirisch-sozialwissenschaftlichen Forschung zurückgegriffen wird (Lipski-Buchholz, 2019, S. 65). Der Forderung Zydati (2012) nach mehr randomisierten, quasi-experimentellen Studien, wurde in den letzten Jahren verstärkt nachgekommen (z.B. Piesche, 2016; Hartmannsgruber; 2014; Haagen-Schtzenhfer, 2011). Dabei konnten umfangreiche Erkenntnisse zur Wirksamkeit des bilingualen Unterrichts in den Fchern Biologie und Physik gewonnen werden. Fr das Fach Chemie fehlen entsprechende Untersuchungen bislang. Vorliegende Arbeit soll diese Lcke schlieen. Ein quasi-experimentelles Setting hat im Vergleich zum experimentellen Setting sowohl Vor- als auch Nachteile. Felduntersuchungen, bei denen die Schler in ihrer gewohnten Umgebung untersucht werden, zeichnen sich durch ein hohes Ma an externer Validitt aus. Dies bedeutet, dass "die gefundenen Ergebnisse auf andere Personen, Situationen und Zeitpunkte generalisiert werden knnen", was bei experimentell gewonnenen Ergebnissen nicht immer der Fall ist (Bortz & Dring, 2002, S. 57ff.). Auf der anderen Seite zeichnen sie sich im Vergleich zu experimentellen Studien jedoch durch eine deutlich geringere interne Validitt aus, was bedeutet, dass „Vernderungen in den untersuchten abhngigen Variablen nicht eindeutig auf den Einfluss der unabhngigen Variablen zurckzufhren sind“ (Bortz & Dring, 2002, S. 56f.). Das liegt daran, dass im Unterrichtsprozess eine Reihe weiterer Variablen das Ergebnis beeinflussen knnen, die oft nicht vollstndig kontrolliert werden knnen. Der Identifizierung und Kontrolle mglicher Strvariablen kommen in quasi-experimentellen Untersuchungen eine groe Bedeutung zu. Dazu werden in der vorliegenden Studie verschiedene Verfahren, wie z.B. *Parallelisierung* oder die Entwicklung von *matched samples* eingesetzt (vgl. Kapitel 6). Das Modul wurde an der Schule des Untersuchungsleiters von diesen selbst unterrichtet und wissenschaftlich ausgewertet. Aus der resultierenden Nhe des Forschers zum Untersuchungsgegenstand ergeben sich Vor- aber auch Nachteile. Zum einen kann er die Durchfhrung der Untersuchung genau verfolgen, weitere Strvariablen erkennen und zustzliche Informationen etwa durch unmittelbare Rckfragen an die Schler zeitnah gewinnen. Zum anderen luft er aber Gefahr, durch persnliche Vorannahmen sowohl die Durchfhrung als auch die Auswertung der Intervention unbewusst zu beeinflussen. An dieser Stelle wird eine gewisse Nhe des hier gewhlten Ansatzes zur Aktionsforschung deutlich. Diese zeichnet sich nach Altrichter und Posch (2007, S. 13ff.) durch verschiedene Merkmale aus. Dazu gehren das „Aufgreifen von Fragestellungen aus der Praxis“, die „Forschung durch

Betroffene“ und das „In-Beziehung-Setzen von Aktion und Reflexion“. Das letzte Merkmal, die Möglichkeit gewonnene Erkenntnisse unmittelbar in praktisches Handeln umzusetzen, trifft für die vorliegende Studie jedoch nicht zu. Die auf wissenschaftlichen Verfahren beruhende Erhebung und Auswertung von Daten benötigt nämlich Zeit. Aus ihren Ergebnissen ergeben sich zwar Hinweise für die unterrichtliche Praxis, die Möglichkeit einer direkten Umsetzung bietet sich jedoch nicht. Altrichter und Posch (2007, S. 13ff.) aber sehen gerade in dem iterativen Prozess sich ständig wiederholender „Zyklen von Erkenntnisgewinnung und praktischem Handeln“ das Charakteristikum von Aktionsforschung²⁶. Methodologisch ist die vorliegende Untersuchung deshalb von der Aktionsforschung abzugrenzen. Im Zentrum steht hier die Überprüfung unterschiedlicher Hypothesen zur sachfachlichen, (fremd-) sprachlichen und motivationalen Wirksamkeit des bilingualen Chemieunterrichts. Damit handelt es sich in erster Linie um eine explanative Studie. An anderer Stelle geht es darum, erste Einblicke in das Forschungsfeld zu gewinnen, wie etwa im Verlauf der Pilotphase. Bevor die Daten der Hauptuntersuchung ausgewertet werden, werden die Stichproben ausführlich anhand verschiedener Parameter beschrieben. Somit enthält die Studie auch explorative und deskriptive Abschnitte (Bortz & Döring, 2002, S. 54f.). Die Studie verfolgt einen empirischen Ansatz, wobei die Ergebnisse des erteilten Unterrichts als Datengrundlage einer Analyse nach wissenschaftlichen Standards dienen. Die gewonnenen Daten sind dabei überwiegend quantitativer Natur und werden mit den Methoden der quantitativen Forschung erhoben und ausgewertet. Dabei handelt es sich vor allem um Zahlenwerte, welche über Ratingskalen (Fragebögen) oder die Vergabe von Punkten (Tests) gewonnen werden. Durch sie wird die Ausprägung eines bestimmten Merkmals beschrieben. Der quantitative Ansatz ermöglicht es, in der vorliegenden Untersuchung große Datenmengen zeitökonomisch und objektiv auszuwerten. Daraus ergibt sich die Möglichkeit, anhand statistischer Verfahren zuverlässige und generalisierbare Aussagen über eine vorab definierte Grundgesamtheit zu treffen. Auf der anderen Seite sind quantitative Erhebungsverfahren immer auch mit Einschränkungen bei der Erfassung komplexer Phänomene verbunden. Das liegt daran, dass eine Quantifizierung von Daten Kategorisierungsprozesse voraussetzt und nur solche Daten erfasst werden, welche sich durch diese Kategorien abbilden lassen. So können beispielsweise bei der Suche nach Gründen für eine

26 Lediglich in der Pilotphase wurde von der Möglichkeit persönliche Rückfragen an die Untersuchungsteilnehmer zu stellen und die Ergebnisse unmittelbar in die Gestaltung von Arbeitsmaterialien oder Erhebungsinstrumente einfließen zu lassen, Gebrauch gemacht. Dies führte dazu, dass das Design der Untersuchung in der Pilotphase mehrfach überarbeitet werden konnte (vgl. Kapitel 6). Der Pilotphase kommt deshalb auch ein explorativer Charakter zu.

bestimmte Einstellung zum bilingualen Unterricht Variablen wie fachliches Interesse und Selbstkonzept anhand von Rating-Skalen erhoben werden. Andere Variablen, wie etwa das soziale Ansehen der Unterrichtsform werden jedoch nicht erfasst und können deshalb nicht mit einer positiven Einstellung in Zusammenhang gebracht werden.

5.2.2 Forschungshypothesen und ihre Testung

Die vorliegende Untersuchung verfolgt, wie bereits dargestellt, in erster Linie einen explanativen Ansatz. Das dabei zum Einsatz kommende hypothesentestende Verfahren soll deshalb im Folgenden genauer erläutert werden. Bezugnehmend auf die drei Fragekomplexe wurden aufgrund der aktuellen Forschungsergebnisse zum bilingualen Unterricht unterschiedliche Hypothesen formuliert. Ihre Entwicklung wird ausführlich in den Unterkapiteln 5.3.1.1, 5.3.2.1 und 5.3.3.1 beschrieben. Bei der Formulierung wissenschaftlicher Hypothesen ist zu beachten, dass diese allgemein formulierte Aussagen darstellen, die eine logische Beziehung beinhalten und falsifizierbar sind. Solche Beziehungen können laut Bortz & Döring (2002, S. 493) Zusammenhänge oder Unterschiede zwischen zwei oder mehr Merkmalen betreffen. In der empirischen Forschung werden diese Hypothesen dann mit statistischen Hypothesentests (Signifikanztests) geprüft. Dazu werden sie anhand von Untersuchungsvariablen operationalisiert. Anschließend wird zu jeder operationalisierten Hypothese eine Gegenhypothese formuliert, die sogenannte Nullhypothese. Eine Hypothese, auch Alternativhypothese genannt, gilt dann als bestätigt, wenn es gelingt, die Nullhypothese zu widerlegen (Bortz & Döring, 2002, S. 496). Diese Überprüfung wird, wie oben erwähnt, mit Hilfe statistischer Verfahren durchgeführt. Die folgende Abbildung zeigt das Vorgehen am Beispiel von Hypothese 6.

Forschungshypothese	
<u>6. Hypothese</u> Fremdsprachlich orientierte Lerner sind im bilingualen Chemieunterricht stärker motiviert als im regulären Chemieunterricht.	Die Hypothese beinhaltet einen Unterschied zwischen zwei Gruppen von Schülern, nämlich den fso-Schülern, die am bilingualen Unterricht teilnehmen, und denen, die am Regelunterricht teilnehmen. Sie besagt, dass erstere motivierter sind als letztere (Unterschiedshypothese).
Operationalisierung	
<u>6.1 Alternativhypothese</u> Die Mittelwerte für die motivationale Disposition der fso-Schüler sind in der Gruppe der bilingualen Schüler größer als die Mittelwerte für die motivationale Disposition der fso-Schüler in der Gruppe der Regelschüler.	Die Hypothese wird nun so umformuliert, dass eine logische Beziehung zwischen der unabhängigen Variable <i>Gruppe</i> und der abhängigen Variable <i>motivationale Disposition</i> aufgestellt wird.
<u>6.2 Nullhypothese</u> Die Mittelwerte für die motivationale Disposition der fso-Schüler sind in der Gruppe der bilingualen Schüler kleiner als bzw. genauso groß wie die Mittelwerte für die motivationale Disposition der fso-Schüler in der Gruppe der Regelschüler.	Die Nullhypothese wird als Falsifizierung der Alternativhypothese formuliert. Gelingt es, die Nullhypothese mithilfe eines entsprechenden Signifikanztests (z.B. t-Test) zu widerlegen, so kann die Alternativhypothese beibehalten werden und gilt als bestätigt.

Tabelle 6: Bildung von operationalisierter Hypothese und Nullhypothese

Bei der im Beispiel formulierten Alternativhypothese handelt es sich um eine *gerichtete* Hypothese, weil angegeben ist, in welche Richtung eine Motivationsänderung erfolgt, nämlich im Sinne einer Steigerung. Außerdem wird keine Aussage über das Ausmaß der Veränderung gemacht, weshalb man auch von einer *unspezifischen* Hypothese spricht (Bortz & Döring, 2002, S. 492ff.). Diese Merkmale müssen bei statistischen Untersuchungen berücksichtigt werden. Die Frage der Gerichtetheit einer Hypothese hat nämlich Einfluss auf das weitere Vorgehen bei der statistischen Testung. So wird beispielsweise eine gerichtete Hypothese einer einseitigen Testung unterzogen. Damit wird der Bereich festgelegt, in welchem der zur Überprüfung von Hypothese 6 verwendete t-Test auf einem bestimmten Signifikanzniveau signifikant wird.

5.2.3 Datenerfassung und Verarbeitung

Die Erhebung, der für die Untersuchung relevanten Variablen erfolgt durch Tests und Fragebögen. Der sachfachliche und der sprachliche Lernzuwachs werden durch Tests im Pre-, Post-, Follow-Up-Design erfasst. Die Daten zu Motivation und Arbeitsverhalten werden über Schülerfragebögen vor und nach dem Modul (Pre- und Postfragebögen) ermittelt. Die Fragebögen dienen auch zur Erfassung weiterer Kontrollvariablen wie Alter, Geschlecht, Migrationshintergrund und

Bildungsniveau des Elternhauses. Außerdem kommen Test zur allgemeinen Leistungsfähigkeit in Chemie und Englisch, zu kognitiven Fähigkeiten und zu Lesefähigkeiten zum Einsatz. Auch diese Variablen werden im Rahmen der Untersuchung kontrolliert.

	Erhebungs- instrument	Abkür- zung	Gegenstand der Erhebung	Quelle
1	Chemie- Leistungstest zum Modul	CLT	Leistungen im Bereich Umgang mit Fachwissen bezogen auf das Modul unter Berücksichtigung unterschiedlicher Anforderungsbereiche (Niveau-Test)	In Anlehnung an die im Rahmen des ESNaS-Projekts entwickelten Vorgaben zur Aufgabenkonstruktion für das Fach Chemie, Köller et al. (2008b)
2	Test der allgemeinen Chemie- Leistungsfähigkeit	ACLT	Leistungen im Bereich Umgang mit Fachwissen bezogen auf Inhalte aus Klasse 7 und 8.1 unter Berücksichtigung unterschiedlicher Anforderungsbereiche (Niveau-Test)	In Anlehnung an die im Rahmen des ESNaS-Projekts entwickelten Vorgaben zur Aufgabenkonstruktion für das Fach Chemie, Köller et al (2008b)
3	Test zum deutschen Fachwortschatz des Moduls	DFWT	Deutscher Fachwortschatz, der in der Unterrichtsreihe neu eingeführt wird	Selbsterstellt als diskreter, selektiver und kontextunabhängiger Wortschatztest nach Read (2011)
4	Test zum englischen Fachwortschatz des Moduls	EFWT	Englische Fachbegriffe, deren deutsche Entsprechungen in der Unterrichtsreihe neu eingeführt werden	Selbsterstellt als diskreter, selektiver und kontextunabhängiger Wortschatztest nach Read (2011)
5	Test zum zusätzlich erworbenen englischen Fachwortschatz des Moduls	ZEFWT	Englische Fachbegriffe, deren deutsche Entsprechungen bereits vor Beginn der Unterrichtsreihe eingeführt wurden	Selbsterstellt als diskreter, selektiver und kontextunabhängiger Wortschatztest nach Read (2011)
6	Test zum zusätzlich erworbenen englischen	ZEAWT	Englischer Allgemeinwortschatz der Unterrichtsreihe	Selbsterstellt als diskreter, selektiver und kontextunabhängiger

	Allgemeinwortschatz des Moduls			Wortschatztest nach Read (2011)
7	Fragebogen zur Motivation und zu demografischen Angaben	FB (FB1 / 2)	Einschätzung der Schüler zu ihrer Motivation (Interesse, Kompetenzerleben, Nutzen, Druck) sowie demografische Angaben (Migrationshintergrund, akademischer Hintergrund, Alter, Geschlecht und fachliche Orientierung)	<u>Motivation</u> Basierend auf der Kurzsкала zur Erfassung der intrinsischen Motivation (KIM) von Hartmannsgruber (2014), Wilde et al. (2009) <u>Fachinteresse und akademisches Selbstkonzept</u> DESI, PISA, DENOCS <u>Demografische Angaben</u> Kunter et al. (2002) (adaptiert)
8	Kognitiver Fähigkeitstest	KFT	Kognitive Leistungen zur verbalen, non-verbalen und quantitativen Dimension der Intelligenz	Heller und Perleth (2000)
9	C-Test Englisch	CT	Allgemeine Leistungsfähigkeit im Englischen	KESS 7/8 (Bos et al. 2009)
10	Lesegeschwindigkeits- und -verständnistest	LGVT 5-12+	Leseleistungen im Deutschen in Bezug auf die Geschwindigkeit, das Verständnis und die Genauigkeit	Schneider, Schlagmüller und Ennemoser (2017)

Tabelle 7: Erhebungsinstrumente der Studie

Alle Datensätze wurden mit einem anonymisierten Code versehen, damit eine Zuordnung zu den einzelnen Testpersonen gelingt. Der Code besteht aus den beiden Anfangsbuchstaben des Vornamens des Vaters, den beiden Anfangsbuchstaben des Vornamens der Mutter, den beiden Ziffern des Geburtsmonats und den beiden letzten Ziffern des Geburtsjahrs. Außerdem war die Klassenbezeichnung durch die Ziffer der Jahrgangsstufe und den Buchstaben zur Kennzeichnung der Klasse anzugeben (vgl. Riese und Reinhold 2014, S. 260). Es gibt insgesamt vier Erhebungsphasen. Zu Beginn des Schuljahres werden Daten zu wichtigen Kontrollvariablen erhoben und ausgewertet. Aufgrund der Ergebnisse erfolgt die Zuteilung der Klassen zur Untersuchungs- bzw. zur Kontrollgruppe (Parallelisierung). Unmittelbar vor Modulbeginn werden die Pretests sowie die erste Fragebogenbefragung durchgeführt. Im Anschluss an das Modul erfolgen die Posttests sowie die zweite Fragebogenbefragung. Sechs Wochen später folgen die Follow-Up-Tests.

Design der Untersuchung: Forschungsfragen und Methodologie

Phase	Zeitpunkt	Test
1	Beginn Klasse 8	Test zur allgemeinen Chemie-Leistungsfähigkeit
		C-Test zur allgemeinen Leistungsfähigkeit im Englischen
		Kognitiver Fähigkeitstest
		Lesegeschwindigkeits- und -verständnistest
2	Unmittelbar vor Modulstart	Prefragebogen zur Motivation
		Pretest zum deutschen Fachwortschatz des Moduls
		Pretest zum englischen Fachwortschatz des Moduls
		Pretest zum zusätzlich erworbenen englischen Fachwortschatz des Moduls
		Pretest zum zusätzlich erworbenen englischen Allgemeinwortschatz des Moduls
		Chemieleistungstest zum Modul (Pretest)
Modulunterricht (ca. 5 Wochen)		
3	Unmittelbar im Anschluss an das Modul	Postfragebogen zur Motivation
		Posttest zum deutschen Fachwortschatz des Moduls
		Posttest zum englischen Fachwortschatz des Moduls
		Posttest zum zusätzlich erworbenen englischen Fachwortschatz des Moduls
		Posttest zum zusätzlich erworbenen englischen Allgemeinwortschatz des Moduls
		Chemieleistungstest zum Modul (Posttest)
4	Sechs Wochen später	Follow-Up-Test zum deutschen Fachwortschatz des Moduls
		Follow-Up-Test zum englischen Fachwortschatz des Moduls
		Follow-Up-Test zum zusätzlich erworbenen englischen Fachwortschatz des Moduls
		Follow-Up-Test zum zusätzlich erworbenen englischen Allgemeinwortschatz des Moduls
		Chemieleistungstest zum Modul (Follow-Up-Test)

Tabelle 8: Zeitlicher Verlauf der Datenerhebung

Die Chemietests werden stets im Anschluss an die Wortschatztests durchgeführt, da erstere Informationen zu verschiedenen Fachbegriffen enthalten, die Gegenstand der Fachwortschatztests

sind. Alle Tests liegen in zwei Formen (A und B) vor, die den Probanden im Wechsel vorgelegt werden, um Lerneffekte durch häufige Testung zu vermeiden. Das bedeutet, dass Probanden, die im Pretest Test A bearbeitet haben, im Posttest Test B vorgelegt bekommen und erst im Follow-Up-Test wieder Test A. Alle Tests und Fragebogenerhebungen werden im Klassenverband durchgeführt und in Papierform administriert. Die Zuordnung der Probanden zu den Gruppen A und B wird jeweils im Sitzplan dokumentiert, der Raum sowie die Sitzordnung bleiben für alle Erhebungen in einer Klasse gleich. Dadurch wird gewährleistet, dass Schüler die Testformen A und B im Wechsel erhalten und nicht etwa mehrfach hintereinander die gleiche Testform bearbeiten. Bei der Terminierung der Tests wird außerdem beachtet, dass diese nur in den ersten vier Unterrichtsstunden geschrieben werden. Damit sollen Konzentrations- und Motivationsverluste zum Ende des Schultages hin vermieden werden. Die so gewonnenen Daten werden in Excel-Tabellen überführt und mit der Statistiksoftware SPSS ausgewertet. Bei Mehrfachtestungen nach dem Pre-, Post-, Follow-Up-Design kann es vorkommen, dass Schüler bei einzelnen Tests fehlen und die entsprechenden Daten deshalb nicht vorliegen. Das Fehlen erfolgte in der Regel krankheitsbedingt oder aufgrund einer Befreiung zur Teilnahme an schulischen oder außerschulischen Aktivitäten. Das Fehlen von Daten geht stets mit einer Verringerung der Reliabilität einher und kann das Ergebnis verfälschen. Wenn Daten jedoch, wie in den beschriebenen Fällen, zufällig fehlen, so ist dieser Fehler als gering einzustufen, da sich daraus kein systematischer Einfluss auf das Ergebnis ergibt (Rumlich 2016, S. 261). In der vorliegenden Studie beträgt die Gesamtzahl der Untersuchungsobjekte der Hauptuntersuchung 149 (N). Die Untersuchungsgruppe enthält 74 (n (UG) = 74), die Kontrollgruppe 75 (n (KG) = 75) Probanden. Beide Gruppen sind also etwa gleich groß. Die Stichprobenumfänge reichen damit laut Bortz und Döring (2002, S. 613) für die drei im Rahmen der vorliegenden Untersuchung eingesetzten statistischen Analyseverfahren t-Test, Korrelationsberechnung und Varianzanalyse (ANOVA) aus, wenn mittlere Effekte auf dem Signifikanzniveau 0,05 abgesichert werden sollen²⁷.

27 Zur Absicherung eines in einer Population gültigen mittleren Effekts ($d = 0,8$) durch einen t-Test für unabhängige Stichproben benötigt man laut Bortz und Döring (2002, S. 613) auf dem 0,05-Signifikanzniveau pro Stichprobe $N = 50$ Untersuchungsteilnehmer. Bei abhängigen Stichproben werden weniger Untersuchungsobjekte benötigt. Für Korrelationsuntersuchungen ergibt sich bei einem mittleren Effekt ($r = 0,3$) eine Gruppengröße von $N = 64$, wenn diese Effekte mit $\alpha = 0,05$ abgesichert werden sollen. Bei einer zweifaktoriellen ANOVA mit Messwiederholung, also einem 2×3 Plan (Pre-, Post-, Follow-Up-Testplan mit Kontroll- und Untersuchungsgruppe) genügen $N = 52$ Probanden (bzgl. Faktor Zeit) bzw. $N = 64$ Probanden (bzgl. Faktor Gruppe), um auch einen mittleren Interaktionseffekt ($r = 0,5$) statistisch mit $\alpha = 0,05$ abzusichern.

5.2.4 Die Bildung von Interessensgruppen

Die Auswertung der Daten zur sachfachlichen, (fremd-)sprachlichen und motivationalen Wirksamkeit des bilingualen Moduls erfolgte auf unterschiedlichen Ebenen. Zum einen werden die Gruppen als Ganzes betrachtet und die Ergebnisse der jeweils drei Klassen zusammengefasst, sodass die Daten aus Untersuchungs- und Kontrollgruppe direkt miteinander verglichen werden können. Zum anderen werden auf der Grundlage der fachlichen Interessen der Schüler für beide Gruppen insgesamt vier Untergruppen gebildet. Dabei wird unterschieden, ob sich Schüler mehr für Fremdsprachen (fremdsprachlich orientierte [fso-] Schüler) oder mehr für Naturwissenschaften (naturwissenschaftlich orientierte [nwo-] Schüler) interessierten. Außerdem gibt es zwei weitere Untergruppen mit Schülern, die sich für beide Fachbereiche gleichermaßen stark (beidseitig interessierte [bso-] Schüler) bzw. gleichermaßen wenig (beidseitig wenig interessierte [bswo-] Schüler) interessierten. Bei der Auswertung werden schließlich die Ergebnisse der einzelnen Untergruppen miteinander verglichen. Insgesamt werden für die Hauptuntersuchung acht Untergruppen gebildet, vier Interessengruppen je Unterrichtssprache.

	Untersuchungsgruppe (N)	Kontrollgruppe (N)
Klasse	8a`,8b`8c`` <sup>28</sup> (74)	8e`,8d``,8e`` (75)
Orientierung	fso (23)	fso (18)
	nwo (32)	nwo (31)
	bso (16)	bso (19)
	bswo (1)	bswo (3)
	keine Angabe (2)	keine Angabe (4)

Tabelle 9: Einteilung der Probanden in Interessensgruppen²⁹

Bei der Zuteilung der Klassen zu den beiden Gruppen wurde darauf geachtet, dass die entstehenden Gruppen etwa gleichgroß und in Bezug auf das Geschlecht etwa gleich zusammengesetzt waren. Auf andere Merkmale konnte aus organisatorischen Gründen keine Rücksicht genommen werden.

²⁸ Die Hauptuntersuchung wurde in zwei Schuljahren durchgeführt. Die mit ` gekennzeichneten Klassenbezeichnungen beziehen sich auf das Schuljahr 2018/19 und die mit `` gekennzeichneten Klassenbezeichnungen auf das Jahr 2020/21.

²⁹ In der Tabelle ist die Zahl aller Probanden angegeben, die an der Untersuchung teilgenommen haben. Die jeweils für die Auswertung verwendete Anzahl von Datensätzen ist in Einzelfällen geringer.

Damit die Daten beider Gruppen dennoch vergleichbar sind, wurden verschiedene Kontrollvariablen erhoben (vgl. Abschnitt 5.4).

Die Einteilung in vier Interessensgruppen beruht auf der Selbsteinschätzung der Schüler. Dazu wurde ihnen auf dem Prefragebogen jeweils folgende Frage zum Ankreuzen vorgelegt:

Welche der vier Aussagen trifft am ehesten auf dich zu? Kreuze an! Ich mag Naturwissenschaften ☐ lieber als Sprachen. ☐ nicht so gerne wie Sprachen. ☐ genauso gerne wie Sprachen. ☐ genauso wenig wie Sprachen.
--

Abbildung 5: Fragebogenitem zur Ermittlung des Interessensschwerpunkts

Anhand der Ergebnisse wurden die Schüler anschließend der entsprechenden Gruppe zugeteilt. Da sich die vier Antwortalternativen gegenseitig ausschließen und vom Probanden eine Entscheidung für eine der vier Antworten verlangt wird, ergeben sich vier überschneidungsfreie Untergruppen. Lipski-Buchholz (2019, S. 76) geht bei ihrer Einteilung in Interessensgruppen ähnlich vor, unterscheidet jedoch darüber hinaus nach drei weiteren Interessen, nämlich für die Fächergruppen der künstlerisch-musischen Fächer, der Gesellschaftswissenschaften und Sport. Verriere (2014, S. 85) wählt einen anderen Weg. Sie lässt die Schüler einzelne Schulfächer in der Reihenfolge ihrer Beliebtheit auflisten und ermittelt anhand der Angaben, ob diese der Mathematik zu- / abgewandt bzw. dem Englischen zu- /abgewandt sind. Die sich ergebenden vier Untergruppen sind - wie sie feststellt - jedoch nicht überschneidungsfrei und so finden sich beispielsweise unter den mathematikzugewandten Schülern sowohl solche, die englischzugewandt sind als auch jene, die englischabgewandt sind (vgl. dazu Lipski-Buchholz, 2019, S. 212)³⁰.

5.3 Fokus der Untersuchung

Im folgenden Abschnitt werden die Hypothesen zu den drei Forschungsschwerpunkten der vorliegenden Untersuchung formuliert und die Instrumente vorgestellt, mit denen diese überprüft werden sowie die Verfahren, mit denen die Daten anschließend ausgewertet werden. Um die Genese

30 Abendroth-Timmer (2007, S. 144), die in ihrer Untersuchung zur Motivation im bilingualen Unterricht ebenfalls Interessen für Sachfächer berücksichtigt, die nicht zur mathematisch-naturwissenschaftlichen Fächergruppe gehören, unterscheidet drei Interessensgruppen, nämlich sprach-, sachfach- und neutralorientierte Schüler. Auch bei ihr erfolgt die Einteilung aufgrund der Selbsteinschätzung der Schüler und liefert überschneidungsfreie Gruppen.

der Forschungshypothesen zu verdeutlichen, erfolgt einleitend jeweils eine kurze Zusammenfassung des aktuellen Forschungsstands. Zunächst werden die Leistungen im Sachfach (Abschnitt 5.3.1), anschließend der (fremd-)sprachliche Lernerfolg (Abschnitt 5.3.2) und zuletzt der Bereich der Motivation (Abschnitt 5.3.3) in den Blick genommen.

5.3.1 Leistung im Sachfach

Für den bilingualen Unterricht in den Naturwissenschaften sind die Forschungsergebnisse in Bezug auf den sachfachlichen Lernerfolg im Modulunterricht teilweise widersprüchlich. Das liegt mitunter daran, dass die Bedingungen, unter denen sie erzielt wurden, sehr unterschiedlich sind.

Insbesondere bei älteren Schülern konnte gezeigt werden, dass sie im bilingualen Unterricht in Bezug auf den Wissenserwerb durchaus vergleichbare Leistungen erzielen können wie im Regelunterricht. Haagen-Schützenhöfer (2011) und Rodenhauser (2016) konnten für die Fächer Physik und Biologie keine statistisch signifikanten Unterschiede in Bezug auf das Wissen von bilingual und auf Deutsch unterrichteten Oberstufenschülern feststellen. Bei jüngeren Schülern mit geringeren Fremdsprachenkenntnissen sind die Forschungsergebnisse jedoch uneinheitlich.

Während Koch (2005) keine Unterschiede im konzeptuellen Wissen von Fünftklässlern feststellt, die an einer bilingualen Arbeitsgemeinschaft in den Naturwissenschaften teilgenommen hatten, finden Hartmansgruber (2014) und Piesche (2016) bei den von ihnen untersuchten Sechstklässlern statistisch signifikante Unterschiede zwischen dem Wissen bilingual und auf Deutsch unterrichteten Schülern. Für Schüler der Mittelstufe fehlen empirische Untersuchungen bislang. Diese Lücke soll mit Hilfe der vorliegenden Untersuchung geschlossen werden. Da sachfachliche Kompetenzen sich nicht auf das Abrufen von Wissen beschränken, sondern auch den Transfer und die Anwendung dieses Wissens bedeuten, soll in dieser Studie, anders als in den zitierten Arbeiten zum sachfachlichen Lernerfolg auch der Frage nachgegangen werden, inwieweit die Leistungen der bilingualen Schüler denen der Regelschüler im Hinblick auf höhere Denkprozesse vergleichbar sind. Insgesamt fehlt es laut Heine (2013) noch an Untersuchungen, die Hinweise auf den Kompetenzerwerb im Sachfach und nicht ausschließlich auf den Wissenszuwachs geben (vgl. Heine 2013, S. 218).

5.3.1.1 Hypothesen und ihre Testung

Vor dem Hintergrund der geschilderten Forschungsergebnisse zum bilingualen naturwissenschaftlichen Unterricht und der sich daraus ergebenden offenen Fragen in Bezug auf das Fach Chemie, werden für den Bereich des sachfachlichen Lernens drei Hypothesen formuliert. Diese sollen im Rahmen der Untersuchung überprüft werden.

Hypothesen

1. Die auf Deutsch unterrichteten Schüler sind den bilingual unterrichteten Schülern im Umgang mit dem vermittelten chemischen Fachwissen überlegen.
2. Die bilingual unterrichteten Schüler erzielen im kognitiv anspruchsvolleren Bereich (*higher order thinking skills*) bessere Leistungen als die auf Deutsch unterrichteten Schüler.
3. Hohe fremdsprachliche Kompetenzen führen zu besseren Leistungen im sachfachlichen Test. Bei Unterschreitung eines Mindestniveaus an fremdsprachlicher Kompetenz sind die sachfachlichen Leistungen im bilingualen Unterricht schlechter als im Regelunterricht. Überschreiten Schüler jedoch ein oberes Kompetenzniveau in der Fremdsprache, so sind die sachfachlichen Leistungen im bilingualen Unterricht besser als im Regelunterricht (*double threshold hypothesis*).

Abbildung 6: Hypothesen zum sachfachlichen Lernerfolg

Variablen

Zur Bestimmung der sachfachlichen Leistungen werden Tests durchgeführt, die den Umgang mit dem erworbenen Fachwissen gemäß der drei Anforderungsbereiche der Bildungsstandards erfassen. Je höher der Punktwert im Test, desto stärker ausgeprägt ist das Merkmal. Die Testergebnisse werden für jeden Schüler einzeln ermittelt. Außerdem werden Gruppenmittelwerte berechnet und miteinander verglichen. Zur Bestimmung des Lernzuwachses bzw. des Vergessens im Anschluss an das Modul wird für jeden Schüler die Differenz zwischen Post- und Pretestergebnis bzw. zwischen Post- und Follow-Up-Testergebnis ermittelt. Zusätzlich werden für jeden Schüler Punktwerte für die Testleistung in den einzelnen Anforderungsbereichen ermittelt. Die Werte für die Anforderungsbereiche II und III werden zusammengefasst und dienen als Maß dafür, inwieweit der Umgang mit chemischem Wissen in Bezug auf höhere Denkopoperationen gelingt. Auch hier werden aus den Einzelwerten anschließend Gruppenmittelwerte gebildet und miteinander verglichen. Fremdsprachliche Kompetenzen werden anhand von C-Tests erfasst und durch einen Gesamtpunktwert abgebildet (vgl. Abschnitt 5.4.3). Die Orientierung der Lerner wird aufgrund der Selbstauskunft im Fragebogen vorgenommen. Die Schüler werden anschließend einer von vier Interessensgruppen zugeordnet (vgl. Abschnitt 5.2.4).

Operationalisierte Hypothesen

1. Der Anstieg der Gesamtpunktzahl vom Pre- zum Posttest ist bei den auf Deutsch unterrichteten Schülern im Mittel größer als bei den bilingual unterrichteten Schülern.
2. Der Anstieg der Punktzahl in den Anforderungsbereichen II und III vom Pre- zum Posttest ist bei den bilingual unterrichteten Schülern im Mittel größer als bei den auf Deutsch unterrichteten Schülern.
3. Schüler, deren Ergebnis im C-Test einen bestimmten Minimalwert unterschreitet, erzielen im Regelunterricht größere sachfachliche Lernzuwächse als im bilingualen Unterricht. Schüler, deren Ergebnis im C-Test einen bestimmten Maximalwert überschreitet, erzielen im bilingualen Unterricht größere sachfachliche Lernzuwächse als im Regelunterricht (*double threshold hypothesis*).

Abbildung 7: Operationalisierte Hypothesen zum sachfachlichen Lernerfolg

5.3.1.2 Theoretischer Hintergrund und Aufbau der Tests

Zur Erfassung der Leistungen im Fach Chemie konnte nicht auf bereits vorliegende standardisierte Tests zurückgegriffen werden. Das lag dran, dass zum Thema *Alkali and alkaline earth metals* keine passenden Tests vorlagen. Deshalb wurden eigene Tests entwickelt und in einer Pilotphase evaluiert. Das methodische Vorgehen bei der Erstellung naturwissenschaftlicher Leistungstests orientiert sich an den Empfehlungen von Riese und Reinhold (2014, S. 257ff.). In den folgenden Abschnitten wird die Konstruktion der Tests und ihre Anbindung an die in den Kernlehrplänen formulierten Kompetenzerwartungen ausführlich beschrieben. Die Evaluation der Tests im Hinblick auf die Gütemerkmale Objektivität, Validität und Reliabilität wird in Abschnitt 6.2 dargestellt.

Bei den eingesetzten Chemietests sollen auch höhere kognitive Leistungen im Umgang mit dem erworbenen Fachwissen erfasst werden. In den nationalen Bildungsstandards für die Naturwissenschaften (KMK 2005) wird diesem Anspruch durch die Formulierung der drei Anforderungsbereiche (I, II und III) Rechnung getragen. Zur Konstruktion von Testaufgaben erwies sich dieses Modell jedoch in der Vergangenheit als ungeeignet, da der Wechsel des Anforderungsbereichs die Variation mehrerer Aufgabenmerkmale gleichzeitig erforderte (vgl. Kauertz et al., 2010). Zur Konstruktion von Aufgaben für die vorliegenden Tests wurde deshalb auf das zur Evaluation der Standards in den Naturwissenschaften entwickelte, fächerübergreifende Kompetenzstrukturmodell zurückgegriffen (Kauertz et al., 2010). Das *ESNaS-Kompetenzstrukturmodell* umfasst die drei Dimensionen *Kompetenzbereich*, *Komplexität des Inhalts* und *kognitive Prozesse* (siehe Abbildung unten).

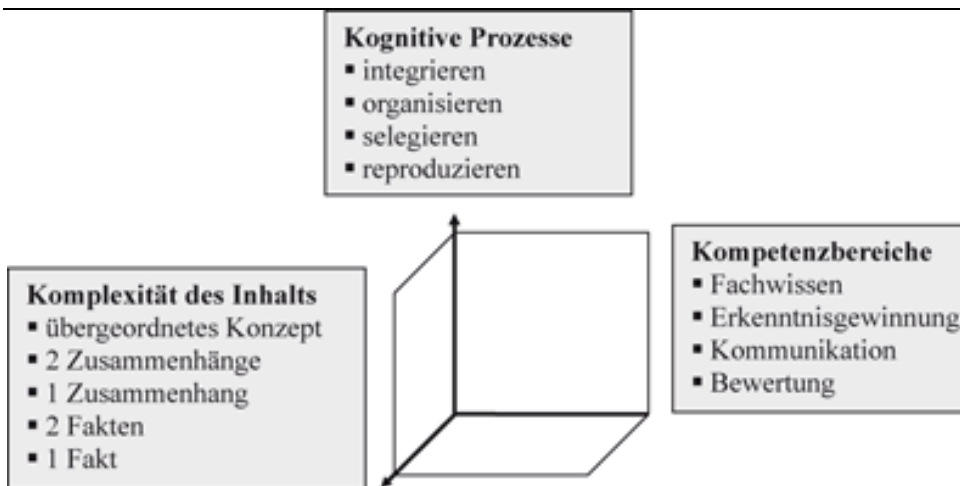


Abbildung 8: Das Kompetenzstrukturmodell des Projekts ESNaS

(entnommen aus Ferber, 2014, S. 19)

In dem Modell wird die Dimension des Kompetenzbereichs direkt aus den Bildungsstandards übernommen, während die Anforderungsbereiche durch die Dimensionen Komplexität des Inhalts und kognitive Prozesse abgebildet werden. Für die vorliegende Untersuchung wird dabei ausschließlich der Kompetenzbereich *Fachwissen* fokussiert und, da es bei den Tests weniger um die Reproduktion von deklarativem Wissen, sondern vielmehr um den Umgang mit diesem Wissen gehen soll, wird im Folgenden die Formulierung *Umgang mit Fachwissen* (UMFW) verwendet (Ropohl, 2010, S. 30). Chemisches Fachwissen wird in den Bildungsstandards anhand von drei *Basiskonzepten* (*Chemische Reaktion*, *Struktur der Materie* und *Energie*) strukturiert, durch welche sich fachspezifische Gemeinsamkeiten unterschiedlicher chemischer Sachverhalte abbilden lassen. Sie erlauben darüber hinaus die Vernetzung fachlicher Inhalte und den Aufbau strukturierten Wissens. Zur Beschreibung der Kompetenzstufen im ESNaS-Kompetenzstrukturmodell wird das Konzept Energie jedoch aufgrund der inhaltlichen Nähe zum Konzept der chemischen Reaktion diesem zugeordnet (Ropohl 2010, S. 29). Darüber hinaus beinhaltet dieses Modell zusätzlich die Bereiche *Phänomene erklären*, *Bezüge finden* und *Abstraktionen nutzen*. In Anlehnung an das ESNaS-Kompetenzstrukturmodell lassen sich die Testaufgaben somit den folgenden drei Inhaltsfeldern zuordnen:

1. Struktur und Eigenschaften / Stoffe und Teilchen (SE)
2. Chemische Reaktion und Energieumsätze (CR)
3. Phänomene erklären, Bezüge finden und Abstraktionen nutzen (PBA)

Die Achse Komplexität des Inhalts dient dazu, die Vernetzung chemischer Wissensinhalte abzubilden. Im ESNaS-Modell wird dazu zwischen den fünf Komponenten *1 Fakt, 2 Fakten, 1 Zusammenhang, 2. Zusammenhänge* und *übergeordnetes Konzept* unterschieden, die den Schwierigkeitsgrad einer Testaufgabe beeinflussen. Dabei meinen Fakten Bezeichnungen, (Fach-)Begriffe, Eigenschaften, Symbole, Tatsachen, Zustände, Elemente, Aspekte oder Objekte, zwischen denen kein expliziter Zusammenhang besteht. Zusammenhänge bezeichnen Beziehungen in Form einer expliziten Abhängigkeit, einer gegenseitigen Bedingung, einer Relation, einer Form-Funktion-Beziehung einer Korrelation, einer erklärten Kausalität oder einer Wechselwirkung. Unter Konzepten werden Prinzipien, Gesetze oder Regelhaftigkeiten verstanden, die situationsunabhängig formuliert und zur Beschreibung von Zusammenhängen verwendet werden können (Köller O. et al., 2008b).³¹ Einem Vorschlag Ferbers (2014, S. 32) folgend, wird die Dimension Komplexität des Inhalts auch für die vorliegende Untersuchung auf die drei Niveaus Fakt (F), Zusammenhang (Z) und übergeordnetes Prinzip (P) reduziert³². Das ESNaS-Modell unterscheidet außerdem vier kognitive Prozesse, die den Umgang mit Wissen beschreiben und den Schwierigkeitsgrad einer Aufgabe bestimmen: *Reproduzieren, selektieren, organisieren* und *integrieren*. Die Kategorien wurden der Informationsverarbeitungstheorie entlehnt und betreffen nach Kauertz et al. (2010, S. 144) das Identifizieren von Informationen (reproduzieren), das Auswählen von Informationen (selektieren), das Strukturieren von Informationen (organisieren) und das Einbinden von Informationen in die Wissensbasis (integrieren)³³. Anders als in den Aufgaben zur Evaluation der Bildungsstandards, bei denen diese Informationen jeweils im Aufgabenstamm der einzelnen Testaufgaben vorgegeben sind, werden Hintergrundinformationen in den vorliegenden Tests nur im zweiten Teil gegeben. Im ersten Teil soll überprüft werden, inwieweit im Unterricht vermittelte, fachliche Informationen tatsächlich memoriert werden³⁴.

31 Eine ausführliche Beschreibung der fünf Komponenten findet sich in Köller et al. (2008b).

32 Ferber (2014) weist darauf hin, dass die statistische Trennung zwischen „einem Fakt“ und „zwei Fakten“ bzw. „einem Zusammenhang“ und „zwei Zusammenhängen“ durch Ropohl (2010) nicht gelingt.

33 Die Prozesse ähneln den Kategorien von Anderson und Krathwohl (2001) zur Analyse von Lernzielen (erinnern, verstehen, anwenden, analysieren, bewerten [evaluieren] und erzeugen).

34 In den Aufgaben zur Evaluation der nationalen Bildungsstandards war dies nicht möglich, da nicht klar war, welche Unterrichtsinhalte an den jeweiligen Schulen behandelt wurden. Es ergaben sich insgesamt nur kleine Schnittmengen der in den einzelnen Lehrplänen der Bundesländer vorgegebenen Fachinhalte (vgl. Ropohl, 2010, S. 30).

Zur Auswertung der Test wurden die beiden Dimensionen Komplexität des Inhalts und kognitiver Prozess wieder den in den Bildungsstandards formulierten Anforderungsbereichen zugeordnet. Die Abbildung zeigt die entsprechende Zuordnung nach Köller et al. (2008a).

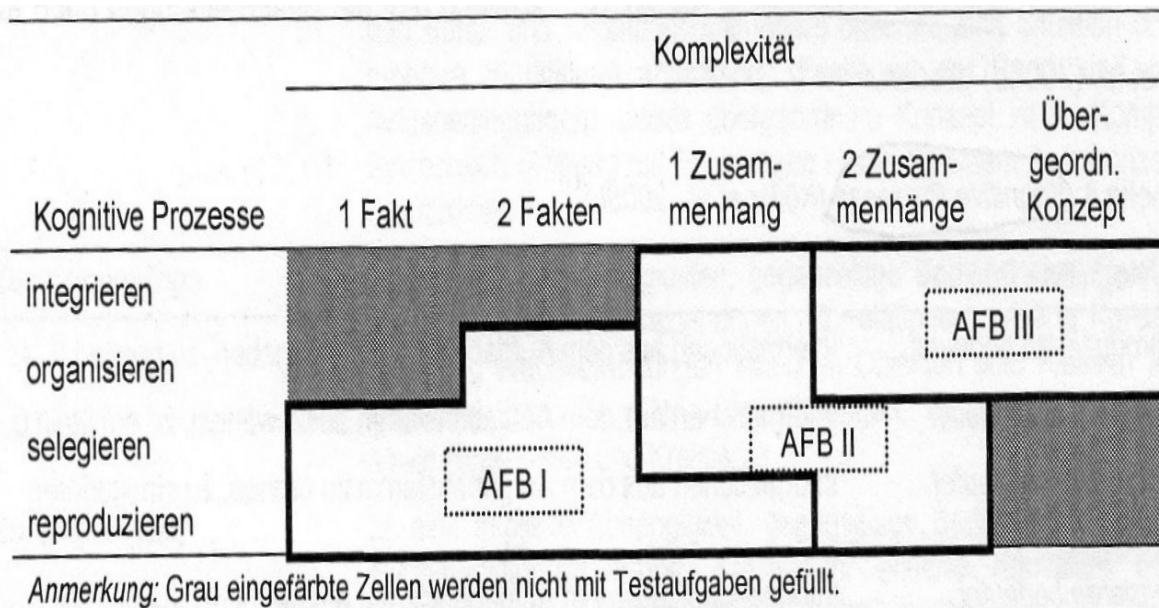


Abbildung 9: Anforderungsbereiche im Kompetenzmodell ESNaS

(entnommen aus Ropohl, 2010, S. 32)

Es wird davon ausgegangen, dass den beiden Dimensionen Komplexität des Inhalts und kognitiver Prozess jeweils schwierigkeiterzeugende Merkmale zur Konstruktion von Testaufgaben entnommen werden können³⁵. Darüber hinaus zeigen viele Untersuchungen, dass auch formale Merkmale einen Einfluss auf die Schwierigkeit von Testaufgaben haben. Dazu gehört etwa das gewählte Aufgabenformat. Im Rahmen der TIMSS-Studie konnte gezeigt werden, dass offene Aufgaben im Mittel signifikant schwieriger sind als geschlossene Aufgaben (vgl. Ropohl, 2010, S. 33). Für das Fach Chemie erweist sich auch die Repräsentationsebene chemischer Inhalte als schwierigkeitsbestimmendes Merkmal. Verschiedene Untersuchungen zeigen, dass es Schülern

³⁵ Ropohl (2010) gelang es im Rahmen des Projekts zur Evaluation der nationalen Bildungsstandards für die Naturwissenschaften in der Sekundarstufe I (ESNaS) die Dimension Komplexität des Inhalts mit Methoden der Probabilistischen Testtheorie erfolgreich als schwierigkeiterzeugende Aufgabenmerkmale zu validieren. Für die Dimension kognitive Prozesse gelang dies nur eingeschränkt, wobei jedoch für Aufgaben mit Vorgabe chemischer Fachinformationen "theoriekonform leichte und schwieriger Prozesse unterschieden werden konnten" (Ropohl 2010, S. 117).

schwerfällt, chemische Sachverhalte auf Teilchenebene zu erklären (Pfeifer, Häuseler, & Lutz, 1992). Ropohl (2010, S. 108) konnte in diesem Zusammenhang zeigen, dass Aufgaben auf Teilchenebene signifikant schwieriger sind als Aufgaben auf Phänomenebene. Zu letzteren zählen auch Aufgaben, die Formel-, Modellwissen oder Mathematisierung verlangen. Da die eingesetzten Tests die kognitiven Anforderungen im Fach Chemie möglichst breit abbilden sollen, werden deshalb folgende Merkmale bei der Konstruktion der Testaufgaben berücksichtigt:

1. Teilbereich (Basiskonzept) (SE, CR, PBA)
2. Komplexität des Inhalts (F, Z, ÜK)
3. Kognitiver Prozess (RE, SE, OR, IN)
4. Anforderungsbereich (AFBI, AFBII, AFBIII)
5. Aufgabenformat (offene, halb-offene und geschlossene) (OA, HOA, GO)
6. Repräsentationsebene (Teilchen- oder Phänomenebene) (TE, PE)

Abbildung 10: Merkmale der Testaufgaben zum Kompetenzbereich UMFW

Der Test besteht aus zwei Teilen und umfasst insgesamt 42 Aufgaben. Damit kann er von Schülern in 33 Minuten bearbeitet werden - ein Zeitumfang, der Schülern aus schriftlichen Leistungsüberprüfungen in den Nebenfächern vertraut ist. Auch die anteilige Berücksichtigung der Anforderungsbereiche kennen die Schüler bereits aus Klassenarbeiten in den sogenannten Hauptfächern. Der erste Teil (Aufgaben 1 – 14, acht Minuten) stellt einen klassischen Wissenstest dar. Hier geht es um die Wiedergabe fachlicher Inhalte aus dem Unterricht. Die Aufgaben zielen überwiegend auf kognitive Prozesse des Reproduzierens und sind ausnahmslos dem Anforderungsbereich I zuzuordnen. Im zweiten Teil (Aufgaben 15 – 42, 25 Minuten) werden kognitiv anspruchsvollere Leistungen (*higher order thinking skills*) verlangt. Die Aufgaben sind den Anforderungsbereichen II und III zuzuordnen. Nach Lienert und Raatz (1994, S. 15) handelt es sich um einen mehrdimensionalen Leistungstest, bei dem unterschiedliche kognitive Leistungen im Umgang mit chemischem Fachwissen getestet werden. Die Aufgaben sind nach aufsteigender Schwierigkeit sortiert. Da ausreichend Zeit zur Bearbeitung zur Verfügung steht, ist nicht die Schnelligkeit, sondern das kognitive Potential entscheidend dafür, ob auch die schwierigeren Aufgaben gelöst werden können. Es handelt sich also um einen *Niveau-* bzw. *Power-Test* (vgl. Lienert und Raatz 199, S. 15). Um sicherzustellen, dass das Nichtlösen der Aufgabe nicht auf das Fehlen des dazu nötigen Fachwissens zurückzuführen ist, wird dieses zu Beginn des zweiten Testteils in einem kurzen Informationstext bereitgestellt. Sämtliche für die Lösung der Aufgaben 15 – 42 relevanten Fachinformationen sind darin enthalten und stehen den Schülern während der

Bearbeitung zur Verfügung. Ein ähnliches Vorgehen wählt Ropohl (2010), der das zur Lösung kognitiv anspruchsvoller Aufgaben nötige Fachwissen jedoch im Aufgabenstamm der jeweiligen Aufgabe vorgibt. Damit erhöht sich der Textumfang der Aufgaben deutlich und, wie schon bei PISA, auch der Einfluss der Lesekompetenz³⁶. Für die vorliegenden Tests, die sich auf einen vergleichsweise kleinen Inhaltsbereich beziehen, werden diese Informationen gesammelt in einem vorangestellten Informationstext gegeben. So soll Redundanz vermieden und der Textumfang möglichst geringgehalten werden. Um den Einfluss der Lesekompetenz zusätzlich zu reduzieren, wird der Text mit Hintergrundinformationen zu Beginn des zweiten Prüfungsteils vom Testleiter vorgetragen. Sämtliche Tests wurden auf Deutsch verfasst. Damit soll zum einen eine Verständlichkeit der Aufgabenstellung sowie eine Vergleichbarkeit der Ergebnisse gewährleistet sein. Zum anderen wird so der Forderung Rechnung getragen, dass das im Rahmen des bilingualen Unterrichts erworbene Wissen auch in der Muttersprache zur Verfügung steht. Dieses Vorgehen findet sich in vielen Untersuchungen zum sachfachlichen Lernerfolg in den Naturwissenschaften (vgl. Piesche, 2016; Hartmannsgruber, 2014 oder Haagen-Schützenhöfer, 2017).

5.3.1.3 Konstruktion von Testaufgaben

Die Aufgabenkonstruktion erfolgte anhand der oben beschriebenen a priori festgelegten Merkmale und im Hinblick auf die konkreten Lernziele der Unterrichtsreihe (vgl. Abschnitt 5.5.1). Zur Unterscheidung der verschiedenen Niveaustufen des kognitiven Umgangs mit dem erworbenen Fachwissen werden die Tests später in Bezug auf die Zahl, der im jeweiligen Anforderungsbereich gelösten Aufgaben, untersucht. Die Konstruktion von Aufgaben zu den drei Anforderungsbereichen erfolgt im Rückgriff auf das ESNaS-Kompetenzstrukturmodell und orientiert sich an den beiden Dimensionen Komplexität des Inhalts und kognitiver Prozess (vgl. Abbildung auf Seite 94). Um die beiden Merkmale richtig zuzuordnen zu können, muss die Lösung der Aufgaben vorab formuliert und in einzelne Lösungsschritte zerlegt werden. Die bei der schrittweisen Lösung auftretende höchste Merkmalsausprägung bestimmt dabei die Einstufung der Aufgabe auf der jeweiligen Dimension des Kompetenzmodells. Bei der Entwicklung der Testaufgaben wurde auf die ausführliche Konstruktionsanleitung zurückgegriffen, die im Rahmen des ESNaS-Projektes von Köller et al. (2008b) erstellt wurde. So wurden insgesamt 42 Aufgaben zum Inhaltsbereich Alkali-

36 In den PISA-Studien aus den Jahren 2000 und 2003 konnte jeweils eine hohe Korrelation zwischen der Lesekompetenz und den naturwissenschaftlichen Kompetenzen festgestellt werden (Baumert et al., 2001; Leutner et al., 2004 zitiert nach Ropohl (2010, S. 35).

und Erdalkalimetalle entwickelt. Die Differenzierung nach Teilbereich (Basiskonzept) und Repräsentationsebene ergab sich in der Regel aus den vorab formulierten Lernzielen. Es wurden sowohl geschlossene als auch halb-offene und offene Aufgaben formuliert. Das Vorgehen soll anhand des Beispiels einer geschlossenen Aufgabe erläutert werden. Geschlossene Aufgaben kommen in den vorliegenden Tests am häufigsten zum Einsatz. Die Vorteile solcher gebundenen Aufgaben, die häufig als *Multiple-Choice-Aufgaben* konzipiert sind (Lienert und Raatz, 1994, S. 19), liegen insbesondere in ihrer zeitökonomischen und vergleichsweise objektiven Auswertbarkeit (Lienert und Raatz, 1994, S. 8). Dabei hat sich die Vorgabe von vier Antwortmöglichkeiten mit jeweils einer richtigen Antwortalternative (*Multiple-Choice-Single-Select*) auch in vielen internationalen Vergleichsstudien durchgesetzt (Ropohl, 2010, S. 46). Die jeweiligen Items bestehen aus der Fragestellung bzw. dem Arbeitsauftrag sowie den anzukreuzenden Antwortmöglichkeiten. Eine zentrale Schwierigkeit bei der Formulierung von MC-Aufgaben ist die Formulierung der Distraktoren, die Probanden, die nicht über die entsprechende Kompetenz bzw. das entsprechende Wissen verfügen, genauso plausibel erscheinen müssen, wie der Attraktor. Köller et al. (2008b) empfehlen Distraktoren, die im Gegensatz zur richtigen Antwort stehen, jedoch über möglichst ähnliche Eigenschaften wie diese verfügen bzw. dem gleichen Kontext entstammen. Auch sprachlich sollen sich Distraktoren möglichst nicht von der richtigen Antwort unterscheiden, etwa den gleichen Textumfang haben und grammatikalisch ähnlich aufgebaut sein. Weitere formale und inhaltliche Aspekte, die bei der Formulierung der Testaufgaben eine Rolle gespielt haben, waren sprachliche Einfachheit, Prägnanz oder inhaltliche Strukturierung. Ein Beispiel für eine geschlossene Aufgabe zeigt die Abbildung unten.

Text	Wie wird Natrium aufbewahrt?
Antwortoptionen	A unter Paraffinöl B in einer Ampulle C unter Spiritus D unter Schutzgas
Codierung	RE – F - SE – GA – PE - AFB I

Abbildung 11: Beispiel für eine geschlossene Aufgabe aus dem Chemiesleistungstest

Es handelt sich um eine Aufgabe, welche die *Reproduktion (R)* von im Unterricht erlerntem Wissen verlangt. Dazu muss ein *Fakt (F)* erinnert werden, nämlich die Art der Aufbewahrung von Natrium unter Paraffinöl. Die Besonderheit der Lagerung von Alkalimetallen stellt ein chemisches Phänomen dar. Laut Kauertz et al. (2010) zeichnet sich die erste Kompetenzstufe des Kompetenzbereiches Umgang mit Fachwissen im Bereich *Struktur und Eigenschaften / Stoffe und Teilchen (SE)* durch das Identifizieren von Fakten und Phänomenen aus.

Es handelt sich um eine *Multiple-Choice-Aufgabe* mit vier Antwortalternativen, also um eine *geschlossene Aufgabe (GA)*. Eine Antwort ist auf der *phänomenologischen Ebene (PE)* möglich und erfordert nicht die zu Hilfenahme eines Teilchenmodells. Sie ist dem *Anforderungsbereich I (AFB I)* zuzuordnen. Dieser beinhaltet laut Bildungsstandard, dass Schüler „Kenntnisse und Konzepte zielgerichtet wiedergeben“ (KMK 2005). Neben geschlossenen Aufgaben kommen im Test auch offene Formate zum Einsatz. Diese sind für Probanden signifikant schwieriger als geschlossene Aufgaben, wie Ropohl (2010) zeigen konnte. Die zusätzliche kognitive Herausforderung besteht für die Probanden darin, Bedeutungszusammenhänge selbst zu konstruieren und aktiv hervorbringen zu müssen (vgl. Hammann und Jördens, 2014, S. 170). Auf der anderen Seite bieten offene Aufgaben nach Hammann und Jörgens (2014, S. 170) den Vorteil "unterschiedliche Verständnisstufen differenziert zu beschreiben". Die Vielfalt der Antwortmöglichkeiten erschwert jedoch die Auswertung offener Aufgaben, worunter häufig auch die Objektivität des Tests leidet. Offene Formate kommen aber auch in großangelegten empirischen Untersuchungen zum Einsatz. Der Naturwissenschaftstest der PISA-Studie 2006 enthielt zu gleichen Anteilen geschlossene, halb-offene und offene Aufgaben (OECD, 2006). Um sicherzustellen, dass das Nichtlösen einer Aufgabe nicht auf sprachliche Probleme zurückzuführen ist, haben die bilingual unterrichteten Schüler auch die Möglichkeit, einzelne englische Formulierungen zu verwenden, ohne dass dies zu Punktabzügen führt. Eine Übersicht über die Vor- und Nachteile offener und geschlossener Aufgaben findet sich bei Lienert und Raatz (1994, S. 26ff.). Offene Aufgaben kommen im vorliegenden Test insbesondere für Aufgaben zum Einsatz, deren Lösung höhere Denkopoperationen verlangt. Es handelt sich um Kurzaufsatz-Antworten, deren Länge zwei bis fünf Sätze beträgt³⁷. Die Auswertung erfolgt anhand eines Codierleitfadens, mit dessen Hilfe die Antworten der Probanden bestimmten Kategorien zugeordnet werden können. Dieser wurde durch einen Experten auf Grundlage fachlicher Überlegungen erstellt und anschließend in einem Expertenteam, bestehend aus zwei weiteren Experten, diskutiert sowie mit den Schülerantworten im Rahmen der Pilotierung verglichen. Dabei kam es vor, dass neue Lösungsmuster auftraten, welche durch das deduktiv entwickelte Kategoriensystem nicht angemessen abgebildet werden konnten. In diesen Fällen wurde der Codierleitfaden durch induktiv, also durch aus den Schülerantworten gewonnenen Kategorien, erweitert (vgl. dazu Abschnitt 6.2). Laut Mayring (2000) enthält ein Codierleitfaden vier Komponenten: *Bezeichnung der Kategorie*, *Definition der Kategorie*, *Codierregel* und *Beispiele*. Außerdem wird angegeben, wie viele Punkte in der jeweiligen Kategorie vergeben werden (*Code*). Die Abbildungen unten zeigen ein Beispiel einer offenen Aufgabe inklusive Codierleitfaden.

37 Lienert und Raatz (1994, S. 21) empfehlen eine maximale Wortanzahl von 150 Wörtern.

Text	In der Nähe des Freiburger Hauptbahnhofs geriet vor einigen Jahren ein Güterwaggon in Brand, der etwa eine halbe Tonne Natrium (Na, <i>sodium</i>) transportierte. Die Feuerwehr löschte unter Missachtung der Gefahrgutkennzeichnung mit Wasser, wobei eine schwere Explosion auftrat, die einen Feuerwehrmann tötete und eine Brücke beschädigte. <u>Erkläre</u> , wie es zu dem Unfall kommen konnte (mögliche Reaktionen bitte erläutern).
Codierung	IN- ÜK - PBA - OA - PE- AFB III

Abbildung 12: Beispiel für eine offene Aufgabe aus dem Chemieleistungstest

Zu Erklärung des Unfalls müssen die Schüler erkennen, dass das Natrium mit dem Löschwasser reagiert hat (Kategorie 1), dass diese Reaktion stark exotherm war (Kategorie 2) und dass es deshalb zur Explosion des entweichenden Wasserstoffs kam (Kategorie 3). Insgesamt können deshalb bei der Aufgabe drei Punkte erzielt werden, wobei jeweils unterschiedliche Formulierungsalternativen akzeptiert werden (Beispiele).

1. <u>Kategorie</u> <u>Reaktion von Natrium mit Wasser</u>	
Code	0
Definition	Hinweis auf die Reaktion von Natrium mit dem Löschwasser fehlt.
Codierregel	Es wird nicht beschrieben, dass das Natrium mit dem Wasser reagiert.
Code	1
Definition	Hinweis auf die Reaktion von Natrium mit Löschwasser vorhanden.
Codierregel und Beispiele	Es wird beschrieben, dass das Natrium mit dem Wasser reagiert. Auch alternative Formulierungen, die den Stoffumsatz beschreiben, werden akzeptiert, z.B. „Natrium verbindet sich mit dem Wasser / H_2O “, "es bilden sich Natriumhydroxid NaOH und Wasserstoff / H_2 "
2. <u>Kategorie</u> <u>Stark exothermer Reaktionsverlauf</u>	
Code	0
Definition	Hinweis auf den stark exothermen Charakter der Reaktion fehlt.
Codierregel	Es wird nicht beschrieben, dass Natrium unter Freisetzung großer Mengen Energie reagiert.
Code	1
Definition	Hinweis auf den stark exothermen Charakter der Reaktion vorhanden.
Codierregel und Beispiele	Es wird beschrieben, dass Natrium stark exotherm mit Wasser reagiert. Statt exotherm werden Formulierungen, wie "Natrium ist unedel / sehr reaktiv, reagiert sehr heftig" akzeptiert. Der Hinweis auf die großen Mengen an Natrium und Wasser ist nicht Voraussetzung für die Vergabe des Punktes. Es reicht der Hinweis auf die große Reaktivität von Natrium mit Wasser.
3. <u>Kategorie</u> <u>Entzündung des Wasserstoffs</u>	
Code	0
Definition	Hinweis auf die Bildung des an der Luft explosiven Wasserstoffs fehlt.
Codierregel	Es wird nicht beschrieben, dass sich an der Luft ein explosives Gemisch mit Wasserstoff bildet.
Code	1
Definition	Hinweis auf die Bildung des an der Luft explosiven Wasserstoffs vorhanden.
Codierregel und Beispiele	Es wird beschrieben, dass sich an der Luft ein explosives Gemisch mit Wasserstoff bildet. Erst durch die spontane Reaktion des Wasserstoffs zu gasförmigem Wasser (Knallgasreaktion) wird die enorme Druckwelle, die zur Beschädigung der Brücke führte, erklärbar. Es reicht jedoch der Hinweis auf die „Explosivität des Wasserstoffs“ bzw. alternativ die Erwähnung der „Knallgasreaktion“.

Abbildung 13: Codierleitfaden zu einer offenen Aufgabe aus dem Chemieleistungstest

Halboffene Aufgaben sind ausschließlich als Ergänzungsaufgaben formuliert, bei der jeweils nur eine Antwort gegeben werden muss (Einfachantworten). Es muss entweder ein Stoff bzw. eine Stoffgruppe, eine Stoffeigenschaft oder eine Reaktionsgleichung angegeben werden. In der Regel sind mehrere Antworten möglich, die sich inhaltlich bzw. sprachlich unterscheiden können.

Sämtliche Antwortalternativen wurden im Vorfeld formuliert und in einem Expertenteam bestehend aus drei erfahrenen Chemielehrern besprochen (deduktives Vorgehen). Im Rahmen der Pilotierung wurden diese anschließend überprüft und gegebenenfalls durch weitere, akzeptable Schülerantworten ergänzt (induktives Vorgehen). Abschließend wurde ein Lösungsbogen verfasst, der alle als richtig zu bewertende Antwortalternativen enthält. Das Vorgehen ist ausführlich in Abschnitt 6.2 beschrieben. Untenstehende Abbildung zeigt ein Beispiel für diesen Aufgabentyp.

Text	Notiere einen Reaktionspartner, mit dem Erdalkalimetalle (<i>alkaline earth metals</i>) besonders gut reagieren!
mögliche Antworten	Sauerstoff, O, O ₂ , Wasser, H ₂ O
Codierung	RE- F- SE- HOA - PE - AFB I

Abbildung 14: Beispiel für eine halb-offene Aufgabe aus dem Chemieleistungstest

Die Reaktionen der Alkalimetalle an der Luft sowie mit Wasser wurden im Unterricht ausführlich besprochen. Neben der Angabe des Stoffnamens wird auch die Formelschreibweise akzeptiert, selbst wenn diese nicht ganz korrekt ist, wie z.B. bei der Antwort "O". Die Aufgaben, bei denen Reaktionsgleichungen angegeben werden müssen, erfordern eine differenziertere Beschreibung und Bewertung der Antwortalternativen. Dies liegt daran, dass die vollständige Lösung dieser Aufgaben gleichzeitig die Angabe korrekter Formeln, Koeffizienten und Aggregatzustände beinhaltet.

Teilpunkte werden vergeben, wenn nicht alle drei Aspekte berücksichtigt wurden. Da im Rahmen der Gruppentestung das Abschreiben vom Nachbarn weitgehend ausgeschlossen werden muss, wurden zwei parallele Testformen (Test A und Test B) konstruiert³⁸. Damit beide Tests möglichst äquivalent sind, wurde bei der Paralleltestkonstruktion nach dem von Bortz & Döring (2002, S. 197) vorgeschlagenen, vierschrittigen Verfahren vorgegangen.

38 Von der Möglichkeit Abschreibeefekte durch Änderung der Reihenfolge der Testaufgaben zu verringern (vgl. Riese & Reinhold, 2014) wurde abgesehen. Zum einen ist diese Methode nicht zuverlässig genug, zum anderen sollten Aufgaben in einem Niveautest nach steigender Schwierigkeit geordnet sein (vgl. Lienert und Raatz, 1994, S. 15).

1. Bildung eines Itempools
2. Itemanalyse
3. Bildung von Item-Zwillingen
4. Konstruktion des Paralleltests

Für die Bildung von Itemzwillingen bei heterogenen Leistungstests spielen nicht nur Reliabilitätsmerkmale wie Schwierigkeit und Trennschärfe eine Rolle, sondern insbesondere die Gleichartigkeit des Testinhalts (vgl. Lienert und Raatz, 1994, S. 306f). Deshalb wurden Itempaare mit exakt der gleichen Merkmalskombinationen (Codierung) entwickelt. Das bedeutet, dass sie in Bezug auf Komplexität der Inhalte, kognitive Prozesse, Teilbereiche/Basiskonzepte, Repräsentationsebenen und Anforderungsbereiche identisch sind. Inhaltliche Unterschiede sind bei den geschlossenen und halboffenen Aufgaben beider Testformen minimal und betreffen zum Beispiel die Wahl des Stoffes oder der betrachteten Stoffeigenschaft bzw. der Richtung, in welche sich eine bestimmte Stoffeigenschaft ändert. Sprachlich sind die Itempaare identisch, formale Unterschiede sind ebenfalls minimal und betreffen z.B. die Reihenfolge der Antwortoptionen. Die Abbildung unten zeigt ein Beispiel für ein solches Itempaar.

Testform A	Gibt man elementares Strontium zu Wasser, so wird die Lösung A sauer B alkalisch C neutral D schwach sauer
Testform B	Gibt man elementares Barium zu Wasser, so wird die Lösung A sauer B neutral C schwach sauer D alkalisch

Abbildung 15: Beispiel für ein Itempaar aus den Paralleltests

Die folgende Tabelle gibt einen Überblick über die Aufgabenmerkmale und die Häufigkeit ihres Vorkommens im Test.

Aufgabenmerkmal	Anzahl der Aufgaben
Kognitiver Prozess	
reproduzieren (RE)	14
selegieren (SE)	0
organisieren (OR)	18
integrieren (IN)	10
Komplexität des Inhalts	
Fakt (F)	8
Zusammenhang (Z)	27
übergeordnetes Prinzip (ÜK)	7
Teilbereich	
Struktur und Eigenschaften / Stoffe und Teilchen (SE)	18
Chemische Reaktion und Energieumsätze (CR)	7
Phänomene erklären, Bezüge finden und Abstraktionen nutzen (PBA)	17
Aufgabenformat	
geschlossen (GA)	26
offen (OA)	5
halboffen (HOA)	11
Repräsentationsebene	
Teilchenebene (TE)	8
Phänomenebene (PE)	34
Anforderungsbereich	
AFB I	14
AFB II	25
AFB III	3
Gesamt	42

Tabelle 10: Aufgabenmerkmale und ihr Auftreten im Chemieleistungstest

5.3.1.4 Auswertung der Ergebnisse

Mit Hilfe des Leistungstests soll der sachfachbezogene Lernzuwachs im Anschluss an das Modul gemessen werden. Dabei werden die Ergebnisse von Kontroll- und Untersuchungsgruppe miteinander verglichen. Für beide Gruppen wird ein Pretest unmittelbar vor dem Modul sowie ein Posttest unmittelbar nach dem Modul durchgeführt. Sechs Wochen später folgt ein Follow-Up Test, durch den ermittelt wird, inwieweit die gelernten Inhalte über einen längeren Zeitraum behalten wurden. Durch die Entwicklung paralleler Testformen ist es möglich, den Übungsfortschritt, der sich bei wiederholter Bearbeitung eines Tests ergibt, gering zu halten (Lienert und Raatz, 1994, S.

297). Die Probanden erhalten für die drei Testungen (Pre-, Post- und Follow-Up-Test) die jeweils andere Testform. Die Klassen werden dazu anhand des Sitzplans in Gruppen eingeteilt, so dass Schüler, die nebeneinandersitzen (Gruppe 1 bzw. 2), jeweils eine andere Testvariante bearbeiten (vgl. Abbildung unten).

Test	Gruppe	Testform
Pretest	1	A
	2	B
Posttest	1	B
	2	A
Follow-Up-Test	wie Pretest	

Abbildung 16: Administration der Testvarianten A und B

Den Aufgaben werden Punkte zugeordnet, die sich am Umfang der Antwort orientieren. Für geschlossene Aufgaben wird 1 Punkt bei einer richtigen Antwort und 0 Punkte bei einer falschen Antwort vergeben. Die Bepunktung offener Aufgaben ergibt sich aus der jeweiligen Codiervorschrift und beträgt, je nach Zahl der zu berücksichtigenden Kategorien, bis zu drei Punkte. Damit ergibt sich bei insgesamt 57 Punkten folgende Verteilung der Punkte auf die AFB:

AFB I	15 Punkte (25 %)
AFB II	34 Punkte (60 %)
AFB III	8 Punkte (15 %)

Zusätzlich werden die Punktzahlen für den Anforderungsbereich I (Reproduktion von Wissen) und die Anforderungsbereiche II und III separat ausgewiesen, damit die Gruppen auch im Hinblick auf Unterschiede im Bereich der höheren Denkopoperationen (AFB II/III) verglichen werden können (Hypothese 4). Je nach Untersuchungsfrage wurde dann unterschiedlich mit den so erhaltenen Daten verfahren. Sollte die Lernerfolg unterschiedlicher Schülergruppen verglichen werden, so wurden Gruppenmittelwerte errechnet und auf statistisch signifikante Unterschiede getestet. Ging es darum, den Einfluss anderer Variablen auf den Lernerfolg zu ermitteln, so wurden anhand der Probandenmittelwerte Korrelations- bzw. Regressionsanalysen durchgeführt (vgl. Abschnitt 7.3).

5.3.2 (Fremd-)sprachlicher Lernerfolg

In der vorliegenden Studie wird die Entwicklung des fremdsprachlichen Wortschatzes von Schülern betrachtet, die an einem bilingualen Modul im Fach Chemie teilgenommen haben. Zahlreiche Untersuchungen dazu belegen, dass der fremdsprachliche Vorsprung der bilingualen Schüler insbesondere im Bereich der Lexik besonders hoch ist (vgl. Kapitel 4). Es zeigt sich, dass bilinguale Schüler, die über einen längeren Zeitraum am bilingualen Unterricht teilgenommen haben, über einen sehr viel größeren und differenzierteren Wortschatz verfügen als Regelschüler (vgl. Kapitel 4). Die meisten Untersuchungen zum Wortschatz beziehen sich auf kontinuierliche Formen und sind querschnittlich angelegt (vgl. Kapitel 4). Longitudinalstudien sind eher selten. Empirische Untersuchungen zum Wortschatz in diskontinuierlichen Formen des bilingualen Unterrichts fehlen bislang. So ist etwa das Ausmaß lexikalischer Zuwächse für den bilingualen Unterricht in Modulform noch nicht erforscht. Wie in der Literatur vielfach berichtet, kam in der Vergangenheit vermehrt die Schulung der muttersprachlichen Fachsprache im bilingualen Unterricht zu kurz, da dieser nicht bi- sondern überwiegend monolingual in der Fremdsprache erteilt wurde. Neben dem fremdsprachigen Wortschatz wird deshalb auch der muttersprachliche Fachwortschatz untersucht. Für die Fächer des mathematisch-naturwissenschaftlichen Fachbereiches wird außerdem bezweifelt, ob sie überhaupt einen Beitrag zur Erweiterung des Allgemeinwortschatzes leisten können, weil ja - so die Annahme - im Unterricht überwiegend fachsprachlich kommuniziert wird (vgl. Kapitel 4). Deshalb wird bei den Tests zum englischen Wortschatz zwischen Allgemein- und Fachwortschatz unterschieden.

5.3.2.1 Hypothesen und ihre Testungen

Aufgrund der Forschungsergebnisse zum Erwerb des mutter- und fremdsprachlichen Wortschatzes im bilingualen Unterricht werden folgende Hypothesen formuliert:

Hypothesen

4. Der deutsche Fachwortschatz wird von den auf Deutsch unterrichteten Schülern im Anschluss an das Modul besser beherrscht als von den bilingual unterrichteten Schülern.
5. Der fremdsprachliche Zugewinn der bilingual unterrichteten Schüler betrifft vorwiegend den Bereich des Fachwortschatzes. Zugewinne im Bereich des Allgemeinwortschatzes sind gering.

Abbildung 17: Hypothesen zum (fremd-)sprachlichen Lernerfolg

Variablen

Zur Erfassung des Lernzuwachses in den Bereichen deutscher Fachwortschatz, englischer Fachwortschatz und englischer Allgemeinwortschatz werden insgesamt vier Wortschatztests durchgeführt. Die Schüler erhalten Punkte für jedes Wort, das sie aufgrund einer vorgegebenen Bedeutung korrekt angeben können bzw. für jedes Wort, dem sie die korrekte Bedeutung zuordnen können. Der so ermittelte Gesamtpunktwert gibt Aufschluss über den Umfang des rezeptiven bzw. produktiven Wortschatzes in den Bereichen Fach- und Allgemeinsprache zum jeweiligen Testzeitpunkt. Die Tests werden insgesamt dreimal - vor dem Modul (Pretest), unmittelbar nach dem Modul (Posttest) und sechs Wochen später (Follow-Up-Test) - durchgeführt. Anschließend werden Gruppenmittelwerte bestimmt und miteinander verglichen. Zur Bestimmung des Lernzuwachses bzw. des Vergessens im Anschluss an das Modul wird für jeden Schüler zusätzlich die Differenz zwischen Post- und Pretestergebnis bzw. zwischen Post- und Follow-Up-Testergebnis ermittelt.

Operationalisierte Hypothesen

4. Der Anstieg der Gesamtpunktzahl im deutschen Fachwortschatztest vom Pre- zum Posttest ist bei den auf Deutsch unterrichteten Schülern im Mittel größer als bei den bilingual unterrichteten Schülern.
5. Bei den bilingual unterrichteten Schülern ist der Anstieg der Gesamtpunktzahl im fremdsprachigen Fachwortschatz größer als im fremdsprachigen Allgemeinwortschatz.

Abbildung 18: Operationalisierte Hypothesen zum (fremd-)sprachlichen Lernerfolg

5.3.2.2 Theoretischer Hintergrund und Aufbau der Wortschatztests

Die eingesetzten Wortschatztests erfassen das Vokabelwissen der Schüler anhand eines diskreten, selektiven und kontextunabhängigen Testverfahrens (Read, 2011). Zur Beurteilung der Testgüte werden dabei die klassischen psychometrischen Merkmale wie Objektivität, Reliabilität und Validität herangezogen. Damit stehen die Tests in der Tradition eines auf Lado (1961, S. 77) zurückgehenden Ansatzes, der darauf abzielt, das Wissen um die Bedeutung einzelner Wörter als individuelle Strukturelemente von Sprache zu erfassen. Die vorliegenden Tests dienen dazu, das Wissen der Schüler zum Wortschatz, der in der Unterrichtreihe vermittelt wurde, zu erfassen. Damit geben sie einen Hinweis darauf, inwieweit die Schüler die Bedeutung einzelner Begriffe verstanden haben (rezeptiver Wortschatz) und ob sie diese in einzelnen, ausgewählten Sätzen sinngebend einsetzen können (produktiver Wortschatz). Sie erlauben jedoch keine Aussage darüber, inwieweit

die Schüler tatsächlich in der Lage sind, dieses Wissen zum Lösen komplexerer Aufgaben, wie das Lesen eines Fachtextes oder das Verfassen eines Laborberichtes einzusetzen³⁹. Dazu wären kommunikative Tests nötig, die entsprechende Schreib- oder Leseaufgaben enthalten. Erst die Analyse der Lesetexte bzw. der von den Schülern produzierten Texte im Hinblick auf den Wortschatz erlaubt es dann, Aussagen darüber zu machen, inwieweit dieser zur Rezeption oder zur Produktion von Texten tatsächlich zur Verfügung steht. In der empirischen Forschung zum Wortschatz kommen deshalb heute überwiegend kommunikative Testformen zum Einsatz (Read, 2011, S: 4). Im Bereich des bilingualen Unterrichts wurden in der Vergangenheit auf unterschiedliche offene Formate wie das Rollenspiel (Burmeister und Daniel (2002) oder freie Schreibaufgaben (Kampmeier 2013) aber auch halb-offene Formate, wie die Satzergänzung (Zydati, 2007) oder der C-Test, bei dem Inhaltswörter getilgt sind (Baker, 2002), zurückgegriffen. Dennoch können auch anhand des Sprachwissens von Lernern Rückschlüsse auf ihre Sprachfähigkeiten gezogen werden. Laut Bachmann (1990, S. 5) ergibt sich sprachliches Können nämlich sowohl aus dem Wissen über die Elemente von Sprache (*language knowledge*) als auch aus der Fähigkeit, diese für kommunikative Zwecke einzusetzen (*strategic competence*). Wenngleich ihre Aussagekraft beschränkt ist, so eignen sich konventionelle Testformen jedoch vor allem dann, wenn das Wissen über bestimmte, etwa im Rahmen einer Unterrichtsreihe vermittelte Vokabeln gezielt untersucht werden soll, wie im vorliegenden Fall (Read, 2011, S. 42f.). Zur Überprüfung der Hypothesen werden insgesamt vier Wortschatztests durchgeführt. Die ersten beiden Tests betreffend den deutschen und den englischen Fachwortschatz, der im Rahmen des Moduls von den Schülern zu erlernen ist. Die Begriffe sind den Schülern auch im Deutschen unbekannt und waren bislang nicht Gegenstand des Chemieunterrichts. Der Test zum deutschen Fachwortschatz wird sowohl von den bilingual als auch von den auf Deutsch unterrichteten Schülern durchgeführt. Der Vergleich der Ergebnisse erlaubt eine Aussage darüber, inwieweit die Verwendung der Fremdsprachen den Erwerb der deutschen Fachterminologie im bilingualen Unterricht beeinflusst. Der englische Fachwortschatztest wird hingegen nur von den bilingual unterrichteten Schülern durchgeführt. Er bezieht sich auf exakt die gleichen Fachbegriffe, wie der Test zum deutschen Fachwortschatz. Der Vergleich der Ergebnisse dieser beiden Tests liefert Informationen darüber, inwieweit der Fachwortschatz in der Fremd- und in der Muttersprache beherrscht wird.

39 Vgl. Read (2011, S. 8) der zwischen *discrete* und *embedded* Tests unterscheidet, je nachdem, ob der Wortschatz isoliert oder als Teil eines größeren Konstrukts, wie z.B. des Leseverstehens, gesehen werden.

Die letzten beiden Tests betreffen den darüberhinausgehenden, zusätzlich erworbenen fremdsprachlichen Wortschatz. Dazu zählen Fachbegriffe, die den Schülern im Deutschen bereits aus dem Chemieunterricht der Klasse 7 bekannt sind (ZEFWT) sowie allgemeinsprachliche Begriffe (ZEAWT). Beide Tests werden ebenfalls nur von den bilingual unterrichteten Schülern absolviert. Anhand der Testergebnisse wird anschließend ermittelt, inwieweit der Wortschatzerwerb im vorliegenden Modul auch den Bereich der allgemeinsprachlichen Lexik betrifft. Es soll also ermittelt werden, welchen Beitrag der bilinguale Chemieunterricht zur Erweiterung des englischen Allgemeinwortschatzes leisten kann. In der folgenden Abbildung sind die beschriebenen Merkmale der vier Tests gegenübergestellt.

Test		Abkürzung	Sprache	Testnehmer	Bedeutung der Begriffe im Deutschen	Wortzahl
1	Deutscher Fachwortschatztest	DFWT	deutsch	UG / KG	unbekannt	20
2	Englischer Fachwortschatztest	EFWT	englisch	UG	unbekannt	20
3	Test zum zusätzlich erworbenen englischen Fachwortschatz	ZEFWT	englisch	UG	bekannt	25
4	Test zum zusätzlich erworbenen englischen Allgemeinwortschatz	ZEAWT	englisch	UG	bekannt	23

Tabelle 11: Inhalt und Umfang der Wortschatztests

Die Aufteilung des englischen Fachwortschatzes auf zwei Tests hat den Vorteil, dass ermittelt werden kann, ob die durch das Modul neu zu erlernenden deutschen Fachbegriffe von den bilingualen Schülern ähnlich gut beherrscht werden wie die englischen Fachbegriffe. Fachbegriffe zeichnen sich durch eine Reihe von Merkmalen aus. In der Regel sind es Wörter von geringer Frequenz, deren Gebrauch meist auf eine bestimmte Domäne beschränkt ist. Fachbegriffe haben darüber hinaus klar umrissene Bedeutungen und stehen in eindeutigen semantischen Relationen zueinander (vgl. Pearson, 1998). Damit tragen sie zu einer möglichst präzisen und ökonomischen Verständigung über ein bestimmtes Sachgebiet bei. In vielen Fällen ist die Zuordnung eines Begriffs zu Alltags- oder Fachsprache jedoch nicht eindeutig, etwa wenn dieser sowohl eine alltags- als auch eine fachsprachliche Bedeutung hat. Häufig ist die Alltagssprache von Fachbegriffen durchsetzt und in fachsprachlichen Kontexten finden sich Anteile von Alltagssprache. Dies gilt insbesondere für

den Unterricht in der Schule. Oft kann dann nur aus dem konkreten Kontext erschlossen werden, ob ein Begriff fach- oder allgemeinsprachlich benutzt wird (Nitz, Nerdel, & Prechtel, 2012). In der vorliegenden Unterrichtsreihe finden sich eine Vielzahl von Begriffen, die sowohl in ihrer wissenschaftlichen als auch in ihrer allgemeinsprachlichen Bedeutung von den Schülern zu erlernen sind und damit auch für die Kommunikation im Alltag zur Verfügung stehen, z.B. *solid, liquid, cloudy, vapour, to dissolve, to melt, to boil* etc. Gerade im Englischen werden wissenschaftliche Fachbegriffe oft auch mit ihrer nicht-wissenschaftlichen Bedeutung benutzt, was im Deutschen nicht immer möglich ist. Im Deutschen wird beispielsweise zwischen dem alltagssprachlichen *Kochen* und dem fachsprachlichen *Sieden* unterschieden, während im Englischen in beiden Fällen das Verb *to boil* verwendet wird. Eine Reihe von Vokabeln, die für die vorliegende Untersuchung dem Fachwortschatz zugerechnet werden, werden somit auch in ihrer alltagssprachlichen Bedeutung gelernt. Dies muss bei der Frage nach dem Beitrag des Unterrichtsfaches zur Förderung der Allgemeinsprache berücksichtigt werden. Der zu erlernende Wortschatz umfasst ausschließlich Inhaltswörter in Form von Nomen, Vollverben, Adjektiven und Adverbien. Diese können einzeln aber auch in Kombination mit anderen Wörtern auftreten, etwa um Kollokationen auszudrücken oder wenn sich daraus neue Bedeutungseinheiten, wie etwa bei *phrasal verbs* ergeben. Im Test taucht immer die Grundform des Wortes auf, flektierte Formen werden nicht einzeln getestet. Abgeleitete Formen werden immer dann separat getestet, wenn davon ausgegangen werden kann, dass die entsprechenden Wortbildungsregeln den Schülern nicht unmittelbar bewusst sind. So tauchen die Begriffe *flame, to inflame* und *flammable* oder *solution, soluble* und *solubility* einzeln in den Tests auf⁴⁰. Gemäß ihrer Bedeutung lassen sich die Vokabeln insgesamt neun unterschiedlichen Inhaltsfeldern zuordnen, wie die folgende Abbildung zeigt.

40 Ausnahmen stellen zum einem die regelmäßige Bildung von Adverbien aus Adjektiven durch Anhängen von *ly* dar. Diese Wortbildungsregel ist auch in Klasse 8 bereits hinreichend bekannt. Deshalb wurden die Vokabeln *vigorous* und *vigorously* nur einmal, und zwar in der Form *vigorous(ly)* notiert. Eine weitere Ausnahme ist die Verwendung von *solid* und *liquid* als Adjektiv und Nomen ohne Veränderung der Wortform. Hier wurde *solid* (Adj., N.) bzw. *liquid* (Adj., N.) notiert.

Fachwortschatz	Allgemeinwortschatz
Labor - Ausrüstung - Tätigkeiten und Verfahren	Küche und Alltag - Ausrüstung und Objekte - Tätigkeiten und Vorgänge
Stoffeigenschaften	Zustand und Eigenschaften von Objekten
Chemische Theorie	Veränderungen und Vorgänge
Naturwissenschaftlicher Erkenntnisweg	Sonstiges
Stoffbezeichnungen	

Tabelle 12: Inhaltsfelder der Wortschatztests

Für die Art und Weise wie der Wortschatz beherrscht werden soll, wird auf die Unterscheidung von rezeptivem und produktivem Wortschatzwissen zurückgegriffen (Read, 2011, S. 154). Rezeptiv wird eine Vokabel beherrscht, wenn der Lerner in der Lage ist, dieser eine Bedeutung zuzuordnen, wenn er ihr begegnet. Von produktiver Beherrschung spricht man, wenn die Vokabel vom Lerner selbst hervorgebracht werden kann, wenn er die entsprechende Bedeutung zum Ausdruck bringen will. In den vorliegenden Tests wird der rezeptive Wortschatz mit Hilfe von Multiple-Choice-Items überprüft. Dazu wird die entsprechende Vokabel in einem einzelnen Satz präsentiert und der Proband wählt zwischen vier Bedeutungen diejenige aus, die am besten passt. Zur Testung des produktiven Wortschatzes wird ein Lückensatz formuliert, wobei in Klammern Hinweise zur Bedeutung des gesuchten Wortes gegeben sind. Der Schüler notiert die Vokabel, die seiner Meinung nach am besten in die Lücke passt. Read (2011, S. 155ff.) weist darauf hin, dass das bloße Wiedererkennen eines Wortes, welches in einem minimalen Kontext präsentiert wird, nicht gleichzusetzen ist mit dem Verstehen eines Wortes in einem größeren Kontext, etwa wenn der Lerner diesem in einem Text begegnet. In der Testsituation wählt der Lerner zwischen vorgegebenen Bedeutungen aus und kann gegebenenfalls über Ausschlussverfahren die Wortbedeutung erschließen. In Texten sind es insbesondere Kontextinformationen, die das Erschließen der Wortbedeutung ermöglichen. So muss das richtige Lösen der Testaufgabe nicht zwangsläufig bedeuten, dass der Lerner die Bedeutung auch in realen Kommunikationssituation zuordnen kann, und umgekehrt. Gleiches gilt für die produktive Beherrschung eines Wortes. Das reine Erinnern eines Wortes aufgrund einer zu ergänzenden Lücke in einem bestimmten Einzelsatz bedeutet nicht, dass der Lerner die entsprechende Vokabel auch für die eigene Sprachproduktionen

gezielt einsetzen kann, wenn die Produktion mehr als ein einzelnes Wort umfasst⁴¹. Die hier gewonnenen Daten können somit lediglich den Lernfortschritt der Schüler im Bereich Wortschatzwissen dokumentieren. Aussagen darüber, ob die Vokabeln auch in realen Kommunikationssituationen zur Verfügung stehen, sind nur eingeschränkt möglich. Die Zuordnung der Vokabel zum rezeptiven bzw. produktiven Wortschatz richtet sich nach ihrem Gebrauchswert. Der getestete produktive Wortschatz enthält zentrale Begriffe der Reihe, von hohem Gebrauchswert für den sachfachlichen und allgemeinsprachlichen Diskurs. Die Schüler sollen später möglichst in der Lage sein, diese Begriffe für eigene Formulierungen produktiv zu nutzen. Der rezeptiv getestete Fachwortschatz enthält weniger zentrale Begriffe der Reihe, von geringerem Gebrauchswert. Hier reicht es, dass Schüler diese Begriffe verstehen, wenn sie sie hören oder lesen. Dabei ist zu beachten, dass der Gebrauchswert von Fachbegriffen domänenspezifisch zu sehen ist. So gibt es Fachbegriffe, die im Bereich des Fachdiskurses besonders häufig gebraucht werden und andere, die eher seltener vorkommen. Es werden nicht alle in der Unterrichtsreihe zu erlernenden Vokabeln getestet, da die Tests dann zu umfangreich wären. Die Auswahl der einzelnen Vokabel für eine selektive, stichprobenartige Testung sowie die Zuordnung zum rezeptiven und produktiven Wortschatz werden in den Kapiteln zu den einzelnen Tests erläutert. Um das Abschreiben zu verhindern, werden zwei Testformen von jedem Test erstellt (Test A und Test B), die sich in der Abfolge der Vokabeln unterscheiden.

5.3.2.3 Der deutsche Fachwortschatztest

Der deutsche Fachwortschatztest enthält Fachbegriffe, die im Rahmen des Moduls neu eingeführt werden, z.B.: *Alkalimetall*, *Hydroxid*, *Sieblöffel*. Er dient dazu festzustellen, ob diese von den bilingual unterrichteten Schülern ebenso gut beherrscht werden, wie von den auf Deutsch unterrichteten Schülern.

Auswahl der Vokabel

Die Fachbegriffe betreffen die beiden Wortfelder Stoffbezeichnungen (SB) und Labor (L). Von diesen werden jedoch solche nicht getestet, deren Namen sich systematisch aufgrund einer Bildungsregel herleiten lassen. Dies betrifft die Namen der zehn Alkali- und Erdalkalimetalloxiden

41 Read (2011, S. 155ff.) unterscheidet beim rezeptiven Wortschatz zwischen dem bloßen Wiedererkennen eines Einzelwortes (*recognition*) und dem Verständnis im Kontext (*comprehension*). Für den produktiven Wortschatz differenziert er analog zwischen dem Erinnern eines einzelnen Wortes (*recall*) und dem Gebrauch der Vokabel im Rahmen der eigenen Textproduktion (*use*). Insofern geht es bei den hier vorliegenden Tests jeweils um *recognition* und *recall*.

und -hydroxiden. Die Bildungsregel lautet: Hänge das Wort Oxid bzw. Hydroxid an den Metallnamen. Die Regeln werden nur anhand eines Beispiels (Metallhydroxid) überprüft. Damit verbleiben insgesamt 20 Vokabeln für den Test. Aufgrund des Gebrauchswertes der Begriffe im fachspezifischen Kontext werden elf Vokabeln dem rezeptiven und neun Vokabeln dem produktiven Wortschatz zugeschrieben.

Wortfeld	Produktiv	Rezeptiv
Stoffbezeichnung	Lithium, Natrium, Kalium, Calcium, Hydroxid, Alkalimetall, Erdalkalimetall, Natronlauge, Kalkwasser ⁴²	Rubidium, Caesium, Beryllium, Strontium, Barium, Kalkmilch, Granalien
Labor		Magnesium, Sieblöffel, pneumatisches Prinzip, Schutzflüssigkeit,

Tabelle 13: Deutscher Fachwortschatz

Konstruktion der Testaufgaben

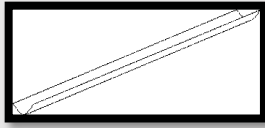
Der rezeptive Wortschatz wird im Multiple-Choice-Verfahren getestet. Dabei ist der Begriff genannt und eine Bedeutung muss aus vier Alternativen ausgewählt werden. Die Bedeutung wird durch eine Abbildung (Typ 1) oder durch eine Umschreibung / chemische Formel (Typ 2) angegeben. Zusätzlich wird ein Beispielsatz präsentiert, der die Verwendung des Begriffs im Satz zeigt. Dieser enthält jedoch keinen Hinweis auf die Wortbedeutung (nicht-definierender Kontext). Aus dem Beispielsatz ist die grammatische Bedeutung (z.B. Wortart) ableitbar. Das Format ist in standardisierten Wortschatztests weit verbreitet (Read, 2011, S. 2).

42 Die Bezeichnung *Kalkwasser* ist den Schülern zwar als Mittel zum Nachweis von Kohlenstoffdioxid bekannt, wird hier aber als Trivialname für eine wässrige Calciumhydroxidlösung konzeptualisiert und ist den Schülern in dieser Bedeutung neu.

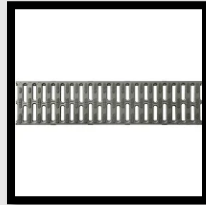
Für das Experiment wurde eine Magnesiumrinne benötigt.

Welches Bild zeigt eine Magnesiumrinne?

A



B



C



D



Abbildung 19: Multiple-Choice Item zum rezeptiven Fachwortschatz (Typ 1)

Das Verfahren ist jedoch problematisch, da die Kenntnis des Begriffs zur Lösung der Aufgabe nicht immer unbedingt nötig ist. Nach dem Ausschlussverfahren kann man immer dann auf die richtige Bedeutung bzw. die richtige Bezeichnung schließen, wenn die anderen drei Distraktoren bekannt bzw. unbekannt sind. Es reicht mitunter aus, den Gegenstand einmal gesehen zu haben, ohne zu wissen, wie er heißt.⁴³ Die Güte der Distraktoren muss deshalb im Rahmen der Pilotierung genau überprüft werden. Während bei Laborgeräten dazu auf Abbildungen zurückgegriffen werden kann, ist das bei Stoffbezeichnungen nicht möglich. Die Angabe der Wortbedeutung ist hier in vielen Fällen problematisch. Dabei geht es um die Frage, welches inhaltliche Wissen vorhanden sein muss, damit der Begriff als „gewusst“ bezeichnet werden kann. Die Frage betrifft die Qualität des Wissens, also die sogenannte Tiefe des Wortschatzwissens (*depth of vocabulary knowledge*) (Read, 2011, S. 25ff.). Die Problematik verdeutlicht folgendes Beispiel:

Er nahm etwas Rubidium.

Rubidium ist ein/e

A durchsichtiger Kristall

B saure Lösung

C reaktives Metall

D roter Farbstoff

Abbildung 20: Multiple-Choice Item zum rezeptiven Wortschatz (Typ 2)

Das Element *Rubidium* hat das Elementsymbol *Rb* und lässt sich eindeutig durch seine Stellung im Periodensystem der Elemente definieren (1. Hauptgruppe, 5. Periode [Ordnungszahl 37]). Diese Daten muss ein Schüler in Klasse 8 jedoch nicht unbedingt kennen. Vielmehr sollte er wissen, dass es sich bei Rubidium um ein Alkalimetall handelt (1. Hauptgruppe) und dass es ein weiches und

⁴³ Vergleiche hierzu Reads (2011, S. 77ff.) Ausführungen zur Problematik der Testung isolierter Vokabeln nach dem Multiple-Choice-Verfahren.

reaktives Metall ist, welches an der Luft leicht Oxide und in Kontakt mit Wasser leicht Hydroxide bildet. Diese Eigenschaften treffen jedoch auf alle Alkalimetalle zu. Tatsächlich unterscheiden sie sich hier nur graduell und das wiederum sollte ein Schüler wissen und entsprechend einschätzen können. Eine derart differenzierte Überprüfung der Wortbedeutung ist jedoch sehr aufwendig und für die vorliegenden Wortschatztests nicht praktikabel. Die meisten der oben genannten Bedeutungsaspekte werden an anderer Stelle, nämlich im Test zum Fachwissen überprüft. An dieser Stelle reicht es deshalb aus, zu wissen, dass es sich bei Rubidium um ein reaktives Metall handelt (rezeptiver Wortschatz). Das Beispiel zeigt, wie eng Fachsprache und Fachwissen miteinander verknüpft sind. Der produktive Wortschatz wird durch Lückensätze getestet, die einen minimalen Kontext erzeugen, der die Verwendung des Begriffs im Satz verdeutlicht. Außerdem werden weitere Hinweise auf die Bedeutung gegeben (z.B. Formel, Definition, Erläuterung).

_____ (K) ist ein weiches und sehr reaktives Metall.

Abbildung 21: Ergänzungsaufgabe zum produktiven Fachwortschatz

Bei den Alkalimetallen, die aufgrund ihrer Wichtigkeit im Unterricht zum produktiven Wortschatz der Schüler gehören (das sind Lithium, Natrium und Kalium), sollte auch das Elementsymbol gewusst werden. Es wird entsprechend häufig im Verlauf der Reihe benutzt und stellt einen eindeutigen Hinweis auf den gesuchten Begriff dar.

Auswertung der Ergebnisse

Für jede richtige Antwort wird ein Punkt vergeben, insgesamt können 20 Punkte erreicht werden. Rechtschreibfehler führen aus Gründen einer größeren Auswertungsobjektivität ebenfalls zum Punktabzug. Aus den Testergebnissen werden Mittelwerte für den Pre-, Post- und Follow-Up-Test in beiden Gruppen gebildet. Außerdem werden für jeden einzelnen Probanden die Zu- bzw. Abnahmen in den Testwerten als Differenz zwischen dem Post- und Pretest, zwischen Follow-Up und Posttest sowie zwischen Follow-Up und Pretest berechnet. Damit ergeben sich Hinweise auf den unmittelbaren Lernzuwachs im Anschluss an das Modul, das Vergessen im Anschluss sowie den mittelfristigen Lernzuwachs nach sechs Wochen.

5.3.2.4 Der englische Fachwortschatztest

Der englische Fachwortschatztest enthält die gleichen Fachbegriffe wie der deutsche Fachwortschatztests. Dadurch ist es möglich, die Ergebnisse beider Tests direkt miteinander zu vergleichen. So lässt sich ermitteln, ob die neu eingeführten deutschen und die englischen Fachbegriffe von den bilingual unterrichteten Schülern gleich gut beherrscht werden. Eine Analyse von Band 4 des in der Sekundarstufe I eingeführten Lehrwerkes *Access G8* von Cornelsen ergab, dass keiner der zu erlernenden Fachbegriffe im Englischunterricht bis Klasse 8 eingeführt wird. Etwa zwei Drittel der englischen Fachbegriffe sind den deutschen sehr ähnlich und unterscheiden sich nur in der Rechtschreibung. Dies betrifft insbesondere die Stoffbezeichnungen, weil die Namen den gleichen Ursprung haben bzw. den gleichen Nomenklaturregeln folgen (vgl. Tabelle unten).

Fachbegriffe	Anzahl	Deutsch	Englisch
Identisch / ähnlich	13	Lithium, Natrium, Calcium, Rubidium, Caesium, Beryllium, Strontium, Barium, (Metall-) Hydroxid, Alkalimetall, Erdalkalimetall, pneumatisches Prinzip, Schutzflüssigkeit,	lithium, sodium, calcium, rubidium, caesium, beryllium, strontium, barium, (metal-) hydroxide, alkali metal, alkaline earth metal, pneumatic principle, protective liquid
verschieden	7	Kalium, Natronlauge, Kalkwasser, Kalkmilch, Magnesiumrinne, Sieblöffel, Granalien	potassium, caustic soda, lime water, lime milk, magnesia tray, sodium spoon, granules

Tabelle 14: Vergleich von deutschem und englischem Fachwortschatz

Konstruktion der Testaufgaben

Die Tests zum deutschen und zum englischen Fachwortschatz sind zwar gleich aufgebaut, dennoch gilt es, bei der Konstruktion der Aufgaben des englischen Fachwortschatztests einige Dinge zu beachten. Die Angabe der deutschen Übersetzung ist z.B. nicht möglich, weil die Begriffe auch im Deutschen neu sind. Es ist denkbar, dass ein Schüler die Bedeutung des englischen Fachbegriffs kennt, nicht aber die des deutschen. Er würde die Aufgabe dann nicht lösen können, obwohl er den englischen Fachbegriff eigentlich kennt. Außerdem muss ausgeschlossen werden, dass Schüler, die den Begriff auf Deutsch beherrschen, aufgrund der von ihnen erkannten Parallelität der Tests die korrekte Lösung im englischen Test erschließen können.⁴⁴ Die Problematik soll an einem Beispiel

⁴⁴ Die bilingual unterrichteten Schüler führen den Test zum deutschen Fachwortschatz vor dem Test zum englischen Fachwortschatz durch. Die umgekehrte Reihenfolge ist nicht sinnvoll, da durch den zusätzlichen Input durch einen

verdeutlicht werden. Bei dem Item zum „Sieblöffel“ ist im deutschen Fachwortschatztest die korrekte Abbildung aus vier Möglichkeiten auszuwählen. Ein Schüler, der den Begriff kennt, kann ihn im deutschen Test der richtigen Abbildung zuordnen. Findet er die gleichen vier Abbildungen im englischen Test wieder, so wird er dem Begriff *sodium spoon* mühelos das richtige Bild zuordnen, auch wenn er diese Vokabel im Englischen gar nicht kennt. Die Aufgaben des englischen Fachwortschatztests müssen deshalb so verändert werden, so dass ein Wiedererkennen der Vokabel ausgeschlossen ist. Deshalb werden ein anderer Beispielsatz sowie andere Antwortmöglichkeiten verwendet (siehe Abbildung unten).

They needed a sodium spoon.

Der Begriff sodium spoon bezeichnet einen Löffel zum

A Abfüllen von Natrium.

B Festhalten von Natrium.

C Verbrennen von Natrium.

D Zerteilen von Natrium.

Abbildung 22: Parallelitem aus dem englischen Fachwortschatztest (Sieblöffel)

Während die Distraktoren im deutschen Fachwortschatztest auf das Aussehen abzielen, geht es im englischen Fachwortschatztest um die Funktion des Laborgeräts. Es wird also auf zwei verschiedene, aber gleichermaßen zentrale Aspekte der Wortbedeutung abgezielt, die beide vom Schüler gewusst werden sollten⁴⁵. Bei Begriffen, die im Deutschen und Englischen sehr ähnlich sind, ist es außerdem denkbar, dass Schüler den deutschen Begriff kennen und die Bedeutung des englischen herleiten können. So kann ein Schüler aus der Kenntnis des Begriffs *spoon* folgern, dass es sich bei dem *sodium spoon* wohl um den *Sieblöffel* handeln muss, insbesondere wenn dieser gerade im vorangegangenen deutschen Fachwortschatztest vorkam⁴⁶. Eine zeitlich verzögerte Darbietung des Tests zur englischen Fachsprache ist somit zusätzlich unumgänglich, um Behaltenseffekte zu minimieren. Zwischen der Darbietung des deutschen und des englischen

vorgeschalteten englischen Fachwortschatztest die bilingualen Schüler unter Umständen einen Vorteil im deutschen Fachwortschatztest gegenüber den auf Deutsch unterrichteten Schülern haben.

45 Inwieweit sich daraus Unterschiede in der Aufgabenschwierigkeit ergeben, wird im Rahmen der Pilotierung untersucht.

46 Solche Ähnlichkeiten sind auch bei einer Reihe anderer Fachbegriffe der Unterrichtsreihe vorhanden, wie z.B. bei Magnesiumrinne und *magnesia tray*, pneumatisches Prinzip und *pneumatic principle* oder Schutzflüssigkeit und *protective liquid*.

Fachwortschatztests liegt deshalb in der Regel mindestens ein Tag. Die folgende Tabelle gibt einen Überblick über alle Begriffe im deutschen und englischen Fachwortschatztest.

Fachwortschatz (20)			
Rezeptiv (11)		Produktiv (9)	
protective liquid	Schutzflüssigkeit	potassium	Kalium
rubidium	Rubidium	lithium	Lithium
sodium spoon	Sieblöffel	calcium	Calcium
strontium	Strontium	hydroxide	Hydroxide
beryllium	Beryllium	sodium	Natrium
lime milk	Kalkmilch	alkaline metals	Alkalimetalle
barium	Barium	alkaline earth metals	Erdalkalimetalle
caesium	Caesium	caustic soda	Natronlauge
magnesia tray	Magnesiumrinne	lime water	Kalkwasser
granules	Granalien		
pneumatic principle	pneumatisches Prinzip		

Tabelle 15: Fachbegriffe im deutschen und englischen Fachwortschatztest

Auswertung der Ergebnisse

Für jede richtige Antwort wird ein Punkt vergeben, insgesamt können, genau wie im Test zum deutschen Fachwortschatz, 20 Punkte erreicht werden. Rechtschreibfehler führen ebenfalls wieder zu einem Punktabzug. Aus den Testergebnissen der Gruppenmitglieder werden Mittelwerte für den Pre-, Post- und Follow-Up-Test in beiden Gruppen gebildet. Diese werden anschließend mit den Ergebnissen aus dem deutschen Fachwortschatztest der Schüler verglichen. Außerdem werden wieder für jeden einzelnen Probanden die Zunahme und die Abnahmen in den Testwerten als Differenz zwischen dem Post- und Pretest (unmittelbarer Zugewinn), zwischen Follow-Up- und Posttest (Vergessen) sowie zwischen Follow-Up- und Pretest (mittelfristiger Zugewinn) berechnet. Es werden nur die Daten jener Probanden berücksichtigt, die alle drei Tests absolviert haben. Für diese Differenzwerte werden wieder die Gruppenmittelwerte gebildet und mit den Werten aus dem deutschen Fachwortschatztest verglichen. Außerdem werden die Mittelwerte für die Zugewinne und das Vergessen im deutschen und englischen Fachwortschatz nach produktiv und rezeptiv zu beherrschenden Vokabeln differenziert.

5.3.2.5 Tests zum zusätzlich erworbenen englischen Wortschatz

Auswahl der Vokabeln und Konstruktion der Aufgaben

Durch den Test zum zusätzlich erworbenen Wortschatz soll überprüft werden, inwieweit der fremdsprachliche Wortschatz der bilingual unterrichteten Schüler eine Erweiterung erfährt, die über die zentrale Fachterminologie der Unterrichtsreihe hinaus geht. Der Fachwortschatzteil umfasst solche Fachbegriffe, die, anders als die Begriffe im Test zum englischen Fachwortschatz (EFWT), bereits im deutschsprachigen Chemieunterricht eingeführt wurden. Die zugrunde liegenden Konzepte sind den Schülern also bekannt. Der allgemeinsprachliche Teil enthält Vokabeln, die für die inhaltliche Arbeit benötigt werden und die aus dem Englischunterricht noch nicht bekannt sind⁴⁷. Außerdem wurde bei der Planung der Unterrichtsreihe festgelegt, ob die Begriffe Teil des produktiven oder Teil des rezeptiven Wortschatzes sein sollen. Diese Entscheidung basierte auf ihrem Gebrauchswert in einem Chemie-spezifischen (Fachwortschatz) bzw. einem allgemeinsprachlichen (Allgemeinwortschatz) Kontext. Damit ergaben sich insgesamt 85 Vokabeln, etwa gleich viele für den Fachwortschatzteil (45) und für den allgemeinsprachlichen Teil (40). Im Test wurden diese nicht vollständig, sondern stichprobenartig überprüft, da der Test ansonsten zu umfangreich gewesen wäre. Im Fachwortschatzteil wurden 25 und im allgemeinsprachlichen Teil 23 Vokabeln getestet, was für Schüler der Klasse 8 durchaus vertretbar ist. Größere Einbußen in der Motivation bzw. Konzentrationsfähigkeit sind bei diesem Umfang noch nicht zu erwarten. Die Vokabeln wurden in vier Gruppen eingeteilt (Fachwortschatz rezeptiv, Fachwortschatz produktiv, Allgemeinwortschatz rezeptiv und Allgemeinwortschatz produktiv). Die Auswahl erfolgte nach dem Zufallsprinzip. Dazu wurden die Vokabeln in jeder Gruppe nummeriert und die Nummern auf Zetteln notiert. Die Zufallsstichprobe erfolgte durch Ziehung der Nummern, die Zusammenstellung wurde im Rahmen der Pilotierung nochmals überarbeitet (vgl. Abschnitt 6.4). Der ZEAW-Test und der ZEFW-Test haben einen ähnlichen Umfang sowie ein annähernd identisches Verhältnis von produktiv und rezeptiv getesteten Vokabeln.

47 Dies wurde anhand des Vokabelverzeichnis des eingeführten Englischlehrwerkes Access A4 G8 (Cornelsen) überprüft.

	Fachwortschatz (ZEFW)		Allgemeinwortschatz (ZEAW)	
	Im Test	Im Modul	Im Test	Im Modul
Auftreten				
Gesamt	25	45	23	40
Rezeptiv	10	12	10	14
Produktiv	15	33	13	26

Tabelle 16: Umfang des zusätzlich erworbenen englischen Wortschatzes

Die folgenden Tabellen geben einen Überblick über die im Test vorkommenden Begriffe und den Wortfeldern, denen sie entstammen.

Wortfeld	Produktiv	Rezeptiv
Labor - Ausrüstung - Tätigkeiten und Verfahren	spatula, test tube, pop test, burner	clay triangle, tripod, set up, evaporating dish, forceps
Stoffeigenschaften	dissolve, solution, density, alkaline	soluble, property
Chemische Theorie	oxidation, redoxreaction, exothermic, equation	compound
Naturwissenschaftlicher Erkenntnisweg	observation	procedure
Stoffbezeichnungen	hydrogen, oxygen	lime water ⁴⁸

Tabelle 17: Fachbegriffe des Tests zum zusätzlich erworbenen Fachwortschatz

48 Die Vokabel *lime water* kommt sowohl im englischen Fachwortschatztest (EFTW) als auch im Test zum zusätzlich erworbenen Fachwortschatz (ZEFWT) vor - jedoch mit unterschiedlichen Bedeutungen. Während der deutsche Begriff (*Kalkwasser*) den Schülern als Bezeichnung für ein häufig benutztes Nachweisreagenz bekannt ist (ZEFWT), lernen die Schüler die zusätzliche Bedeutung als Bezeichnung für eine wässrige Lösung eines Erdalkalimetallhydroxids (*Calciumhydroxidlösung*) im Rahmen dieser Unterrichtsreihe kennen (EFTW).

Wortfeld	Produktiv	Rezeptiv
Küche und Alltag - Ausrüstung und Objekte - Tätigkeiten und Vorgänge	to heat, crust, combustion, to use up	sieve
Zustand und Eigenschaften von Objekten	cloudy, solid, liquid	exposed to, to gleam, to tarnish
Veränderungen und Vorgänge	to rise, to decrease, to escape from	to float, to ignite
Sonstiges	to contain, to observe, to conclude,	vigorously, residue, to be released, sphere

Tabelle 18: Vokabeln des Tests zum zusätzlich erworbenen Allgemeinwortschatz

Wie in den Tests zum deutschen und englischen Fachwortschatz, wird der produktive Wortschatz durch Lückensätze und der rezeptive durch Multiple-Choice-Single-Select-Aufgaben getestet. Da den Schülern die Bedeutung der Begriffe im Deutschen meistens bekannt ist, kann - anders als im EFW-Test - bei der Abfrage der Vokabeln in vielen Fällen auch auf die jeweilige deutsche Übersetzung zurückgegriffen werden. Bei einigen, im Chemieunterricht der Klasse 7 eingeführten Fachbegriffe, deren Beherrschung chemisches Theoriewissen voraussetzt, wie z.B. *compound* (Verbindung), *density* (Dichte) oder *exothermic* (exotherm) ist fraglich, inwieweit die jeweilige fachwissenschaftliche Bedeutung den Schülern tatsächlich noch klar ist, obwohl sie das entsprechende deutsche Pendant angeben können. Der Test gibt in diesen Fällen lediglich einen Hinweis darauf, ob der entsprechende Begriff mit dem gleichen konzeptuellen Wissen verknüpft ist wie seine deutsche Entsprechung, nicht aber, wie präzise dieses konzeptuelle Wissen ist. Genauere Informationen zum chemischen Vorwissen der Schüler liefert der Test zur allgemeinen Leistungsfähigkeit in Chemie (ACLT).

Auswertung der Ergebnisse

Für jede „gewusste“ Vokabel wird ein Punkt vergeben, so dass die Gesamtpunktzahl je Test 25 bzw. 23 beträgt, jeweils zehn Punkte im rezeptiven und 13 bzw. 15 Punkte im produktiven Teil. Rechtschreibfehler führen, wie in den anderen Wortschatztests auch, zu Punktabzug. Die Tests zum zusätzlich erworbenen Fach- und zum Allgemeinwortschatz werden separat ausgewertet. Dazu werden Gruppenmittelwerte für die drei Testzeitpunkte ermittelt und überprüft, ob statistisch signifikante Veränderungen vorliegen. Diese Veränderungen werden auch differenziert nach rezeptivem und produktivem Wortschatz ausgewertet. Auch dazu werden wieder Gruppenmittelwerte berechnet. Anders als in den Tests 1 (DFWT) und 2 (EFWT), bei denen alle zu

erlernenden Fachbegriffe getestet wurden, wird in Test 3 (ZEFWT) und Test 4 (ZEAWT) nur stichprobenartig getestet. Deshalb wird später anhand des Testergebnisses und der Angaben zum Gesamtumfang des zusätzlichen Wortschatzes geschätzt, wie viele Vokabeln insgesamt gelernt wurden. Dazu wird der Quotient aus den im Test „gewussten“ Vokabeln und der Gesamtzahl der Vokabeln im Test mit der Zahl der insgesamt zu erlernenden Vokabeln multipliziert. Zur Bestimmung des insgesamt erworbenen englischen Fachwortschatzes müssen außerdem noch die Zuwächse aus Test 2 (EFWT) berücksichtigt werden (vgl. Abschnitt 7.4.3). Anhand dieser Werte können die Gesamtzuwächse im Fach- und im Allgemeinwortschatz miteinander verglichen werden. So kann auch der durchschnittliche Vokabelzuwachs pro Stunde ermittelt, und mit den vorliegenden Werten für den Fremdsprachenunterricht verglichen werden (siehe Kapitel 8).

5.3.3 Motivation

Die Forschung zum bilingualen Unterricht ist in letzter Zeit verstärkt der Frage nach den Auswirkungen des neuen Lernsettings auf die Motivation nachgegangen. Die Bemühungen sind dabei für die naturwissenschaftlichen Sachfächer mit der Hoffnung verbunden, lernförderliche motivationale Aspekte, wie großes Interesse, positives fachbezogenes Selbstkonzept oder persönliche Nützlichkeitsüberzeugungen von der Fremdsprache auf das Sachfach übertragen zu können. Gleichzeitig stehen Befürchtungen im Raum, bilingualer Unterricht könne bei weniger fremdsprachlich interessierten Schülern auf Ablehnung stoßen und sich negativ auf Motivation und Leistung auswirken. Das *Internal/External-Frame-of-Reference-Modell* von Marsh (1986) erklärt, warum naturwissenschaftlich interessierte Schüler häufig ein geringes Interesse an Sprachen haben, während fremdsprachlich interessierte Schüler sich oft weniger für Naturwissenschaften interessieren. Zur Frage, welchen Einfluss die Zugehörigkeit zu einer bestimmten Interessensgruppe (Orientierung) auf die Motivationen im naturwissenschaftlichen bilingualen Unterricht hat, liegen zwar erste Ergebnisse für die Fächer Biologie, Physik und Mathematik vor (vgl. Piesche, 2016; Rodenhauser, 2016; Verriere, 2014; Lipski-Buchholz, 2019), für das Fach Chemie fehlen entsprechende Daten jedoch bislang.

5.3.3.1 Hypothesen und ihre Testung

Vor dem Hintergrund der vorliegenden Forschungsergebnisse zum bilingualen naturwissenschaftlichen Unterricht und der sich daraus ergebenden offenen Fragen in Bezug auf das Fach Chemie, werden für den Bereich der Motivation zwei Hypothesen formuliert. Diese sollen im Rahmen der Untersuchung überprüft werden.

Hypothesen

6. Fremdsprachlich orientierte Lerner sind im bilingualen Chemieunterricht stärker motiviert als im regulären Chemieunterricht.
7. Naturwissenschaftlich orientierte Lerner sind im regulären Chemieunterricht stärker motiviert als im bilingualen Chemieunterricht.

Abbildung 23: Hypothesen zur Motivation

Variablen

Zur Erfassung der intrinsischen Motivation wird auf die Selbstbestimmungstheorie nach Deci und Ryan zurückgegriffen. Interesse und Kompetenzerleben bilden dabei zusammen mit fünf weiteren Variablen das Konstrukt der *intrinsischen Motivation* ab (Deci & Ryan, 2003)⁴⁹. Für empirische Untersuchungen zur Erfassung der intrinsischen Motivation wird häufig auf die sogenannten "Kurzskala zur intrinsischen Motivation (KIM)" nach Deci und Ryan (2003) zurückgegriffen. Diese umfasst vier der insgesamt sieben Variablen, nämlich *Interesse/Vergnügen*, *wahrgenommene Kompetenz*, *wahrgenommene Wahlfreiheit* und *Druck/Anspannung* (vgl. Hartmansgruber, 2014; Rodenhauser, 2016 oder Wilde, Bätz, Kovaleva & Urhahne, 2009). In der vorliegenden Untersuchung wurde die Subskala wahrgenommene Wahlfreiheit durch die Subskala *Nutzen* ersetzt. Wie bereits gesehen, ist ein beträchtlicher Teil der Motivation für den bilingualen Unterricht instrumenteller Natur, dabei geht es um die Verbesserung fremdsprachlicher Kompetenzen - etwa als Vorbereitung auf das Berufsleben. Inwieweit bilingualer Chemieunterricht aus Sicht der Schüler dazu beitragen kann, soll deshalb an dieser Stelle untersucht werden. Damit ergeben sich für den Bereich der motivationalen Disposition folgende unabhängige Variablen: Interesse, Kompetenzerleben, Nutzen und Druck. Die entsprechenden Daten werden über die Selbsteinschätzung der Schüler (Introspektion) im Rahmen von Fragebogenbefragungen erfasst. Dieses Verfahren ist in der Motivationsforschung weit verbreitet, da motivationale Konstrukte einer direkten Messung nicht zugänglich sind und durch Untersuchungen der Unterrichtsrealität (Außenperspektive) z.B. mittels Videografie nur schwer zu erfassen sind. Insbesondere bei großen Probandenzahlen sind Fragebögen auch einer direkten Befragung im Rahmen von Interviews aus praktischen Gründen vorzuziehen. Auch Dörnyei (2001, S. 189) unterstreicht die Bedeutung von Fragebögen in der Motivations- und Einstellungsforschung und merkt an, dass sie eines der am häufigsten verwendeten Mittel in der Motivationsforschung darstellen. Auch Verriere (2014),

49 Zu den weiteren Variablen gehören: Bemühen, Druck, wahrgenommene Wahlfreiheit, Nutzen und Eingebundenheit (Deci & Ryan, 2003).

Hartmansgruber (2014), Piesche (2016), Rodenhauser (2016) und Lipski-Buchholz (2019) setzen in ihren Studien zur Motivation Fragebögen ein. Um ein gewisses Maß an Vergleichbarkeit zu gewährleisten, kommen in der Regel teilstandardisierte Fragebögen zum Einsatz, und dort wo möglich werden Fragebögen verwendet, die bereits im Rahmen anderer Untersuchungen erprobt wurden.

Operationalisierte Hypothesen

6. Im bilingualen Chemieunterricht sind die Werte für die Motivation im Mittel bei fremdsprachlich orientierten Schülern höher als im regulären Chemieunterricht.
7. Im bilingualen Chemieunterricht sind die Werte für die Motivation im Mittel bei den naturwissenschaftlich orientierten Lernern geringer als im regulären Chemieunterricht.

Abbildung 24: Operationalisierte Hypothesen zur Motivation

5.3.3.2 Theoretischer Hintergrund der Motivation: KIM

Deci und Ryan untersuchen intentionales Handeln im Hinblick auf den Grad der Selbstbestimmtheit. Im Mittelpunkt ihrer Theorie steht das "Selbst", welches drei grundlegenden psychischen Bedürfnissen unterliegt, nämlich dem Bedürfnis nach *Autonomie*, *Kompetenz* und *sozialer Einbindung* (Deci & Ryan, 1993). Lernen findet demnach aus dem Bedürfnis heraus statt, sich als autonom, kompetent und eingebunden in eine soziale Gruppe zu erleben. In Bezug auf das Autonomie-Erleben sehen die Autoren einen Übergang von der Fremd- zur Selbstbestimmtheit, je nachdem, ob die Beweggründe vom jeweiligen Individuum als internal bzw. external wahrgenommen werden:

1. Externale Regulation

Die Handlung ist vollständig fremdbestimmt und erfolgt etwa aufgrund von Belohnung oder Strafe.

2. Introjizierte Regulation

Die Handlung wird als notwendig erachtet, wenngleich sie noch nicht einer eigenen Überzeugung folgt. Dies ist z.B. der Fall, wenn der Schüler bemüht ist, gute Leistungen in der Schule zu erbringen, um den Erwartungen der Eltern gerecht zu werden.

3. Identifizierte Regulation

Die Handlung wird persönlich für wichtig erachtet, der Lerner hat sich damit identifiziert. Das Ablegen des Abiturs kann eine solche Handlung darstellen, wenn damit das Ziel verfolgt wird, ein bestimmtes Studium aufzunehmen.

4. Integrierte Regulation

Die Handlung wird als selbstbestimmt wahrgenommen, da die Handlungsziele in das eigene Wertesystem integriert wurden und zu Zielen der eigenen Persönlichkeitsentwicklung wurden.

Tabelle 19: Unterschiedliche Arten der Regulation nach Deci und Ryan

(zitiert nach Krapp, 2014, S. 203f.)

Menschen haben somit dann das Gefühl selbstbestimmt zu handeln, wenn infolge von Internalisierung und Integration von außen an sie herangetragene Ziele in das eigene Selbstsystem aufgenommen werden (Krapp & Seidel, 2014, S. 204). Der Wunsch, selbstbestimmt zu handeln, stellt für die Autoren dabei, wie bereits dargestellt, ein grundlegendes psychologisches Bedürfnis dar (Bedürfnis nach Autonomie). Damit werden die fließenden Übergänge zwischen der auf Aristoteles zurückgehenden Dichotomie von "extrinsisch und intrinsisch geleitetem Handeln" deutlich (Abendroth-Timmer, 2007, S. 32). In der Motivationspsychologie unterscheidet man dabei zwischen intrinsischer und extrinsischer Motivation, je nachdem, ob die angestrebten Ziele innerhalb oder außerhalb der Handlung liegen (Krapp & Seidel, 2014, S. 194). Intrinsische Motivation ist dadurch gekennzeichnet, dass die Handlung selbst Freude bereitet und einen Tätigkeitsanreiz darstellt. Csikszentmihalyi (1975) benutzt den Begriff *flow-Erlebnis* für Tätigkeiten, die an sich Spaß bereiten und zu einer vollständigen Selbstvergessenheit führen. Die Tätigkeit selbst stellt dann den motivationalen Reiz dar. Extrinsisch motiviert ist ein Lerner, der seine Befriedigung aus den Begleitumständen des Lernens schöpft (z.B. gute Noten oder Lob). Es wird davon ausgegangen, dass der intrinsischen Motivation eine größere Bedeutung für die Induzierung von Lernprozessen zukommt. Sie bewirkt eine tiefere Informationsverarbeitung und

geht mit Interessen und stark affektiven Komponenten einher (Schiefele & Schreyer, 1994). Im Gegensatz zur extrinsischen Motivation wirkt sie zudem langfristiger, da intrinsisch motivierte Lerner nicht ständig neue externe Anreize benötigen. Da intrinsische Motivation korrumpiert werden kann, wenn etwa durch benotete Leistungsüberprüfung negative, lernschädliche affektive Reaktionen hervorgerufen werden können (vgl. Deci & Ryan, 1993, S. 234 f.), rät Abendroth-Timmer dazu (2007, S. 34), intrinsische Motivation zu fördern "indem externe Kontrolle reduziert und Autonomie und kooperatives Lernen verstärkt werden". Die vier Stufen von Deci und Ryan beziehen sich zwar auf den Bereich der extrinsischen Motivation, beschreiben aber durch ihr wachsendes Maß an Selbstbestimmtheit in Wirklichkeit den Übergang zwischen extrinsischer und intrinsischer Motivation.

Situationales Interesse

Das *Interesse* beschreibt das Verhältnis einer Person zu einem Inhalt bzw. zu einem Lerngegenstand. Die Interessentheorie wird deshalb auch Person-Gegenstand-Theorie genannt (*POI* = *person-object theory of interest*) (Krapp & Seidel 2014, S. 205). Interessen können sich dabei kurzfristig auf einen bestimmten Lerngegenstand beziehen oder längerfristig andauern und sich in einer gewissen Interessiertheit an einem bestimmten Unterricht äußern (Abendroth-Timmer, 2007, S. 59). Man unterscheidet deshalb zwischen einem situationalem und einem dispositionalem Interesse. Situationales Interesse ist für Deci und Ryan ein Faktor, der intrinsische Motivation bedingt. Es geht somit um das situationale Interesse wie es in der vorliegenden Untersuchung durch das Modul hervorgerufen wird. Interesse hat eine gefühlsbezogene und eine wertbezogene Komponente. Das Interesse am Unterricht in einem bestimmten Schulfach beruht in der Regel auf einer hohen subjektiven Wertschätzung für das Fach sowie auf einer positiven Gefühlsausprägung. Interessen bilden sich somit dann aus, wenn die entsprechenden Inhalte und Tätigkeiten als hinreichend bedeutsam erachtet werden und mit positiven Emotionen besetzt sind. Darunter versteht man "Gefühlsregungen, die relativ klar bestimmbar sind und sich meist auf einen Auslöser zurückführen lassen" (Krapp & Seidel, 2014, S. 196). Emotionen können sich sehr unterschiedlich auf die Lernleistung aber auch die Lernbereitschaft auswirken. Sie treten immer dann auf, wenn etwas für die betreffende Person von großer Bedeutung ist. Emotionen ermöglichen kognitive Entscheidungsprozesse, die dann in konkretes Handeln umgesetzt werden können. Apelt/Koernig (1994, S. 163) bezeichnen sie als Regulatoren des Handelns, während Emotionen für Kieweg (2003, S. 6) gar die "treibende Kraft für zielgerichtetes Handeln" sind (zitiert nach Abendroth-Timmer, 2007, S. 65).

Für Deci und Ryan ist situationales Interesse eng mit dem Gefühl der Freude und der Zufriedenheit verbunden. Körperlich-seelisches Wohlbefinden setzt außerdem die Erfüllung der Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit voraus. Lernumgebungen, die dieses Erleben ermöglichen, gelten somit als besonders förderlich zur Ausbildung selbstbestimmten und interessengeleiteten Lernens. In der schulischen Praxis sind damit z.B. Forderungen nach einer optimalen Passung des Schwierigkeitsniveaus von Lernaufgaben (Kompetenzerleben), Einsatz von Lernmethoden, die ein hohes Maß an Eigenaktivität und Selbstbestimmung erlauben (Autonomieerleben) sowie kooperativen Lernformen (Erleben sozialer Eingebundenheit) verbunden (Schiefele U. , 2004). Neue Interessen können über Prozesse der Internalisierung und der Identifikation in das persönliche Wertesystem übernommen werden, so dass sich ein permanentes Interesse etwa am Unterrichtsfach entwickeln kann.

Kompetenzerleben

Sich selbst als kompetent zu erleben und den Anforderungen, welche die Umwelt an eine Person stellt, gewachsen zu sein, zählt zu den Grundbedürfnissen der Menschen. Menschen spüren Freude und Zufriedenheit, wenn sie sich selbst als wirksam erleben, sie sind dadurch intrinsisch motiviert. Laut Deci und Ryan (2000, S. 253) ist diese kompetenzbedingte Motivation unabhängig von spezifischen Inhalten, sie ergibt sich schlicht aus der Freude über den aus eigener Kraft herbeigeführten Erfolg. Sie führt dazu, dass Individuen sich immer neuen Herausforderungen und unbekannten Situationen stellen. Entwicklungspsychologisch hat dies dazu geführt, dass die Menschen sich immer besser an die sich verändernde Umwelt anpassen konnten (Deci und Ryan, 2000, S. 253). Empfundene Kompetenz kann zwar durch positives oder negatives Feedback beeinflusst werden, Voraussetzung ist aber, dass Schüler sich selbst als verantwortlich für eine kompetente Handlung fühlen (Deci und Ryan, 2000, S. 234f.). Unterricht, der so gestaltet ist, dass er Schülern immer wieder ermöglicht, ihre eigene Kompetenz zu erleben, kann somit intrinsisch motivieren und langfristig das Interesse der Schüler für das Fach wecken.

Druck

Anspannung und Druck gehören nach Deci und Ryan zu den negativen Prädiktoren intrinsischer Motivation. Laut Krapp (2014, S. 213) führen negative Emotionen wie Stress, Nervosität und Angst insbesondere bei misserfolgs-orientierten Lernern dazu, dass die Aufmerksamkeit von der Bearbeitung der jeweiligen Aufgabe abgelenkt wird. Der Mangel an Konzentration und der sich daraus ergebende defizitäre Einsatz von Lernstrategien erweisen sich dann insbesondere bei anspruchsvollen Aufgaben als leistungsmindernd. Als Folge davon entsteht unter Umständen der

Wunsch, solche angstbesetzten Situationen zu meiden, in denen Misserfolg nicht ausgeschlossen wird. Oft bedeutet das, dass sich die Schüler erhöhten Anforderungen erst gar nicht stellen möchten. Ist dies nicht möglich, so können Anspannung oder Angst insbesondere bei erfolgsorientierten Lernern (Krapp, 2014, S. 214) jedoch auch zu verstärkten Lernanstrengungen führen, die sich lernfördernd auswirken. Nach dem *Yerkes-Dodson-Gesetz* wird davon ausgegangen, dass es in Abhängigkeit vom Schwierigkeitsgrad der gestellten Aufgabe ein optimales Erregungsniveau für das Leistungsverhalten gibt (Zimbardo & Gerrig, 2004). Dieses ist umso geringer, je schwieriger die zu bewältigende Aufgabe ist. Preisfeld (2016, S. 111) stellt auf Grund neurobiologischer Erkenntnisse aus der Gehirnforschung fest, dass stressreiche Situationen aufgrund der hohen Konzentration an Stresshormonen die Aufnahme und Verarbeitung von Informationen behindern. Sanfter Stress (Eustress) kann jedoch aufgrund der nur geringen Ausschüttung solcher Hormone durchaus positive Effekte auf das Lernen haben, etwa, wenn Lernsituationen als Herausforderung betrachtet werden.

Nutzen

Der Wert bzw. *Nutzen* des Gelernten und seine Bedeutung für die Lernmotivation wird in der Motivationspsychologie im Zusammenhang mit den sekundären Lernmotiven diskutiert. Im engeren Sinne wird darunter der Gebrauchswert des Gelernten im Hinblick auf eine spätere berufliche Karriere verstanden. Diese beruflich ausgerichtete Form des Nützlichkeitsmotivs spielt vor allem bei älteren Schülern und insbesondere bei Berufsschülern eine zentrale Rolle, die bereits konkretere Vorstellungen über ihren zukünftigen Beruf haben. Im weiteren Sinne ist das Nützlichkeitsmotiv ein eher allgemein ausgerichtetes, wobei der Nutzen des Gelernten je nach persönlichen Lebensumständen und Bedürfnissen des Einzelnen ganz unterschiedlich wahrgenommen werden kann. Das Nützlichkeitsmotiv gehört damit zu den klassischen Faktoren der extrinsischen Motivation. Deci und Ryan (2003) konnten jedoch zeigen, dass durch Internalisierung der Nützlichkeitsaspekt Teil des individuellen Wertesystems des Lerners wird und das Lernen selbst dann von diesem als wertvoll erachtet wird.

5.3.3.3 Aufbau der Fragebögen

Mit Hilfe der Fragebögen werden Daten zur Motivation und zur Demografie der Probanden erhoben. Die demografischen Angaben dienen vor allem zur Beschreibung der Stichproben, die in Bezug auf verschiedene Kontrollvariablen vergleichbar sein sollen. Dabei werden im Fragebogen Daten zu Alter, Geschlecht, Migrationshintergrund, Bildungsniveau des Elternhauses sowie Inanspruchnahme von Nachhilfeunterricht im Fach Chemie erhoben.

Zusätzlich werden in den Fragebögen beider Gruppen Angaben zum Fachinteresse und zum akademischen Selbstkonzept in den Fächern Chemie und Englisch ermittelt. Diese werden später im Hinblick auf ihren moderierenden Einfluss auf die sachfachliche, (fremd-)sprachliche und motivationale Wirksamkeit des Moduls hin untersucht. In einem weiteren Item vergleichen die Schüler ihr Interesse für Naturwissenschaften und Fremdsprachen. Aufgrund ihrer Angaben werden sie dann in Interessensgruppen (Orientierung) eingeteilt (vgl. Abschnitt 5.2.4). Die Wirksamkeit bilingualen Unterrichts auf sachfachlicher, (fremd-)sprachlicher und motivationaler Ebene erfolgt später differenziert nach Interessensgruppen. Sowohl in der Untersuchungs- als auch in der Kontrollgruppe werden zwei unterschiedliche Fragebögen eingesetzt, der erste (FB1) unmittelbar vor Beginn des Moduls der zweite (FB2) sofort im Anschluss. Jeder Fragebogen besteht aus zwei Teilen, die Items zur Motivation (Teil 1) und zu demografischen Angaben (Teil 2) enthalten. Dazu kommt das Deckblatt, auf welchem die Schüler ihren Code-Namen notieren. Sowohl in der Untersuchungs- wie auch in der Kontrollgruppe beziehen sich die Fragen zur Motivation im Prefragebogen (FB1) auf die Erwartungen der Schüler in Bezug auf das bevorstehende Modul. Im Postfragebogen (FB2) bezieht sich ihre Einschätzung auf das gerade durchgeführte bi- bzw. monolinguale Modul.

Prefragebogen UG (FB1)	
Deckblatt	Erzeugen des Code-Namens
Teil 1	Einstellung der Schüler
	Die Fächer: Fachinteresse und Selbstkonzept
	Fachinteresse Chemie
	Fachinteresse Englisch
	Selbstkonzept Chemie
	Selbstkonzept Englisch
	Der bilinguale Unterricht: Intrinsische Motivation
	Interesse am Unterricht
	Nutzen
	Kompetenzerleben
	Druck
Teil 2	Demografische Angaben
	Geschlecht
	Alter
	Chemienote
	Englischnote
	Interessenschwerpunkt / Orientierung

Tabelle 20: Aufbau des Prefragebogens der CLIL-Schüler

Postfragebogen UG (FB2)	
Deckblatt	Erzeugen des Code-Namens
Teil 1	Einstellung der Schüler
	Der bilinguale Unterricht: Intrinsische Motivation
	Interesse am Unterricht
	Nutzen
	Kompetenzerleben
	Druck
Teil 2	Demografische Angaben
	Externe Hilfe
	Migrationshintergrund
	Akademikerhaushalt
	Schülerurteil
	Fazit zum bilingualen Unterricht

Tabelle 21: Aufbau des Postfragebogens der CLIL-Schüler

Der erste Teil des Fragebogens beginnt dabei immer mit Fragen, die nur ein geringes Maß an emotionalen Reaktionen erwarten lassen, wie z.B. Fragen zu den fachlichen Interessen sowie dem

Nutzen bestimmter Inhalte. Persönliche und vertrauliche Informationen, deren Preisgabe womöglich auch im Bewusstsein der Wahrung der Anonymität eine gewisse Überwindung kostet (z.B.: nach Schulnoten oder dem möglicherweise erteilten Nachhilfeunterricht), folgen erst später. Damit soll vermieden werden, dass die Schüler durch bestimmte Fragen frühzeitig abgeschreckt werden, offen Auskunft zu geben. Im zweiten Teil des Fragebogens werden neben demografischen Angaben auch weitere Variablen zur Beschreibung der Stichproben erfasst. Dabei kommen unterschiedliche Formate, wie Fragen mit Alternativform (Akademikerhaushalt und Geschlecht) oder Fragen nach Zahlenangaben (Alter) zum Einsatz. Im Postfragebogen der bilingualen Schüler wurde außerdem um ein Fazit zum bilingualen Unterricht gebeten. Die Regelschüler waren aufgefordert, das Modul mit dem normalen Chemieunterricht zu vergleichen und mögliche Unterschiede zu beschreiben. Hierzu beantworteten die Schüler jeweils eine offene Frage. Diese wurde bewusst ans Ende des Fragebogens gesetzt, damit die Schüler sich genügend Zeit nehmen können, sie möglichst ausführlich zu beantworten, ohne unter dem Druck zu stehen, noch weitere Fragen beantworten zu müssen. Die Daten wurden ausschließlich für die Pilotgruppe ausgewertet. Bei der formalen Gestaltung der Fragebögen wurde vor allem auf Bearbeitungsfreundlichkeit Wert gelegt⁵⁰. Zur besseren Übersichtlichkeit sind beide Themenblöcke Motivation und Arbeitsverhalten deshalb optisch durch eine Überschrift getrennt, und die zu bearbeitenden Items in Blöcken aus jeweils acht Items präsentiert. Die Items sind dabei so angeordnet, dass sie abwechselnd unterschiedliche Subskalen betreffen, um zu vermeiden, dass inhaltlich ähnliche Items unmittelbar aufeinander folgen, was zu einem „mechanischen“ Abarbeiten der Fragen führen kann. Kurze Bearbeitungshinweise unter Angabe eines Beispiels finden sich auf dem Deckblatt und betreffen dort die Erstellung des Code-Namens. Weitere Hinweise erfolgen zu Beginn des eigentlichen Fragebogens und betreffen dort die Bearbeitung von Rating-Items. Der Angabe der zu verwendenden Hilfsmittel erfolgt ebenfalls gut sichtbar auf dem Deckblatt.

5.3.3.4 Erstellung von Items

Die zur Erfassung der Motivation eingesetzten Items stammen teilweise aus bereits erprobten Fragebögen, wie der "Kurzskala intrinsische Motivation" (Wilde, Bätz, Kovaleva & Urhahne, 2009). Außerdem wurde auf Items aus den Studien von Abendroth-Timmer (2007), Hartmannsgruber (2014) und Lipski-Buchholz (2019) zurückgegriffen. Einige Items wurden um- bzw. neuformuliert, um sie an das vorliegende Forschungssetting des bilingualen Chemieunterrichts anzupassen. Dieses Vorgehen ist in der Motivationsforschung nicht unüblich (Dörnyei Z. , 2001, S.

50 Ausführlich Hinweise zur formalen Gestaltung von Fragebögen finden sich bei Dörnyei & Taguchi (2010, S. 12ff.).

190) Ein ähnliches Vorgehen findet man auch bei Lipski-Buchholz (2019) und Hartmannsgruber (2014). Das abgebildete Beispiel verdeutlicht die Vorgehensweise.

Items	Quelle
1. Die Tätigkeit in der Ausstellung hat mir Spaß gemacht. 2. Ich fand die Tätigkeit in der Ausstellung sehr interessant. 3. Die Tätigkeit in der Ausstellung war unterhaltsam.	"Kurzsкала intrinsische Motivation"(Wilde, Bätz, Kovaleva & Urhahne, 2009) ⁵¹
1. Der NWA-Unterricht zum Thema Wasser hat mir Spaß gemacht. 2. Ich fand den NWA-Unterricht zum Thema Wasser sehr interessant. 3. Der NWA-Unterricht zum Thema Wasser war unterhaltsam.	Für den bilingualen Biologieunterricht adaptierte Version der KIM nach Hartmannsgruber (2014)
1. CAE hat mir Spaß gemacht. 2. Ich fand CAE sehr interessant. 3. CAE war langweilig. 4. Ich fand CAE nicht interessant.	Für den bilingualen Chemieunterricht von Hülnden adaptierte Version

Abbildung 25: Konstruktion der Items zur Subskala Interesse / Vergnügen

Die Skala wurde auf vier Items erweitert, wobei zwei positiv und zwei negativ formuliert wurden. Die Erhöhung der Itemanzahl und die Verwendung negativ formulierter Items tragen laut Dörnyei (2010) dazu bei, die Reliabilität einer Skala zu erhöhen. Die Abkürzung CAE bedeutet "Chemie auf Englisch"; diese Formulierung wurde im Instruktionstext des Fragebogens erläutert und durchgängig verwendet. Den Probanden werden Fragen Rating-Skalen vorgelegt, bei denen sie bewerten müssen, wie weit eine bestimmte Aussage auf sie zutrifft. Die sogenannte Likert-Skala umfasst vier Stufen: "Stimmt völlig", "Stimmt eher", "Stimmt wenig" und "Stimmt nicht" sowie die Kategorie "Ich weiß es nicht." Eine noch feinere Abstufung mit mehr Stufen erscheint in Anbetracht des Alters der Probanden und des Umfangs des Tests unangemessen und hätte wahrscheinlich Ermüdungserscheinungen zur Folge, wodurch die Qualität der Daten leiden würde (vgl. Rumlich 2016, S. 277). Es wurde bewusst eine gerade Anzahl von Stufen gewählt, damit die Probanden sich entscheiden, ob sie der Aussage zustimmen oder sie abzulehnen. Ein routinemäßiges Ausweichen auf einen mittleren Skalenwert ist dann nicht möglich. Auch die Entscheidung, dass man es nicht zu

⁵¹ Der Fragebogen von Wilde et al. (2009) wurde ursprünglich für den Museumsbesuch entwickelt.

beurteilen vermag, ist mit Option 5 ("Ich weiß es nicht") möglich, muss aber bewusst getroffen werden. Die Probanden haben also die Freiheit, bei Items, für die sie auch nach reiflicher Überlegung zu keiner sie selbst überzeugenden Einschätzung kommen, auf eine solche Entscheidung zu verzichten. Das entsprechende Item wird dann bei der Ermittlung des entsprechenden Skalenwertes nicht berücksichtigt. Damit soll vermieden werden, dass sie sich willkürlich für ein Item entscheiden. Zur Erhöhung der Reliabilität werden zur Erfassung eines Merkmals nicht einzelne Items, sondern *Multi-Item-Skalen* verwendet, die aus vier Items bestehen. Mehr als vier Items sind im vorliegenden Fall nicht sinnvoll, da sie eine deutliche Erhöhung der Bearbeitungszeit zur Folge hätten. Da die Schüler neben den Fragebögen auch sachfachliche bzw. fremdsprachliche Tests absolvieren, wurde darauf geachtet, dass die Fragebögen in maximal zehn Minuten bearbeitet werden können. Auch Dörnyei (2001) weist darauf hin, dass die Qualität der Daten leiden kann, wenn die Fragebögen zu umfangreich sind, und empfiehlt die Verwendung von drei bis vier Items zur Erfassung einer Skala. Die Items variieren in ihrer Wortwahl, was dazu beitragen soll, die Zahl von Fehlern aufgrund von Missverständnissen zu verringern. Solche Missverständnisse können sich aufgrund des subjektiven Verständnisses der Probanden auch bei sorgfältig formulierten Items ergeben. Die Items werden den Schülern unsortiert vorgelegt. Dadurch soll ein einseitiges Ankreuzverhalten vermieden werden (Dörnyei & Taguchi, 2010, S. 25).

5.3.3.5 Auswertung der Ergebnisse

Die numerische Codierung erfolgte mit Werten von 4 bis 1 von "Stimmt völlig", "Stimmt eher", "Stimmt wenig" bis "Stimmt nicht". Unbearbeitete Items sowie die Antwort "Ich weiß es nicht" wurden mit 0 codiert. Für negativ gepolte bzw. invertierte Items, also solche, die nicht in Schlüsselrichtung formuliert sind, erfolgte die Codierung gegenläufig, also von 1 ("Stimmt völlig") bis 4 ("Stimmt nicht"). Dieses Vorgehen schlägt auch Bühner (2011, S: 217f.) vor. Es soll im Folgenden am Beispiel eines Items der Skala „Interesse bilingualer Unterricht“ erläutert werden.

Item IBU 4				
„CAE war langweilig.“	Stimmt nicht	Stimmt wenig	Stimmt eher	Stimmt völlig
Codierung	1	2	3	4
Umcodierung	4	3	2	1

Abbildung 26: Umcodierung eines negativ gepolten Items

Schüler mit großem Interesse am bilingualen Unterricht in Chemie würden der Aussage „CAE war langweilig“ sicher nicht zustimmen. Stimmt also jemand der Aussage nicht zu, so wird die Antwort

mit „4“ und nicht mit „1“ codiert. Der Wert „4“ bedeutet jeweils, dass ein Merkmal besonders stark ausgeprägt ist. Hier bedeutet es, dass ein großes Interesse am bilingualen Chemieunterricht vorliegt, weil der Aussage „CAE war langweilig“ nicht zugestimmt wurde. Für die Auswertung per Hand wurden Auswertungsschablonen angefertigt. Diese Folien waren transparent und wurden über den jeweiligen Antwortbogen gelegt. So konnte die Codierung direkt abgelesen und auf dem Antwortbogen vermerkt werden. Negativ gepolte Items waren auf diesen Schablonen bereits umcodiert, so dass die umcodierten Werte anschließend direkt in die entsprechende Excel-Tabelle übernommen werden konnten. Für jede Skala wurden Mittelwerte über die Anzahl der beantworteten Items berechnet. Konnten von einem Probanden weniger als drei Items einer Skala verwendet werden, so wurde der entsprechende Wert als zu wenig reliabel betrachtet und nicht aufgenommen. Dies war z.B. dann der Fall, wenn Items von Probanden übersehen, mehrfach bewertet oder mit "Ich weiß es nicht" markiert wurden. Anhand der Items wurden jedem Merkmal vier Ausprägungsstufen zugeordnet, welche Punkten von 1 (nicht ausgeprägt) bis 4 (stark ausgeprägt) entsprachen. So ergaben sich je Merkmal zwischen 4 und 16 Punkten. Da nicht alle Probanden immer alle vier Items einer Subskala bearbeitet hatten, wurde der Mittelwert über die Zahl der tatsächlich bearbeiteten Items der Skala gebildet. Dieser Wert lag wieder zwischen 1 und 4. Eine mittlere Merkmalsausprägung liegt beim Wert 2,5 vor und nicht bei dem Wert 2. Werte von 2 und weniger bedeuten, dass das entsprechende Merkmal wenig ausgeprägt ist ("Stimmt wenig" bzw. "Stimmt nicht"). Erst drei Punkte bedeuten eine Zustimmung des Schülers, da er dann im Mittel "stimmt eher" ankreuzt. Je nach Untersuchungsfrage wurde dann unterschiedlich mit den so erhaltenen Daten verfahren. Sollte die Motivation unterschiedlicher Schülergruppen verglichen werden, so wurden Gruppenmittelwerte errechnet und auf statistisch signifikante Unterschiede getestet. Ging es darum, den Einfluss der Motivation auf andere Variablen zu ermitteln, so wurden anhand der Probandenmittelwerte Korrelations- bzw. Regressionsanalysen durchgeführt (vgl. Abschnitt 7.5).

5.4 Erfassung weiterer Kontrollvariablen

Weil die Probanden für die vorliegende Untersuchung aus sechs Klassen kommen, die vollständig entweder der Untersuchungs- oder der Kontrollgruppe zugeordnet wurden, ist eine Randomisierung der Stichproben nicht gegeben. Damit erhöht sich die Wahrscheinlichkeit, dass Unterschiede im Lernerfolg oder in der Motivation auf die unterschiedliche Zusammensetzung der Gruppen zurückzuführen sind. Um dennoch eine gewisse Vergleichbarkeit beider Stichproben zu gewährleisten, werden diese im Hinblick auf bestimmte Störvariablen verglichen. Dies geschieht mit Hilfe von Fragebögen für demografische Merkmale (Abschnitt 5.3.3.3), Fachinteresse

(Abschnitt 5.4.5) und akademisches Selbstkonzept (Abschnitt 5.4.5) sowie anhand weiterer Tests. Im folgenden Abschnitt werden zunächst diese Tests vorgestellt. Dabei handelt es sich um Tests zu kognitiven Fähigkeiten (Abschnitt 5.4.1), zur Lesekompetenz (Abschnitt 5.4.2), zum Leistungsvermögen in Englisch (Abschnitt 5.4.3) sowie zum Leistungsvermögen in Chemie (Abschnitt 5.4.4). Abschließend werden die Konstrukte Fachinteresse und Selbstkonzept erläutert und ihre Erfassung mittels Rating-Skalen beschrieben (Abschnitt 5.4.5).

5.4.1 Kognitive Fähigkeiten

Intelligenz und schulischer Erfolg sind eng miteinander verbunden. Artelt und Wirth (Krapp & Seidel et al., 2014, S. 181) sehen die Vorteile intelligenter Schüler darin, dass sie sich „schneller auf Aufgaben einstellen können, über effektivere Problemlösestrategien verfügen, leichter lösungsrelevante Regeln erkennen und über größere Verarbeitungskapazität und elaboriertere Lern- und Gedächtnisstrategien verfügen“. Insbesondere bei jüngeren Schülern ist die Intelligenz deshalb ein guter Prädiktor für schulischen Lernerfolg (Köller & Baumert, 2008). Der Einfluss kognitiver Fähigkeiten auf die Leistungen in den naturwissenschaftlichen Fächern (Steinkamp & Maehr, 1983) sowie in den Fremdsprachen (Dallinger, 2015; Nold, 2003) konnte bereits mehrfach auch empirisch nachgewiesen werden. Intelligenz ist im Gegensatz zu domänspezifischen Kompetenzen, die in bestimmten Situationen mit bestimmten Anforderungen erworben werden, nicht lernbar (vgl. Ropohl, 2010, S. 50). Sie gilt als kontextunabhängig und zeitlich stabil (Hartig & Klieme, 2006, S. 131). Bei Intelligenz handelt es sich um ein vielschichtiges Konstrukt und so existieren unterschiedlicher Modelle, welche die einzelnen Facetten abbilden (vgl. Artelt & Wirth, 2014, S. 181). Dem hier verwendeten kognitiven Fähigkeitstest liegt die Unterscheidung zwischen verbaler, quantitativer und figuraler Intelligenz zugrunde. Er wurde auf der Basis des *Lorge-Thorndike Intelligence Tests* entwickelt und dient der Bestimmung kognitiver Fähigkeiten, die insbesondere für schulisches Lernen relevant sind (Heller und Perleth, 2000, S. 3). Es wird allgemein angenommen, dass insbesondere verbale kognitive Fähigkeiten einen starken Einfluss auf das Erlernen einer Fremdsprache haben (vgl. Rumlich, 2016). Im Rahmen der DESI-Studie konnte jedoch nachgewiesen werden, dass auch non-verbale und figurative kognitive Fähigkeiten die Leistungen in der Fremdsprache beeinflussen (Helmke, Schrader, Wagner, Nold, & Schröder, 2008). In Studien zum naturwissenschaftlichen Leistungsvermögen werden oft non-verbale kognitive Fähigkeiten erhoben (vgl. z.B. Ferber, 2014, S. 36; Ropohl, 2010, S.36). Für die vorliegende Arbeit sollen sowohl verbale als auch non-verbale kognitive Fähigkeiten erfasst werden. Dabei kommt eine Kurzform des kognitiven Fähigkeitstests (KFT) für Schülerinnen und Schüler von Heller und Perleth (2000) zum Einsatz. Die psychometrische Qualität des Tests wurde bereits in zahlreichen

statistischen Untersuchungen nachgewiesen (Heller und Perleth, 2000). Im Kontext des bilingualen Unterrichts wurde der kognitive Fähigkeitstest unter anderem von Dallinger (2015), Fehling (2008), Piesche (2016) und Rumlich (2016) eingesetzt. Der Test wurde für Klasse 8 in der verkürzten Version verwendet, der die drei Subskalen zur Messung verbaler, quantitativer und figuraler Fähigkeiten umfasst. Die gekürzte Version wurde gewählt, um die Gesamtbelastung der Schüler durch die Tests möglichst gering zu halten. Für die vorliegende Untersuchung wurden von den insgesamt sechs Subtests nur je einer pro Skala bestimmt. Damit erfasst der Kurztest alle drei Subskalen des kognitiven Fähigkeitstests nach Heller und Perleth (2000) und zwar über die Tests Q1 (quantitativer Teil), V3 (verbaler Teil) und N1 (nonverbaler Teil). Die gleiche Kurzform kam auch bei der Entwicklung und Evaluation von Testaufgaben zur Erfassung der Kompetenz „Umgang mit Fachwissen“ im Rahmen des ESNaS-Projektes zur Anwendung. Ropohl (2010, S. 97) konnte dabei eine hohe Korrelation zwischen den dort entwickelten Aufgaben und der kognitiven Fähigkeit beobachten. Beim Test V3 zur verbalen Intelligenz handelt es sich um einen Test zum Erkennen von Wortanalogien. Er umfasst 20 Aufgaben, wobei die Schüler aus fünf möglichen Antworten jenes Wort finden sollen, das zu einem vorgegebenen Wort in einer bestimmten semantischen Beziehung steht. Die Art dieser Beziehung wird jeweils anhand eines Beispielpaares vorgegeben. Der Test Q1 erfasst die quantitative Dimension kognitiver Fähigkeiten und enthält 25 Aufgaben. Die Schüler müssen dazu beurteilen, in welcher Relation – kleiner, größer oder gleich groß – zwei Mengen zueinanderstehen. Diese umfassen Punkte, Linien, Zahlen, einfache mathematische Terme, geometrische Angaben zu Fläche, Streckenlänge, Winkel und Rauminhalt sowie unterschiedliche Größeneinheiten (z.B. L und mL; cm^2 und m^2). Es gibt jeweils drei Antwortoptionen, wobei die Schüler sich durch Ankreuzen für eine Antwort entscheiden. Die Test N1 erfasst Fähigkeiten im Bereich der nonverbalen Intelligenz. Dabei müssen die Probanden zu drei oder vier vorgegebenen Figuren das gemeinsame Merkmal erkennen und aus fünf gegebenen Alternativen die Figur auswählen, die das gleiche Merkmal aufweist. Die Merkmale, die in den insgesamt 25 Aufgaben vorkommen, betreffen Form, Farbe, Schraffur oder Lage der Figuren. Der KF-Test umfasst somit insgesamt 70 Aufgaben, die in 39 Minuten bearbeitet werden müssen. Inklusiv der Testeinführung von 4:00 Minuten ist die Bearbeitung somit in einer Schulstunde ermöglicht.⁵²

52 Für die einzelnen Testteile werden 9:30 Minuten (V3), 13:30 Minuten (Q1) und 12:00 Minuten (N1) benötigt.

5.4.2 Lesekompetenz

In zahlreichen Studien konnte der Einfluss der Lesekompetenz auf die Testleistung in mathematisch-naturwissenschaftlichen Tests nachgewiesen werden. Sowohl in der PISA-2000- (Baumert et al., 2001) als auch in der PISA-2003-Studie (Leutner et al., 2004) konnten hohe latente Korrelationen zwischen Lesekompetenz und naturwissenschaftlicher Kompetenz festgestellt werden. Dies ist verständlich, da zur Beantwortung von Testaufgaben zuerst immer Informationen aus einem Begleittext entnommen werden müssen (vgl. Ropohl, 2010, S. 51). Auch für die hier vorgelegten sachfachlichen Tests (CL-Test und ACL-Test) werden Lesekompetenzen benötigt. Sie spielen insbesondere bei der Bearbeitung des zweiten Teils der Tests eine Rolle, bei denen es um die Anwendung und den Transfer des chemischen Fachwissens geht. Das hierzu benötigte Fachwissen wird vorab in einem längeren Informationstext bereitgestellt, auf den die Schüler immer wieder bei der Bearbeitung der Aufgaben zurückgreifen, um bestimmte Information zu finden (siehe Abschnitt 5.3.1.2)⁵³. Das Leseverstehen stellt einen aktiven Konstruktionsprozess dar, bei dem der Leser Textinhalte entnehmen, mit den umgebenden Informationen und seinem Weltwissen in Verbindung setzen muss (vgl. Artelt et al., 2001; Lenhard et al., 2017). Dabei unterscheiden sich Schüler deutlich in ihrer Lesefähigkeit, da die zugrunde liegenden Prozesse individuell unterschiedlich stark automatisiert sind (Schneider et al., 2017, S. 11). Je schneller Schüler die relevanten Informationen aus einem Text aufspüren und verstehen, desto besser werden sie wahrscheinlich in den Chemietests abschneiden. In der vorliegenden Untersuchung kommt der *Lesegeschwindigkeits- und -verständnistest für die Klassen 5-12* von Schneider et al. (2017) zur Anwendung. Der Test ist vielfach als Gruppentest im Schulbereich erprobt und gilt als zuverlässig. Auch zur Beurteilung der Ergebnisse im Chemieleistungstest wurde der LGVT 5-12+ bereits erfolgreich eingesetzt (Ferber 2014). Dem Test liegt die Annahme zugrunde, dass sich Lesekompetenz aus Lesegeschwindigkeit und Leseverständnis zusammensetzt. Außerdem erfasst der Test in seiner Neuauflage von 2017 auch die sogenannte Lesegenauigkeit, welche „eine Aussage darüber erlaubt, unter welchem Verhältnis von Geschwindigkeit und Genauigkeit eine bestimmte Leseverständnisleistung erzielt wurde“ (Schneider et al., 2017, S.15). Nachdem sie ein Übungsbeispiel bearbeitet haben, lesen die Schüler einen Fließtext und entscheiden an 47 Stellen durch Unterstreichen, welches von drei gegebenen Wörtern am besten in den Textzusammenhang passt. Für jedes korrekt unterstrichene Wort werden

53 In der vorliegenden Untersuchung wurde versucht, den Einfluss der Lesekompetenz zu verringern, indem die nötigen Fachinformationen nicht im Aufgabenstamm der einzelnen Aufgaben gegeben wurden, sondern als Zusammenfassung vor Bearbeitung aller Testaufgaben vorlagen (vgl. Abschnitt 5.3.1.2). Dadurch gelang es, den Umfang der zu lesenden Texte deutlich zu reduzieren.

dabei zwei Punkte vergeben. Die Ratekorrektur erfolgt, indem für falsch angegebene Wörter ein Punkt abgezogen wird (Leseverständnis). Die insgesamt gelesene Wortzahl wird ermittelt, indem die Wörter bis zur letzten bearbeiteten Lücke gezählt werden (Lesegeschwindigkeit). Die Lesegenauigkeit berechnet sich schließlich als prozentualer Anteil der richtig gelösten Aufgaben. Der Test wird als Gruppentest durchgeführt mit einer Bearbeitungszeit von 12 Minuten (inklusive Instruktionszeit). Es werden zwei Paralleltexte eingesetzt (Rosenkohltext: 1989 Wörter und Laufburschentext: 2155 Wörter), um das Abschreiben vom Nachbarn möglichst zu verhindern.

5.4.3 Allgemeine Fremdsprachenkompetenz

Der Einfluss fremdsprachlicher Fähigkeiten auf den Lernerfolg im bilingualen Unterricht wurde in der Forschung intensiv diskutiert. Im Allgemeinen wird davon ausgegangen, dass sich hohe Fremdsprachenkompetenzen positiv auf die sachfachlichen Leistungen auswirken (vgl. Marsh et al., 2000; Zydati, 2007). In Anlehnung an die *double threshold hypothesis* von Cummings (1978) wird in neueren Forschungen zu modularen Varianten von CLIL auch der Frage nachgegangen, über welches Mindestniveau an Fremdsprachenkompetenzen Schüler verfügen müssen, um erfolgreich am CLIL-Unterricht teilnehmen zu können (vgl. Abschnitt 4.1.4). Für die vorliegende Untersuchung ist die allgemeine Fremdsprachenkompetenz deshalb in zweifacher Hinsicht von Bedeutung. Zum einen als Störvariable, weil der Vergleich sachfachlicher Leistungen von CLIL- und nicht-CLIL-Schülern nur dann verallgemeinerbare Schlüsse zulässt, wenn die Gruppen auch im Hinblick auf ihre Fremdsprachenkompetenz vergleichbar sind. Zum anderen, weil der Frage nach einem Mindestniveau fremdsprachlicher Kompetenz hier unmittelbar nachgegangen werden soll, ab dem CLIL-Schüler im naturwissenschaftlichen Anfangsunterricht vergleichbare sachfachliche Leistungen erbringen können wie nicht-CLIL-Schüler. Zur Bestimmung der allgemeinen Fremdsprachenkompetenz kommen in dieser Studie C-Tests zum Einsatz. C-Tests haben sich als Instrumente zur globalen Sprachstandsfeststellung, die unabhängig vom vorangegangenen Unterricht bzw. der individuellen Lerngeschichte sind, bewährt und werden oft als Einstufungstests etwa an Universitäten verwendet (Grotjahn R., 2002, S. 211). Auch wenn C-Tests nur geringe Augenscheinvalidität besitzen, sind gute C-Tests nach Grotjahn (2002, S. 211) ein zuverlässiger Prädiktor für allgemeinsprachliche Kompetenz. Dies zeigen auch Untersuchungen, in denen eine hohe Korrelation der C-Test-Ergebnisse mit den Ergebnissen des TOEFL-Tests (Grotjahn, 2011, S. 135) ermittelt wurde. Hastings (2002, S. 66) erklärt die Anforderungen eines C-Tests folgendermaßen:

„A C-test measures the ability to apply and integrate contextual, semantic, syntactic, morphological, lexical and orthographic information and knowledge pertaining to a particular written language.“

C-Tests eignen sich insbesondere für empirische Untersuchungen, weil sie objektive und hoch reliable Instrumente sind, die leicht auszuwerten sind und somit eine ökonomische Messung erlauben (Rumlich, 2012, S. 175). In vielen großangelegten Studien, in denen die Leistungen in der Fremdsprache erfasst werden wie z.B. KESS 7, COMBIH, LAU und DESI, wird deshalb auf C-Tests zurückgegriffen (vgl. Bos et al., 2009; Dallinger, 2015; Lehmann et al., 2011). C-Tests stellen eine besondere Form der *Cloze-Tests* dar, wobei etwa jedes vierte Wort eines Textes zur Hälfte gelöscht ist und ergänzt werden muss. In klassischen C-Tests kommen in der Regel vier bis sechs thematisch unterschiedliche, kurze Texte von etwa gleicher Länge und gleicher Zahl an Lückenwörtern vor. Die in der vorliegenden Studie verwendeten C-Tests der Hamburger Behörde für Schule, Jugend und Berufsbildung wurden für Schüler der Klassen 7 und 8 konzipiert und wurden unter anderem im Rahmen von Schulleistungstests, wie der KESS-Studie (Bos et al., 2009) oder der LAU-Studie (Lehmann, Peek, Gänsfuß, & Husfeldt, 2011), eingesetzt und können daher als gut erprobt gelten. Rumlich (2016) greift in seiner Untersuchung zur allgemeinen Fremdsprachenkompetenz bilingualer Lerner (DENOCS) für Schüler der Klasse 8 ebenfalls auf vier C-Tests dieser Reihe zurück und konnte deren gute psychometrische Eigenschaften nachweisen (Rumlich, 2016, S. 269ff.). Für diese Arbeit wurden die gleichen vier C-Tests ausgewählt. Diese enthalten zusammen insgesamt 97 Lückenwörter. In den einzelnen Tests sind jeweils 24 bzw. 25 Lücken zu ergänzen, wofür die Schüler jeweils fünf Minuten Zeit haben. Damit ergibt sich eine maximale Bearbeitungszeit von 20 Minuten, was für Schüler der Klasse 8 durchaus angemessen ist. Für jede richtige Ergänzung wurde ein Punkt vergeben, wobei auch Rechtschreibfehler zu einem Punktabzug führten. Grotjahn (1995, S. 43) schlägt dieses Verfahren vor, um die Auswertungsobjektivität zu erhöhen und Unklarheiten zu vermeiden, was als Fehler zählt und was nicht. Die vier C-Tests wurden in unterschiedlicher Reihenfolge, einer nach dem anderen an die Schüler ausgeteilt, so dass nebeneinandersitzende Schüler nicht voneinander abschreiben können⁵⁴.

54 Testform A enthielt die vier C-Tests in folgender Reihenfolge: *Bike Tour*, *Crossing the Channel*, *The Underground* und *Crime Prevention*. Bei Testform B war die Abfolge: *Crime Prevention*, *The Underground*, *Crossing the Channel* und *Bike Tour*.

5.4.4 Allgemeine Leistungsfähigkeit in Chemie

Der Einfluss sachfachlicher Kompetenzen auf das Lernen im bilingualen Unterricht wurde zuletzt verstärkt in den Blick genommen (vgl. Badertscher und Bieri, 2009). Dabei beschränken sich diese Ansätze jedoch auf das deklarative Fachwissen der Schüler, weil sich dies als ein starker Prädiktor für Lernleistungen erwiesen hat. In der konstruktivistischen Theorie des Lernens wird dieser Befund darauf zurückgeführt, dass neues Wissen sich umso leichter in bereits vorhandene Wissensstrukturen einbauen lässt, je differenzierter diese sind und je mehr Anknüpfungspunkte sich deshalb ergeben (Renkl A. , 1996). Hier wird das sachfachliche Leistungsvermögen mit Hilfe eines allgemeinen Chemie-Leistungstests (ACLT) erfasst. Der Test berücksichtigt neben dem chemischen Wissen (AFB I) auch höhere kognitive Fähigkeiten, welche die Anwendung und den Transfer dieses Wissens betreffen (AFB II/III). Der Test soll Leistungen in Bezug auf den Kompetenzbereich „Umgang mit chemischem Fachwissen“ erfassen. Er wurde aus diesem Grund nach den gleichen Prinzipien wie die Leistungstests zum Modul konstruiert. Dem ACL-Test liegt ebenfalls das ESNaS-Kompetenzmodell zugrunde, das zur Erfassung der verschiedenen Kompetenzstufen Fähigkeiten beschreibt, die unterschiedliche kognitive Leistungen erfordern (Walpuski, et al., 2010, S. 177). Die Testaufgaben wurden deshalb nach den entsprechenden Vorgaben konstruiert⁵⁵. Auch der Umfang des Tests entspricht dem des Chemietests zum Modul, die Verteilung der insgesamt 57 Testpunkte auf die drei Anforderungsbereiche war ebenfalls analog. Durchführung und Bewertung erfolgten nach den gleichen Kriterien. Anders als der Test zum Modul, beschreibt der ACL-Test jedoch das Leistungsvermögen der Schüler vor Modulbeginn und bezieht sich deshalb auf Inhalte aus Klasse 7 und Klasse 8.1. Um eine möglichst breite Streuung der Inhalte zu gewährleisten, werden alle vier in Klasse 7 unterrichteten Inhaltsfelder, nämlich *Stoffe und Stoffveränderung*, *Stoff- und Energieumsätze bei chemischen Reaktionen*, *Luft und Wasser* sowie *Metalle und Metallgewinnung* erfasst. Aufgrund der Korrelationsanalyse, die im Rahmen der Pilotierung durchgeführt wurde, konnte eine hohe Korrelation zwischen den Ergebnissen im allgemeinen Chemie-Leistungstest und den Ergebnissen im Chemie-Leistungstest zum Modul festgestellt werden (vgl. Abschnitt 6.3). Das spricht dafür, dass beide Tests das Merkmal „Umgang mit chemischem Fachwissen“ abbilden. Damit können die Ergebnisse des ACL-Tests zur Deutung der Ergebnisse in den CL-Tests herangezogen werden.

55 Da die Einführung der Formelschreibweise erst zu Beginn von Klasse 8 erfolgt, ist der Anteil an Aufgaben, die eine Lösung auf Teilchenebene erfordern, im Test zum Modul etwas höher ist als im ACL-Test zu Beginn von Klasse 8.

5.4.5 Interesse und Selbstkonzept

Das individuelle Interesse für ein Unterrichtsfach sowie das fachbezogene, akademische Selbstkonzept sind für die vorliegenden Überlegungen von Bedeutung, weil sie einen Einfluss auf die sachfachliche, (fremd-)sprachliche und motivationale Wirksamkeit des bilingualen Unterrichts haben können (vgl. Kapitel 4).

Dispositionales Interesse

Nach der Person-Gegenstands-Theorie wird Interesse als eine Beziehung zwischen einer Person und einem bestimmten Gegenstand des schulischen Lernens, also z.B. einem bestimmten Unterrichtsfach gesehen (vgl. Abschnitt 5.3.3.2). Dieses Verhältnis ist von Emotionen und einer kognitiven Bewertung geprägt und kann sich zu einer stabilen, situationsunabhängigen Disposition entwickeln. Über die Art der Entstehung dispositionaler Interessen ist jedoch nicht viel bekannt. Sie bilden sich im Kindes- und Jugendalter und verfestigen sich etwa bis zum 16. Lebensjahr, wobei sich Schüler verstärkt einzelnen Fächern zu- und von anderen abwenden. Es scheint, dass sich fachliches Interesse positiv auf die Motivation aber auch auf die Lernleistungen der Schüler auswirkt. Interessierte Lerner investieren mehr Zeit in den Lernprozess. Außerdem zeigen sie eine stärkere Fokussierung der Aufmerksamkeit, eine Erhöhung der Konzentration, eine größere Behaltensleistung, ein tiefergehendes Textverständnis sowie den Einsatz höherwertiger Strategien zum Wissenserwerb (Krapp & Seidel, 2014, S. 212). Im Extremfall erreichen Lerner den Zustand des *Flows*, der höchste Lerneffizienz ermöglicht, weil Lerner hoch konzentriert sind, einen Bezug zu Raum und Zeit verlieren und in der jeweiligen Aufgabe aufgehen (Csikszentmihalyi, 1975). Krapp (2014, S. 213) resümiert dazu, dass eine auf Selbstbestimmung und Interessen beruhende Lernmotivation optimale Voraussetzungen bietet, "wenn das Ziel darin besteht, tiefgründig verankertes und somit relativ dauerhaft gespeichertes Wissen zu erwerben." In einer von Schiefele, Krapp und Schreyer (1993) durchgeführten Meta-Analyse zur Beziehung von schulbezogenem (thematischen) Interesse und Schulnoten als Indikator des Schulerfolgs, konnte über alle Fächer und Schulstufen hinweg eine mittlere Korrelation von $r = 0.30$ ermittelt werden (zitiert nach Krapp, 2014, S. 209). Auch Schiefele und Schaffner (2015) konnten in Untersuchungen zu schulischen Leistungen eine kleine bis mittelstarke Korrelation der Leistungen und der intrinsischen Motivation nachweisen. Insbesondere bei Schülern der Oberstufe ist das Interesse dabei ein Prädiktor für die gezeigte Leistung. Bei Schülern der Unter- und Mittelstufe ist dies nicht der Fall. Der Einfluss des Interesses auf die Leistung scheint sich erst mit zunehmendem Alter herauszubilden (Köller et

al., 2001, S. 461)⁵⁶. Die Erfassung des Fachinteresses erfolgt über die Prefragebögen anhand eines Rating-Verfahrens, wobei wieder Multi-Item-Skalen auf der Basis vierstufiger Likert-Skalen zum Einsatz kamen. Vorhandene Skalen zum Fachinteresse konnten dabei nicht direkt aus anderen Untersuchungen übernommen werden und wurden den Bedürfnissen der Studie angepasst. Die Items wurden weitgehend aus den für die DESI- und PISA-Studien entwickelten Items erstellt und anschließend pilotiert (vgl. Abschnitt 6.5 sowie Anhang). Zwei Items waren positiv und zwei Items negativ formuliert und wurden anschließend umcodiert. Die Auswertung der Ergebnisse erfolgte nach dem gleichen Verfahren wie bei der Erfassung der Subskalen der intrinsischen Motivation (vgl. Abschnitt 5.3.3.5). Das folgende Beispiel verdeutlicht die Vorgehensweise.

Item-Name	Item-Text	Item-Text und Quelle
	Einleitungssatz: Hier würden wir gerne etwas über die Fächer Englisch und Chemie erfahren [...]:	
FIEN 1	Das Fach Englisch macht mir Spaß.	Englisch macht mir Spaß. (DESI)
FIEN 2	Mit Englisch beschäftige ich mich gerne.	In meiner Freizeit beschäftige ich mich manchmal mit Englisch über die Hausaufgaben hinaus. (DESI)
FIEN 3	Ich mag das Fach Englisch nicht.	Keine Quelle, vom Autor entwickelt
FIEN 4	Englisch zu lernen ist langweilig.	Ich finde das Fach Englisch langweilig. (DESI)

Abbildung 27: Items zur Erfassung der Variable Interesse am Fach Englisch

Selbstkonzept

Die Persönlichkeit bzw. das *Selbstkonzept*, also die Vorstellung, die der Lernende von sich selbst und insbesondere von seinen Fähigkeiten hat, spielen für die Motivation von Lernprozessen eine

⁵⁶ Nold (2003, S. 167) konnte keinen signifikanten Einfluss des Interesses auf die Leistungen in der Fremdsprache beobachten, während die DESI- sowie die PISA-Studie zur Leistung im Fach Mathematik kleine Effekte des Interesses auf die Leistungen jüngerer Schüler feststellen konnten (Chiu und Xihua, 2008, S. 331 ff., zitiert nach Rumlich, 2016).

zentrale Rolle (Weiner, 1986, S. 286). Ein positives akademisches Selbstkonzept, d. h. das Gefühl, den Herausforderungen des Unterrichts gewachsen zu sein, führt zu mehr Selbstvertrauen. Untersuchungen bestätigen, dass sich dies positiv auf die Lernbereitschaft, das Engagement, die Ausdauer und den Lernerfolg auswirkt (vgl. Rumlich 2016, S. 128ff.). Helmke und Weinert (1997, S. 112) sehen im akademischen Selbstkonzept sogar den wichtigsten affektiv-motivationalen Prädiktor für schulischen Erfolg, unabhängig vom jeweiligen Fach. Akademische Selbstkonzepte sind dabei stark vom jeweiligen Unterrichtsfach abhängig. Sie entwickeln sich im Laufe der Sekundarstufe aufgrund der Lernerfahrungen in den einzelnen Fächern. Dabei spielen die erzielten Leistungen sowie insbesondere der Vergleich mit den Mitschülern eine entscheidende Rolle. Auch die Erwartungen und Rückmeldungen von Lehrern, Eltern und Mitschülern haben einen Einfluss auf das persönliche Urteil, so dass dieses nicht selten den gängigen stereotypen Vorstellungen folgt. Dabei konnte in Untersuchungen wiederholt nachgewiesen werden, dass Jungen bei gleichem Leistungsvermögen ein besseres akademisches Selbstkonzept in den mathematisch-naturwissenschaftlichen Fächern haben als Mädchen und vice versa (Möller & Trautwein, 2015, S. 191). Bei vielen Lernern entwickelt sich das mathematisch-naturwissenschaftliche Selbstkonzept unabhängig oder sogar im Kontrast zum sprachlichen Selbstkonzept, selbst dann, wenn die Leistungen in den beiden Aufgabenbereichen vergleichbar sind. Im Rahmen des reziproken internalen und externalen Referenzmodells (*internal / external frame of reference (I/E) model*, bzw. *reciprocal I/E model*) nach Marsh (1986) führt man diese Wahrnehmung auf Kontrasteffekte zurück, die sich ergeben, wenn Schüler ihre persönlichen Leistungen in den einzelnen Fächern miteinander vergleichen (internale Bezug). Gute Leistungen im Fach Mathematik führen damit nicht nur zu einem verbesserten mathematisch-naturwissenschaftlichen Selbstkonzept, sondern gleichzeitig zu einem verschlechterten sprachlichen Selbstkonzept. Das liegt offensichtlich daran, dass Lerner davon ausgehen, dass ein hohes Leistungsvermögen im mathematischen Bereich notwendigerweise ein geringeres Leistungsvermögen im sprachlichen Bereich zur Folge hat (reziprokes Verhältnis). Ein hohes fachspezifisches Selbstkonzept stellt nach Krapp & Ryan (2002, S. 71) zwar keine hinreichende aber dennoch eine notwendige Bedingung für ein gesteigertes Interesse dar. Da das Kompetenzerleben eine Voraussetzung für intrinsische Motivation und eine daraus resultierende Interessenbildung darstellt, wundert es nicht, dass die Einschätzung der eigenen akademischen Fähigkeiten einen hoch signifikanten Einfluss auf das Interesse hat, wie von Zaunbauer et al. (2009, S. 159) ermittelt. Das bedeutet, dass einseitige Interessen, wie sie im Konzept des sprachlich- bzw. des naturwissenschaftlich orientierten Lernalters zum Ausdruck kommen (siehe Abschnitt 5.2.4), Hand in Hand gehen mit einem besonders ausgeprägten Selbstkonzept im favorisierten Fachbereich.

Die Erfassung des akademischen Selbstkonzepts für die Fächer Englisch und Chemie erfolgt über die Prefragebögen nach einem Rating-Verfahren, wobei erneut Multi-Item-Skalen auf der Basis vierstufiger Likert-Skalen eingesetzt wurden. Zwei Items waren positiv und zwei Items negativ formuliert und wurden anschließend umcodiert. Vorhandene Skalen zum Selbstkonzept konnten nicht direkt aus anderen Untersuchungen übernommen werden, weil diese häufig andere Fächer betreffen oder weil sie auf nur drei Items pro Skala zurückgreifen (z.B. PISA). Die Items wurden weitgehend auf der Grundlage der für die DESI- und DENOCS-Studie entwickelten Items erstellt und anschließend pilotiert (vgl. Abschnitt 6.5 sowie Anhang). Die Auswertung der Ergebnisse erfolgte nach dem gleichen Verfahren wie bei der Erfassung der Subskalen der intrinsischen Motivation (vgl. Abschnitt 5.3.3.5). Das folgende Beispiel verdeutlicht die Vorgehensweise.

Item-Name	Item-Text	Item-Text und Quelle
	Einleitungssatz: Hier würden wir gerne etwas über die Fächer Englisch und Chemie erfahren [...]:	
SKEN1	In Englisch kann ich dem Unterricht nicht folgen.	Keine Quelle, vom Autor entwickelt
SKEN2	Englisch ist sehr schwer für mich.	Ich finde das Fach Englisch schwer. (DENOCS)
SKEN3	Den Unterricht in Englisch kann ich gut verstehen.	Keine Quelle, vom Autor entwickelt
SKEN4	In Englisch habe ich keine Probleme die Dinge zu verstehen.	Was wir in Englisch durchnehmen, verstehe ich meistens. (DESI)

Abbildung 28: Items zur Erfassung der Variable Selbstkonzept im Fach Englisch

5.4.6 Demografische Merkmale

Zu den demografischen Merkmalen, die über die Fragebögen erhoben werden, zählen *Alter*, *Geschlecht*, *Migrationshintergrund*, *Bildungsniveau des Elternhauses* sowie Inanspruchnahme von *Nachhilfeunterricht im Fach Chemie*. In zahlreichen großangelegten Studien (PISA oder TIMSS) konnte nachgewiesen werden, dass die Leistungen in den Naturwissenschaften nicht nur von Alter und Geschlecht, sondern auch vom Bildungsniveau des Elternhauses und einem möglichen Migrationshintergrund abhängen. Bei der Auswertung des sachfachlichen Lernerfolgs müssen diese Faktoren deshalb berücksichtigt werden (siehe Abschnitt 7.1.1). Zusätzlich wird ermittelt, ob die

Schüler externe Hilfe in Form von Nachhilfeunterricht in Anspruch nehmen, weil auch dies möglicherweise den Lehrerfolg beeinflusst.

5.5 Das Modul: *Alkali and alkaline earth metals*

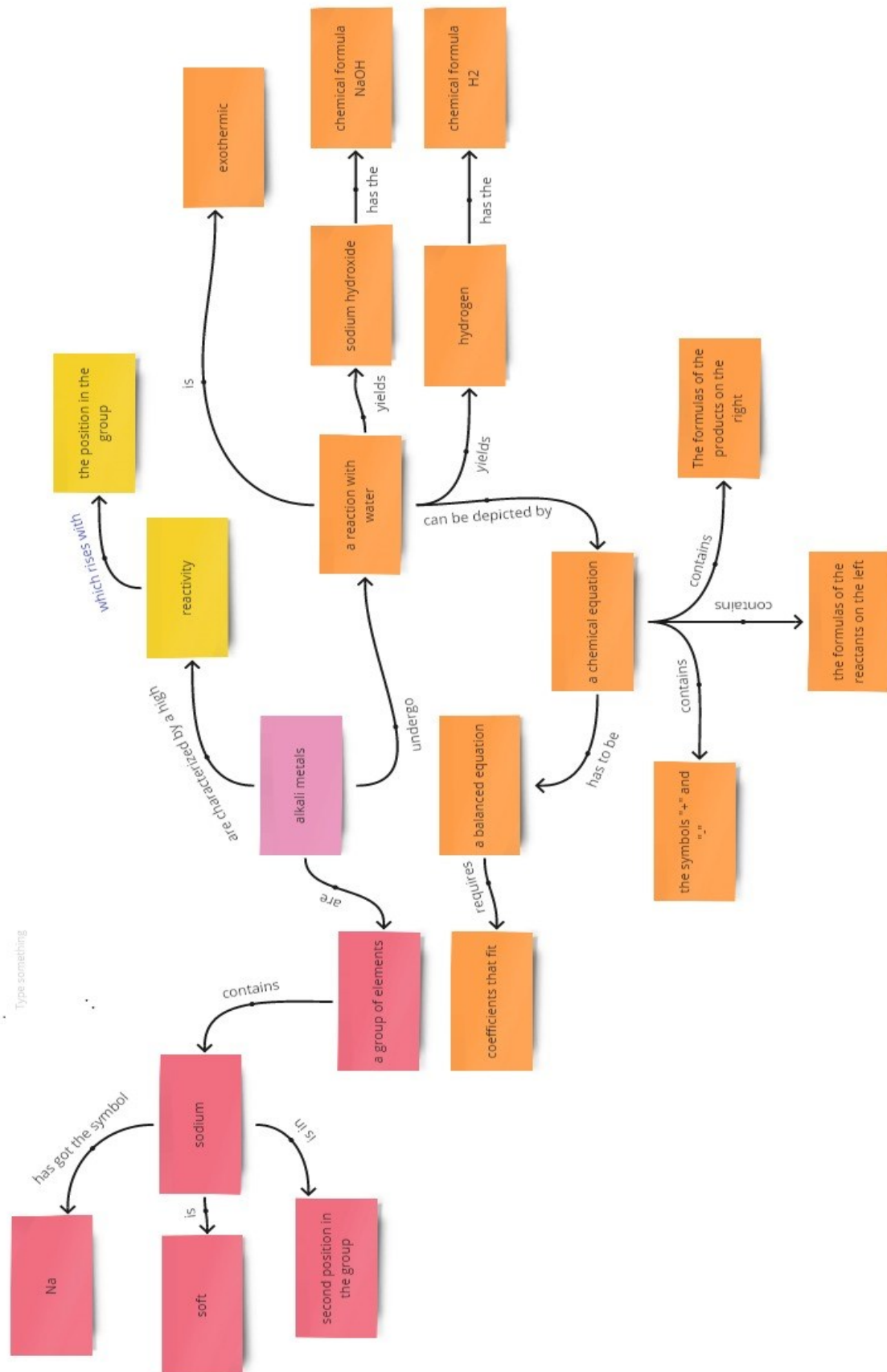
Im folgenden Abschnitt wird die der vorliegenden Untersuchung zugrunde liegende Unterrichtsreihe vorgestellt, mit dem Ziel, den didaktisch-methodischen Hintergrund des Chemiemoduls zu präzisieren, auf dessen Grundlage die Daten zur sachfachlichen, (fremd-) sprachlichen und motivationalen Wirksamkeit des bilingualen Chemieunterrichts gewonnen werden. Zunächst erfolgt dazu eine Einordnung des Themas in den Chemielehrplan sowie eine Darstellung der chemischen Inhalte, die Grundlage der Leistungstests sind (Abschnitt 5.5.1). Anschließend wird die Gestaltung des Unterrichts anhand der in Abschnitt 3.3 entwickelten Überlegungen zur Methodik bilingualen Chemieunterrichts präzisiert (Abschnitt 5.5.2).

5.5.1 Themen und curriculare Einbettung

Die Unterrichtsreihe *Alkali and alkaline earth metals* gehört zum Inhaltsfeld *Elementfamilien, Atombau und Periodensystem* des Kernlehrplans Chemie für die Sekundarstufe I in Nordrhein-Westfalen. Die Reihe bildet den Einstieg in das neue Inhaltsfeld, wobei die Entwicklung des Konzeptes der Elementfamilie am Beispiel der Alkali- und Erdalkalimetalle im Vordergrund steht. Die Schüler sollen erkennen, dass Elemente sich aufgrund ähnlicher physikalischer und chemischer Eigenschaften gruppieren lassen und an welchem Ort Elementfamilien im Periodensystem der Elemente zu finden sind. Später lernen sie diese Gemeinsamkeiten auf Übereinstimmungen im Atombau zurückzuführen, die sich aus der Stellung des Elements im PSE ergibt. In der auf diese Unterrichtsreihe folgenden Sequenz wird die Elementfamilie der Halogene eingeführt, bei der die am Beispiel der Alkali- und Erdalkalimetalle erworbenen Kenntnisse zum Konzept der Elementfamilie weiter vertieft werden. In der vorliegenden Reihe geht es darum, die wichtigsten Vertreter der Alkali- und Erdalkalimetalle kennenzulernen, gemeinsame Stoffeigenschaften, wie metallischer Charakter, hohe Reaktivität in Bezug auf die Reaktionen mit Wasser und Sauerstoff sowie relativ geringe Härte zu identifizieren. Außerdem sollen die Schüler erkennen, wie sich diese Eigenschaften in Bezug auf die Stellung im Periodensystem systematisch ändern. Zusätzlich werden die beiden Elementfamilien verglichen, d.h. es werden Gemeinsamkeiten und Unterschiede im physikalischen und chemischen Verhalten gegenübergestellt. Ihr Wissen über die Stoffe erwerben die Schüler dabei über chemische Experimente, die als Lehrerdemonstrationsversuch bzw. Schülerexperiment durchgeführt werden. Zur Auswertung der Experimente greifen die Schüler auf

ihr Vorwissen in den Bereichen chemische Reaktion, Energieformen, Metalle, saure und alkalische Lösungen sowie Nachweisreaktionen zurück. Außerdem wenden sie ihre Kenntnisse im Aufstellen von Reaktionsgleichungen an, und formulieren Gleichungen für die Reaktionen der Alkali- und Erdalkalimetalle mit Sauerstoff und Wasser. Die theoretischen Grundlagen hierzu wurden in der vorangehenden Unterrichtsreihe gelegt. Über Fachwissen zu verfügen, bedeutet gemäß den Bildungsstandards für den mittleren Schulabschluss im Fach Chemie (KMK, 2005, S. 8) „Wissen über chemische Phänomene“ und das „Verständnis von Begriffen, Gesetzmäßigkeiten, Prinzipien“ sowie von „Modellen“ zur Beschreibung von Stoffen und ihrer Veränderung. Im Rahmen des ESNaS-Kompetenzstrukturmodells, welches auch der Untersuchung zugrunde liegt, werden für die drei unterschiedlichen Arten von Wissen die Begriffe „Wissen um Fakten“ (Phänomenen), „Wissen um Zusammenhänge“ (Gesetzmäßigkeiten) sowie „Wissen um übergeordnete Prinzipien bzw. Konzepte“ (Begriffe und Prinzipien) verwendet (Ropohl, 2010, S. 29ff.). In der vorliegenden Unterrichtsreihe werden die Konzepte *Alkali-* und *Erdalkalimetall* als Teil des übergeordneten Konzepts der *Elementfamilie* eingeführt. Zur Wiedergabe konzeptuellen Wissens werden häufig sogenannte *concept maps* verwendet. Die Darstellung mittels *concept map* verdeutlicht, wie aus Sicht des Konstruktivismus semantisch-konzeptuelles Wissen im Gehirn gespeichert ist (vgl. Abschnitt 3.2.2.1). Das Erstellen von *concept maps* wurde von J. D. Novak (2006) beschrieben, der versuchte, das naturwissenschaftliche Wissen von Studenten darzustellen. Novaks Arbeit geht dabei auf die kognitiven Theorien von D. Ausubel (1968) zurück, der die Bedeutung von Vorwissen für den Wissenserwerb untersuchte. Die Konzepte, welche meist in Kreisen oder Kästchen angegeben werden, bilden die Knoten eines Netzwerkes, bei dem diese über eine Vielzahl unterschiedlicher Verbindungen (Kanten) miteinander verbunden sind. Diese Verbindungen werden als Pfeile dargestellt, wobei die Art der Verbindung, also die logische Beziehung zwischen den Konzepten jeweils angegeben ist. Die entsprechende Proposition ergibt sich, wenn man den betreffenden Netzausschnitt (Zweig) von einem Knoten ausgehend in Pfeilrichtung bis zum nächsten Knoten liest. *Concept maps* sind hierarchisch strukturiert und müssen von oben nach unten gelesen werden. Die folgende Abbildung zeigt einen Ausschnitt aus der *concept map* der *alkali metals*. Der Ausschnitt zeigt auch Beispiele für die drei unterschiedlichen Arten von Wissen, welche das übergeordnete Konzept der Alkalimetalle (in Pink) abbilden. In Rot sind einzelne *Fakten*, in Gelb ist ein *Zusammenhang* und in Orange ist ein *übergeordnetes Konzept*⁵⁷ dargestellt.

57 Für Köller et al. (2008b) sind übergeordnete Konzepte dadurch charakterisiert, dass sie als Prinzip, Gesetz oder Regelhaftigkeit von der gegebenen Situation unabhängig formulierbar sind und mit ihrer Hilfe Zusammenhänge beschrieben werden können.

Abbildung 29: Ausschnitt aus der *concept map* der *alkali metals*

Die Schüler verfügen zu diesem Zeitpunkt allerdings bereits über chemisches Vorwissen, auf das im Laufe der Reihe zurückgegriffen wird:

Oxidation, Redoxreaktion, Reaktion, Reaktionsgleichung, Metall, Reaktivität, exotherme Reaktion, Dichte, Rauch, Löslichkeit, Gas, Härte, Schmelzpunkt, alkalischer Charakter, Lösung, Löslichkeit, Suspension, Indikator, Rückstoßprinzip, Energie, Aktivierungsenergie, Reaktionsenergie, Eindampfen, Flamme, Aggregatzustand, pneumatisches Prinzip, Verbrennung, Auftrieb, Verhältnisformel

Abbildung 30: Weitere Konzepte, auf die im Modul zurückgegriffen wird

Gefordert ist im Unterricht jedoch nicht nur die reine Reproduktion von Wissen, sondern der Umgang damit unter Berücksichtigung höherer Denkopoperationen. Kauertz et al. (2010) sprechen deshalb in ihrem Kompetenzmodell auch nicht von "Fachwissen", sondern vom "Umgang mit Fachwissen". Unterrichtsmethodisch ist die Schulung höherer Denkopoperationen im bilingualen Unterricht an das Einüben unterschiedlicher Diskursfunktionen gebunden (siehe Abschnitt 3.3.2). Die in der Unterrichtsreihe zum Einsatz kommenden Diskursfunktionen sind in der folgenden Tabelle aufgelistet.

Makrofunktion	beschreiben, berichten deuten, vergleichen
Mikrofunktion	nennen, benennen, beschreiben, berichten, zusammenfassen erklären, begründen, schlussfolgern, deuten, Hypothesen bilden, verallgemeinern Meinung ausdrücken

Abbildung 31: Diskursfunktionen zum Modul

Die Unterrichtsreihe umfasst insgesamt zehn Stunden. Ein Überblick über die Themen und Inhalte der Reihe befindet sich im Anhang. Die Unterrichtsreihen unterscheiden sich in der Unterrichtssprache.

5.5.2 Methodische Gestaltung des Unterrichts

Der Unterricht folgt einer kontextorientierten Herangehensweise (vgl. Abschnitt 3.3.1), die das Alltagswissen der Schüler miteinbezieht und einen direkten Bezug zur Lebenswirklichkeit der

Schüler ermöglichen soll. Dazu werden zunächst Mineralien wie Natrium, Magnesium und Calcium vorgestellt, die Bestandteile vieler Mineralwasser sind. Ihr Vorkommen und ihre Bedeutung für den menschlichen Körper werden erläutert, um anschließend, ausgehend von den Elementen der Elementfamilie der Alkali- und der Erdalkalimetalle, zu den einzelnen Verbindungen der Mineralstoffe zu kommen. Das vorliegende Modul dient dazu, das Konzept der Elementfamilie zu entwickeln und stoffliches Wissen zu den Alkali- und den Erdalkalimetallen zu erwerben und dieses in Beziehung zu setzen. Deutlich soll auch werden, inwieweit der Aufbau des PSE sich an den Eigenschaften der Elemente orientiert. Ein Bezug zum Atombau ist erst nach Einführung des Schalenmodells möglich. Als Datengrundlage zur Generierung von strukturiertem, stofflichem Wissen dienen in der vorliegenden Reihe zahlreiche Experimente, die als Lehrerdemonstrations- bzw. als Schülerexperiment durchgeführt werden. Sie sind teilweise recht spektakulär, wie die Zugabe von Kalium zu Wasser, und bieten sich wegen ihrer auf Schüler motivierenden Wirkung an. Außerdem vermitteln sie Informationen über unterschiedliche Wahrnehmungskanäle (z.B. sehen, riechen, hören, fühlen). Damit regen sie in besonderem Maße zu sozialer Interaktion in Form eines kommunikativen Austauschs mit den Mitschülern und dem Lehrer an, wodurch Assimilations- und Akkomodationsprozesse in Gang gesetzt werden. Der Erwerb des Konzeptes Alkali- bzw. Erdalkalimetall vollzieht sich dabei ausgehend vom Konzept der Metalle als *Konzeptwechsel*. Das Konzept der Metalle ist den Schülern aus Klasse 7 bekannt und auf dieser Lernstufe noch stark durch alltagsweltliche Vorstellungen geprägt. Ihr Vorwissen erfährt insofern eine Erweiterung, als sie erkennen, dass Metalle auch reaktiv, sehr leicht und sehr weich sein können. Diese Erkenntnisse können die Schüler, im Sinne des *entdeckenden Lernens*, den Experimenten selbst entnehmen und mit den Mitschülern und dem Lehrer diskutieren (*Phänomenorientierung*). So erhalten sie Gelegenheit, neues Wissen im Sinne eines aktiven Konstruktionsprozesses an bereits vorhandene Wissensstrukturen anzuknüpfen und weiterzuentwickeln. Das gilt in besonderem Maße für das arbeitsteilig durchgeführte Schülerexperiment zum Calcium, bei dem es auch darum geht, die bei den Alkalimetallen gewonnenen Erkenntnisse auf die Nachbargruppe der Erdalkalimetalle zu übertragen. Die Gestaltung der Unterrichtsreihe orientiert sich dabei an einer moderaten Form des Konstruktivismus, dem sogenannten wissensbasierten Konstruktivismus, bei dem sich Phasen der Instruktion und der Konstruktion abwechseln (Reinmann & Mandl, 2006). Dabei dienen instruktionale Phase dazu, eine Wissensbasis zu schaffen, die es Lernern ermöglicht, eigenständig Probleme zu lösen und eigenes Wissen zu konstruieren. Damit durch die fremdsprachliche Hürde die anspruchsvolle inhaltliche Arbeit nicht gefährdet wird (*cognitive overload*), folgen die Einzelstunden der im Rahmen der *CLIL-Matrix* (Coyle et al., 2012) vorgeschlagenen Abstimmung fremdsprachlicher und sachfachlich-kognitiver Anforderungen. Das bedeutet, dass in kognitiv

anspruchsvollen, mündlichen Erarbeitungsphasen ein Eingreifen durch den Lehrer im Sinne von *prompting* oder *bridging* oder dem kurzfristigen Wechsel in die Muttersprache (*code switching*) nötig werden. Sicherungsphasen, in denen der kognitive Anspruch in der Regel geringer ist, erfolgen hingegen, wenn möglich, ausschließlich in der Fremdsprache. Hier werden auch sprachbezogene Fragen thematisiert, wie etwa die adäquate Formulierung. Neben der schriftlichen Sicherung an der Tafel findet im vorliegenden Modul eine zusätzliche Sicherung durch geschlossene, halb-offene und offene, schriftliche Aufgaben statt. Diese sind von den Schülern selbstständig und meist zu Hause zu bearbeiten, und bieten insbesondere langsameren Schülern die Gelegenheit, inhaltliche Zusammenhänge in ihrem eigenen Lerntempo zu verstehen und konzeptuelles Wissen zu konstruieren. Das Medium, in dem gelernt wird, ist somit in erster Linie die Fremdsprache. Neue Konzepte sollen in dieser erarbeitet und geübt werden, auch, um die kognitiven Vorteile eines Konzepterwerbs in der Fremdsprache aufgrund der vermuteten „tieferen Verarbeitung“ nutzbar zu machen. Damit entspricht der Spracheinsatz dem Typ A nach Diehr (2012, S. 27), wie er auch für den Modulunterricht von ihr vorgeschlagen wird.

6 Die Pilotphase

Ziel der Pilotphase ist es, das Forschungsdesign einer kritischen Überprüfung zu unterziehen und mögliche Mängel im Hinblick auf die Hauptuntersuchung zu beseitigen. Dazu wurden insbesondere die Verfahren der Datenerhebung und der Datenauswertung untersucht. Gleichzeitig liefern die Ergebnisse aus den Pilotgruppen auch einen ersten Einblick in das Forschungsfeld und können dabei helfen, Forschungsfragen zu präzisieren bzw. zu verwerfen, weitere Störvariablen zu identifizieren und Abläufe zu optimieren. Dem explorativen Charakter der Pilotphase kommt gerade bei Felduntersuchungen eine besondere Bedeutung zu, da mögliche Störvariablen nicht immer im Vorfeld klar zu erkennen und zu kontrollieren sind. Nach einigen grundlegenden Informationen zur Testgüte und zu den Rahmenbedingungen der Pilotstudie (Abschnitt 6.1) folgt die detaillierte Pilotierung der selbstentwickelten Erhebungsinstrumente. Sie betrifft den Chemietest zum Modul (Abschnitt 6.2), den allgemeinen Chemie-Leistungstest (Abschnitt 6.3), die Wortschatztests (Abschnitt 6.4) und die Fragebögen (Abschnitt 6.5). Es sei darauf hingewiesen, dass in diesem Abschnitt aus Platzgründen nur die Ergebnisse der Pilotierung wiedergegeben werden.

6.1 Testgüte und Rahmenbedingungen

Die Erfassung der im Rahmen der vorliegenden Studie untersuchten Variablen erfolgt mit Hilfe von Tests und Fragebögen. Wo es möglich ist, wird auf bereits erprobte und in der Regel standardisierte Erhebungsinstrumente zurückgegriffen (KFT, LGVT 5-12+, C-Test). War dies nicht möglich, wurden eigene Instrumente in Anlehnung an bereits bestehende Tests entwickelt und evaluiert (Chemietests, Wortschatztests, Fragebögen). Die Qualität der Erhebungsinstrumente hat maßgeblichen Einfluss auf die Qualität der Daten, welche die Grundlage der Auswertung darstellen. Deshalb werden alle eingesetzten Erhebungsinstrumente im Hinblick auf ihre Güte untersucht. Die wichtigsten Gütekriterien der klassischen Testtheorie sind laut Bortz & Döring (2002, S. 193ff.) *Objektivität*, *Reliabilität* (Zuverlässigkeit) und *Validität* (Gültigkeit).

Gütekriterien

Objektiv ist ein Test, wenn seine Ergebnisse unabhängig vom einzelnen Testleiter und Auswerter sind. Die *Objektivität* eines Tests hängt insbesondere von einer objektiven Testdurchführung und einer ebensolchen Auswertung der Aufgaben ab. Eine objektive Durchführung bedeutet, dass die Art der Durchführung von Testung zu Testung nicht variiert und somit für alle Testpersonen gleich ist. Daher werden zu jedem Test die Instruktionen des Testleiters schriftlich festgehalten. Weiterhin muss die Auswertung auf objektive Weise erfolgen. Eine bestimmte Schülerlösung sollte immer gleich bewertet werden, unabhängig davon, wer die Aufgabe auswertet. Die Auswerter erhalten deshalb genaue Anweisungen, wie eine Aufgabe zu bewerten ist. Dazu erhalten sie, falls nötig, einen Lösungsbogen, der Musterlösungen und konkrete Hinweise zur Bewertung enthält.

Reliabel ist ein Test, wenn er das, was er erfasst, zuverlässig (d.h. genau und stabil) erfasst. *Reliabilität* (Zuverlässigkeit) bezeichnet die Genauigkeit, mit der ein Test das zu messende Merkmal messen kann. Je nach Test wurden unterschiedliche Verfahren zur Reliabilitätskontrolle eingesetzt (Bortz & Döring, 2002). Gültig ist ein Test, wenn er das misst, was er messen soll. Die *Validität* (Gültigkeit) wird oft als das wichtigste Güte Merkmal eines Tests bezeichnet. Ein Test kann sich als objektiv und reliabel erweisen, dennoch ein anderes als das ursprünglich intendierte Konstrukt messen. Genau dieser Frage wird deshalb im Rahmen von Validitätsanalysen nachgegangen. Auch hier kommen, abhängig vom Test, unterschiedliche Verfahren zum Einsatz (Bortz & Döring, 2002).

Pilotgruppe

Zur Überprüfung der Testgüte wurde im Rahmen einer Pilotierung ($N = 52$) eine ausführliche Itemanalyse durchgeführt. Dabei spielen die Formulierungen der einzelnen Items und der Instruktionstexte eine wichtige Rolle. Es muss sichergestellt sein, dass sie von den Probanden verstanden werden. Die Bearbeitungszeit wird ebenfalls ermittelt. Diese darf insgesamt nicht zu hoch sein, um eventuelle Ermüdungserscheinungen zu vermeiden. Die Pilotgruppe umfasst insgesamt 52 Schüler. Damit ist auch die Berechnung psychometrischer Kennwerte, wie z.B. Trennschärfen, zuverlässig möglich. Riese und Reinhold (2014, S. 264) empfehlen hierzu eine Gruppengröße von mindestens 40 Schülern. Damit die Erkenntnisse auf die Probanden der Hauptgruppe übertragbar sind, wird die Pilotgruppe außerdem so ausgewählt, dass sie dieser möglichst ähnlich ist. Die Pilotphase fand im Schuljahr vor der Hauptuntersuchung statt. Es wurden, wie in der Hauptuntersuchung, zwei achte Klassen derselben Schule ausgewählt, wobei die eine bilingual (Untersuchungsgruppe) und die andere auf Deutsch (Kontrollgruppe) unterrichtet wurde. Beide Gruppen wurden in der Modulphase vom gleichen Chemielehrer unterrichtet. In der Kontrollgruppe wurde das Modul im Rahmen des regulären Chemieunterrichts eingesetzt. Es umfasste in dieser Gruppe insgesamt fünf Wochen mit je zwei Wochenstunden. In der Untersuchungsgruppe fand das Modul im Rahmen des regulären Englischunterrichts der Klasse statt, der dazu in die Fachräume der Chemie verlegt wurde, da der Untersuchungsleiter nicht der Chemielehrer aber der Englischlehrer dieser Klasse war. Aufgrund der Vierstündigkeit des Englischunterrichts ergab sich ein etwas anderer zeitlicher Verlauf des Modulunterrichts. Dieser betrug in der Untersuchungsgruppe knapp drei Wochen mit jeweils vier Stunden Unterricht. Der Chemieunterricht wurde in dieser Zeit parallel von einer Kollegin erteilt, die die Inhalte der Modulphase aussparte.

6.2 Chemie-Leistungstest

Die eingesetzten Chemietests dienen dazu, den Umgang mit dem im Modul erworbenen Fachwissen gemäß den in den Kernlehrplänen formulierten drei Anforderungsbereichen zu erfassen. Die Aufgaben wurden anhand der vorab formulierten Lehrziele sowie der im Rahmen des ESNaS-Projektes entwickelten Konstruktionsanweisung für Testaufgaben entwickelt (vgl. Abschnitt 5.3.1.3). Im Folgenden werden die Ergebnisse der Evaluation der Tests in Bezug auf die Gütemerkmale Objektivität, Validität und Reliabilität sowie die sich hieraus ergebenden Implikationen für die Hauptstudie beschrieben.

6.2.1 Objektivität

Zur Gewährleistung einer möglichst objektiven Testdurchführung wurde eine Testleiterinstruktion verfasst. Diese beinhaltet alle durch den Testleiter zu gebenden Anweisungen bezüglich der Bearbeitung der Aufgaben, der zu verwendenden Hilfsmittel sowie der einzuhaltenden Zeitvorgaben. Zur Objektivierung der Auswertung wurde ein Lösungsbogen erstellt, welcher die Lösungen zu den einzelnen Aufgaben sowie Hinweise zur Bewertung von Schülerantworten enthält. Anders als bei den geschlossenen Aufgaben vom Typ *Multiple-Choice-Single-Select*, die sich durch eine hohe Auswertungsobjektivität auszeichnen, muss für die halb-offenen und insbesondere für die offenen Aufgaben möglichst klar definiert sein, welche Antworten akzeptabel sind, und wie diese bewertet werden (vgl. Riese und Reinhold, 2014, S. 262). Bei den halb-offenen Aufgaben (Lückentext mit Ergänzung) wurden im Rahmen der Pilotierung alle Schülerlösungen untersucht und mit den durch die Experten im Vorfeld festgelegten Lösungsmöglichkeiten verglichen. Neue und von den Experten ebenfalls für akzeptabel befundene Schülerlösungen wurden anschließend in den Lösungsbogen aufgenommen⁵⁸. Die Überarbeitung des Codierleitfadens für die offenen Aufgaben (*Kurzaufsatz-Aufgabe*) gestaltete sich als deutlich umfangreicher, weil die Schülerlösungen nicht nur Formulierungsalternativen enthielten, sondern zum Teil ganz andere Lösungswege eingeschlagen wurden als von den Experten vorgesehen. Das Vorgehen verdeutlicht folgendes Beispiel:

Item 41, Test B

41. Bei Experimenten mit elementarem Barium sollen stets Schutzhandschuhe getragen und Hautkontakt mit dem Metall vermieden werden. Erkläre diese Schutzmaßnahmen möglichst genau (mögliche Reaktionen bitte erläutern).

Abbildung 32: Offene Aufgabe aus dem Chemieleistungstest

Für die vollständige Lösung der Aufgabe gibt es drei Punkte, da sich die Argumentation auf drei Kategorien reduzieren lässt:

1. Die Haut ist durch Wasser feucht.
2. Barium reagiert mit Wasser zu Bariumhydroxid.
3. Bariumhydroxid ist stark alkalisch bzw. ätzend.

⁵⁸ Zum Verfahren der deduktiven und induktiven Kategorienbildung vergleiche auch Hammann und Jördens (2014, S. 172ff.).

Die Schüler der Pilotgruppe begründeten die Gefährlichkeit von Barium aber auch anhand der Reaktivität des Metalls, das unter Hitzeentwicklung mit Wasser reagiert, was zu Verbrennung führen kann. Deshalb wurde alternativ zum Hinweis auf den alkalischen Charakter des Metalls auch der Hinweis auf den stark exothermen Verlauf der Reaktion akzeptiert. Für die Kategorie 3 der Codiervorschrift sind deshalb zwei Alternativen denkbar:

3a) Bariumhydroxid ist stark alkalisch bzw. ätzend.

oder

3b) Die Reaktion ist stark exotherm bzw. setzt große Wärmemengen frei.

Damit konnte die volle Punktzahl von drei Punkten entweder durch die Formulierung von Kategorie 3a) oder 3b) erreicht werden. Die Codiervorschrift wurde anschließend entsprechend verändert und die Endfassung in das Lösungsheft zum Test aufgenommen. Insgesamt wurden auf diese Art und Weise alle 16 Vorschriften für offene und halb-offene Aufgaben überarbeitet.

6.2.2 Validität

Im Rahmen der Validitätsprüfung soll überprüft werden, ob die Chemietests tatsächlich das testen, was sie zu testen vorgeben, nämlich den Umgang mit chemischem Fachwissen in Bezug auf die drei Anforderungsbereiche. Die Aufgaben beider Testvarianten (A und B) wurden durch einen erfahrenen Chemielehrer erstellt und anschließend in einem Expertenteam, bestehend aus zwei weiteren Chemielehrern mit mehrjähriger Berufserfahrung, besprochen. Dabei wurde diskutiert, inwieweit durch die Aufgaben chemisches Fachwissen adressiert wird. Für jede Aufgabe wurde das zur Lösung nötige Fachwissen benannt. Für jede der jeweils 42 Aufgaben aus den Tests A und B wurde festgestellt, dass sie Fachwissen adressieren und welcher Art dieses Wissen ist. Zusätzliche Hinweise auf die Validität des Tests können sich dann ergeben, wenn die Testergebnisse mit anderen Kriterien korrelieren, die ebenfalls einen Hinweis auf die Leistungsstärke in Chemie geben, wie z.B. die Vornote im Fach Chemie (Lienert und Raatz, 1994, S. 220ff)⁵⁹. Die entsprechenden Berechnungen von Kendall-Tau-b zeigen, dass die Chemienote ($r = 0,482$, $p = 0,001$) aus Klasse 7 hochsignifikant mit dem Ergebnis des Posttests korreliert. Im nächsten Schritt soll geklärt werden,

⁵⁹ Man geht dabei davon aus, dass auch das externe Kriterium in Bezug auf die zu messende Variable valide ist.

Inwieweit allerdings die Schulnote im Fach Chemie die Fähigkeit zum Umgang mit chemischem Fachwissen valide abbildet, ist fraglich.

inwieweit auch die drei Anforderungsbereiche durch die Tests abgebildet werden. Zur Klärung dieser Frage wurden die einzelnen Testitems im Rahmen eines Expertenratings, einem Team aus drei Chemielehrern mit mehrjähriger Berufserfahrung, zur Einschätzung vorgelegt. Die Experten ordneten jedem einzelnen Item einen Anforderungsbereich zu. Unterstützend dazu erhielten sie eine Übersicht über die in den Klassen 7 und 8.1 behandelten Inhalte. Anschließend wurde die Interrater-Korrelation der drei Rater durch Berechnung von Fleiss' Kappa bestimmt (vgl. Bortz & Döring, 2002). Es zeigt sich, dass eine hoch signifikante Übereinstimmung vorliegt, die nach Landis & Koch (1977) einigermaßen gut ist ($Kappa = 0,342$, $r = 0,079$, $z = 4,303$, $p < 0,01$).

6.2.3 Reliabilität

Zur Prüfung der Reliabilität (Zuverlässigkeit) eines Tests stehen unterschiedliche Verfahren zur Verfügung. Zu den gängigsten Methoden zählen die *Testwiederholungsmethode*, die *Paralleltestmethode*, die *Testhalbierungsmethode* sowie die *Konsistenzanalyse* (Lienert und Raatz 1994, S. 180). Für die vorliegenden Tests wurden Konsistenzanalysen durchgeführt. Die *innere Konsistenz* kann anhand der Berechnung des Kennwertes Cronbachs Alpha bestimmt werden.

Anforderungsbereich	Test A	Test B
AFB I	0,825	0,737
AFB II	0,806	0,800
AFB III	0,626	0,666
AFB I, II, III	0,885	0,881

Abb. Innere Konsistenz des Chemie-Leistungstests (Cronbachs Alpha)

Die Werte von Cronbachs Alpha liegen überwiegend im Guten ($>0,8$) bzw. im akzeptablen ($>0,7$) Bereich ⁶⁰. Sie deutet darauf hin, dass die Testaufgaben innerhalb einer Testform von den Probanden ähnlich gelöst wurden. Dies gilt sowohl für die Aufgaben zu den einzelnen Anforderungsbereichen als auch für den Gesamttest. Im nächsten Schritt wird überprüft, inwieweit die einzelnen Testitems

⁶⁰ Der Wert für den Anforderungsbereich III fällt in beiden Testformen etwas niedriger aus. Während in den Anforderungsbereichen I und II nur geschlossene und halb-offene Aufgaben zum Einsatz kommen, wird der Anforderungsbereich III ausschließlich durch offene Aufgaben erfasst. Die ermittelten Werte können ein Hinweis auf die geringere Reliabilität offener Aufgabenformate sein.

sich im Hinblick auf ihre *Schwierigkeit* und ihre *Trennschärfe* unterscheiden. Dazu werden Itemzwillinge aus den Testformen A und B betrachtet. Items mit geringen oder negativen Trennschärfen werden genauso überarbeitet wie Items, deren Schwierigkeiten sich in beiden Testformen deutlich unterscheiden. Für jedes Item wurde der korrigierte Trennschärfekoeffizient als Produkt-Moment-Korrelation zwischen den Punktwerten pro Item und dem korrigierten Gesamttestwert ermittelt (Bortz, 2002, S. 219). Die Trennschärfen liegen im Mittel bei $r = 0,41$ (Test A) bzw. $r = 0,35$ (Test B) und damit im mittelmäßigen bis guten Bereich (Bortz, 2002, S. 219), wenngleich sich 30 der insgesamt 84 Items als problematisch erwiesen, da ihre Trennschärfen unter 0,3 lagen. Anschließend wurden die Trennschärfen und die Schwierigkeiten der Items analysiert. Dabei wurde berücksichtigt, dass ein zu geringer Schwierigkeitsgrad genauso wie ein zu hoher Schwierigkeitsgrad zu einer geringen Trennschärfe führen (Decken- bzw. Bodeneffekt). Will man jedoch den gesamten Schwierigkeitsbereich in einem Niveau-Test erfassen, so müssen Trennschärfeeinbußen in Kauf genommen werden (Bortz, 2002, S. 219). Auch Rumlich (2014, S. 272f.) weist darauf hin, dass besonders schwere bzw. besonders leichte Items mit geringeren Trennschärfekoeffizienten notwendig sind, um auch zwischen leistungsstarken und besonders leistungsstarken bzw. zwischen leistungsschwachen und besonders leistungsschwachen Probanden unterscheiden zu können. Lediglich negative Trennschärfekoeffizienten sind nach Rumlich inakzeptabel (Rumlich, 2014, S. 273). Es wurden deshalb nur solche Items überarbeitet, deren Trennschärfe besonders gering ($TS < 0,1$) bzw. negativ war. Leichte Items, die nicht trennscharf waren, wurden nicht zwangsläufig überarbeitet. Dies ist bei einem Test mit einer relativ großen Zahl von 42 Items jedoch kein Problem. Um Mängel in der Itemformulierung bestimmen zu können, wurde bei den MC-Aufgaben auch auf die Ergebnisse der Distraktorenanalyse zurückgegriffen. Dabei wurden sowohl auf die Ergebnisse der Pre- als auch der Prosttests berücksichtigt. Insbesondere die Ergebnisse der Pretests waren hilfreich dabei mangelhafte Distraktoren aufzuspüren, etwa wenn die Lösungshäufigkeiten deutlich von der für den Pretest zu erwartenden Ratewahrscheinlichkeit von 25 Prozent abwichen. Offene und halboffene Items wurden einzelnen Schülern zum Teil nochmals vorgelegt, um zu klären, warum eine Lösung besonders leichtgefallen war. Das folgende Beispiel verdeutlicht die Vorgehensweise.

Item	Item aus Pilotfragebogen - Erläuterungen zur Überarbeitung - überarbeitetes Item für die Hauptuntersuchung
15 AB	Item15B Welche Formel hat Magnesiumhydroxid? A MgOH_2 B MgOH C MgO_2H D $\text{Mg}(\text{OH})_2$
	Die Distraktorenanalyse zeigte, dass sich kein Schüler für die Varianten A (MgOH_2) und C (MgO_2H) entschieden hat. Die gleiche Beobachtung macht man beim analogen Item in Test A, wo es um die Formel von Berylliumhydroxid ging. Hier wurden weder BeOH_2 (Antwort A) noch BeO_2H (Antwort D) als mögliche Formeln von den Schülern in Betracht gezogen. Alle diese Formeln zeichnet aus, dass die Buchstabenkombination OH, welche für die Hydroxidgruppe steht, aufgrund des eingefügten bzw. angehängten Index 2, nicht mehr als solche erkennbar ist. Deshalb wurden die Distraktoren modifiziert und das modifizierte Item 15 A, das sich als trennscharf erwiesen hatte, für beide Tests verwendet.
	Item 15 (Test A und B) Welche Formel hat Berylliumhydroxid? A $\text{Be}_2(\text{OH})_2$ B $\text{Be}(\text{OH})_2$ C BeOH D Be_2OH

Abbildung Überarbeitung des Items 15 B

6.3 Allgemeiner Leistungstest in Chemie

Der ACL-Test dient dazu, den Leistungsstand der Schüler in Bezug auf den Umgang mit dem chemischen Fachwissen der Klassen 7 und 8.1 zu erfassen. Dieser wird im Rahmen der vorliegenden Untersuchung als Kontrollvariable erhoben und soll dazu beitragen, die Ergebnisse zum sachfachlichen Lernerfolg und zur Motivation besser zu verstehen. Der ACL-Test wurde nach den gleichen Vorgaben wie der CL-Test konstruiert. Im Folgenden werden die Ergebnisse der Evaluation des Tests in Bezug auf die Gütemerkmale Objektivität, Validität und Reliabilität beschrieben. Dabei kommen die gleichen Verfahren wie bei der Pilotierung der CL-Tests zum Einsatz.

6.3.1 Objektivität

Auch für den ACL-Test wurde eine Testleiterinstruktion zur Erhöhung der Durchführungsobjektivität verfasst. Auch beim ACL-Test wurden sämtliche Aufgaben sowie das Lösungsheft inklusive der Codiervorschriften für offene Aufgaben überarbeitet. Dies erfolgte nochmals auf der Grundlage des Lösungsverhaltens der Pilotschüler. Dabei wurde auch auf die Ergebnisse der Distraktorenanalyse bzw. der mündlichen Rückmeldungen der Schüler zurückgegriffen. Das Verfahren wurde bereits für den CL-Test in Abschnitt 6.2 beschrieben. Insgesamt wurden auf diese Art und Weise alle 16 Codiervorschriften für offene und halb-offene Aufgaben überarbeitet.

6.3.2 Validität

Ziel der Validitätsprüfung ist es sicherzustellen, dass der Test tatsächlich das Merkmal Umgang mit chemischem Fachwissen der Klassen 7 und 8.1 erfasst. Das Verfahren wurde bereits bei der Validitätsprüfung des CL-Tests in Abschnitt 6.2.2 beschrieben. Die Aufgaben beider Testvarianten (A und B) wurden wieder durch einen erfahrenen Chemielehrer erstellt und anschließend in einem Expertenteam, bestehend aus zwei weiteren Chemielehrern mit mehrjähriger Berufserfahrung, besprochen. Dabei wurde diskutiert, inwieweit durch die Aufgaben chemisches Fachwissen adressiert wird. Für jede Aufgabe wurde das zur Lösung nötige Fachwissen benannt. Für jede der jeweils 42 Aufgaben aus den Tests A und B wurde festgestellt, dass sie Fachwissen adressieren und welcher Art dieses Wissen ist. Zur Klärung der Frage, inwieweit die drei Anforderungsbereiche durch die Tests abgebildet werden, wurden die einzelnen Testitems wieder im Rahmen eines Expertenratings dem Chemielehrerteam vorgelegt, mit der Bitte, jedes Item, so weit möglich, einem der drei Anforderungsbereiche zuzuordnen. Anschließend wurde die Interrater-Korrelation der drei Rater durch Berechnung von Fleiss' Kappa bestimmt (vgl. Bortz & Döring, 2002). Es zeigt sich, dass eine hoch signifikante Übereinstimmung vorliegt, die nach Landis & Koch (1977) im moderaten Bereich liegt ($Kappa = 0,505$, $r = 0,068$, $z = 7,477$, $p < 0,01$). Wie schon beim CL-Test, wird auch für den ACL-Test die Chemienote aus Klasse 7 als Außenkriterium zur Validierung des Tests herangezogen und die entsprechende Korrelation berechnet. Dadurch soll geprüft werden, inwieweit die Leistungen im Umgang mit Fachwissen der Klassen 7 und 8.1 der Chemienote aus Klasse 7 entsprechen. Die Korrelation war mit $p < 0,01$ hoch signifikant und lag mit $r = 0,419$ im mittleren bis guten Bereich.

6.3.3 Reliabilität

Wie schon beim CL-Test erfolgte die Prüfung der Reliabilität des Tests anhand der Berechnung von Cronbachs Alpha zur Bestimmung der inneren Konsistenz.

Anforderungsbereich	Test A	Test B
AFB I	0,721	0,686
AFB II	0,689	0,743
AFB III	0,705	0,503
AFB I, II, III	0,705	0,644

Abb. Innere Konsistenz des allgemeinen Leistungstests (Cronbachs Alpha)

Die Werte von Cronbachs Alpha liegen überwiegend im akzeptablen ($>0,7$) Bereich⁶¹. Der Wert für den Anforderungsbereich III in Test B muss jedoch als schlecht bezeichnet werden ($>0,5$). Dieser Teil enthält ausschließlich offene Aufgaben und muss im Rahmen der Pilotierung sorgfältig überarbeitet werden. Die ermittelten Werte für Cronbachs Alpha deuten darauf hin, dass die Testaufgaben innerhalb einer Testform von den Probanden ähnlich gelöst wurden. Dies gilt sowohl für die Aufgaben zu den einzelnen Anforderungsbereichen als auch für den Gesamttest.

Im nächsten Schritt wird überprüft, inwieweit sich die einzelnen Testitems im Hinblick auf ihre Schwierigkeit und Trennschärfe unterscheiden. Dazu werden, wie schon bei der Reliabilitätsanalyse des CL-Tests, Itemzwillinge aus den Testformen A und B betrachtet. Items mit geringen oder negativen Trennschärfen werden genauso überarbeitet wie Items, deren Schwierigkeiten sich in beiden Testformen deutlich unterschieden. Bei den Items, deren Trennschärfe unter 0,3 lag, wurde wieder zunächst auf die Schwierigkeit geguckt. War die mangelnde Trennschärfe auf den geringen bzw. hohen Schwierigkeitsgrad zurückzuführen, so wurden die Items, wie schon bei der Pilotierung des CL-Tests, nicht notwendigerweise überarbeitet, da der Test den gesamten Schwierigkeitsbereich abdecken soll. Nur Items, deren Trennschärfen nahe bei null ($TS < 0,1$) oder sogar negativ waren, wurden genauer auf Mängel in der Aufgabenstellung untersucht. Items mit Trennschärfen zwischen 0.1 und 0.3 wurden nicht verändert, da die mangelnden Trennschärfen darauf zurückzuführen waren, dass die Items leicht waren (LH zwischen 0.75 und 0.95). Der Tests ist deshalb tendenziell leicht bis mittel-schwer. Items mit geringerer Trennschärfe werden also in Kauf genommen, was bei der großen Zahl an Items (42 Items) vertretbar ist. Die restlichen Items wurden, wie schon beim CL-Test, im Rückgriff auf die Ergebnisse der Distraktorenanalyse und der mündlichen Rückmeldungen einzelner Schüler

⁶¹ Zur Bewertung von Cronbachs Alpha vergleiche George und Mallery (2002, S. 231).

im Hinblick auf mögliche Mängel überprüft und gegebenenfalls überarbeitet. Das Verfahren wurde bereits bei der Reliabilitätsprüfung des CL-Tests ausführlich beschrieben.

6.4 Wortschatztests

Die Pilotphase dient dazu, die eingesetzten Tests im Hinblick auf mögliche Mängel in Bezug auf Konstruktion, Durchführung und Auswertung zu untersuchen. Die Wortschatztests der Pilotphase unterschieden sich jedoch teilweise von den in der Hauptuntersuchung verwendeten. Dies liegt daran, dass die ursprüngliche Zielsetzung, nur den Erwerb des deutschen Fachwortschatzes zu untersuchen, erweitert wurde, mit dem Ziel, auch den Erwerb des fremdsprachlichen Wortschatzes zu erfassen und dabei zwischen Fach- und Allgemeinwortschatz zu differenzieren. Die Entwicklung der Wortschatztests vollzog sich dabei in mehreren Schritten. Die endgültige Fassung der Tests für die Hauptuntersuchung wurde bereits ausführlich in Abschnitt 5.3.2 erläutert. Die folgende Abbildung gibt einen Überblick über die unterschiedlichen Phasen der Testentwicklung.

Die Pilotphase

Nr.	Test	Begriffe	Wortumfang (rezeptiv/produktiv)
Pilotphase			
Pretests			
1a	Deutscher Fachwortschatz	Im Deutschen bereits eingeführte sowie noch nicht eingeführte Fachbegriffe	13/23
Posttests / Follow-Up-Test			
1b	Deutscher Fachwortschatz	Im Deutschen bereits eingeführte sowie noch nicht eingeführte Fachbegriffe	13/23
2	Englischer Fachwortschatz	Im Deutschen bereits eingeführte sowie noch nicht eingeführte Fachbegriffe	13/23
3	Zusätzlich erworbener englischer Wortschatz	Im Deutschen bereits eingeführte sowie noch nicht eingeführte Fachbegriffe und im Englischen noch nicht eingeführter Allgemeinwortschatz	13/23
Hauptuntersuchung			
1'	Deutscher Fachwortschatz	Im Deutschen noch nicht eingeführte Fachbegriffe	11/9
2'	Englischer Fachwortschatz	Im Deutschen noch nicht eingeführte Fachbegriffe	11/9
3'	Zusätzlich erworbener englischer Fachwortschatz	Im Deutschen bereits eingeführte Fachbegriffe	10/15
4'	Zusätzlich erworbener englischer Allgemeinwortschatz	Im Englischen noch nicht eingeführter Allgemeinwortschatz	10/13

Tabelle 22: Entwicklung der Wortschatztests

6.4.1 Güte der Tests: Objektivität, Validität und Reliabilität

Für alle Wortschatztests wurde eine Testleiterinstruktion verfasst, die eine möglichst hohe Durchführungsobjektivität gewährleisten soll. In den Tests kommen ausschließlich geschlossene (MC-Aufgaben) und halb-offene (Lückensätze) Aufgabenformate zum Einsatz. Für die halb-offenen Aufgaben wurde ein Lösungsbogen entwickelt, der alle korrekten Antwortalternativen beinhaltet. Es wurden nur die dort angegebenen Lösungen akzeptiert, fehlerhafte Rechtschreibung führte ebenfalls dazu, dass das jeweilige Wort als falsch gewertet wurde. Die Auswertungsobjektivität kann somit insgesamt als hoch gelten. Durch die Tests soll das Vokabelwissen der Schüler im Hinblick auf die rezeptive und die produktive Beherrschung überprüft werden. Das hierzu gewählte Verfahren der selektiven und kontextunabhängigen Testung zeichnet sich dabei laut Read (2011) durch eine hohe Augenscheinvalidität aus. Dies gilt insbesondere für die Tests zum produktiven Wortschatz (vgl. Read, 2011). Die Validität, der im Test zum rezeptiven Wortschatz eingesetzten MC-Aufgaben hängt jedoch stark von der Qualität der Distraktoren ab. Bei diesem Format könnten Probanden nämlich auch auf andere Strategien zurückgreifen, wie z.B. das Ausschlussverfahren. Deshalb wurden im Rahmen der Pilotierung auch Distraktorenanalysen durchgeführt. Für die Überarbeitung der Items waren zahlreiche Gründe verantwortlich. Zu diesen gehört neben dem Austausch ungeeigneter Distraktoren auch die Vereinheitlichung der sprachlichen Form oder das Anpassen des Itemformats. Darüber hinaus wurden auch die Lösungshäufigkeiten der Items untersucht. Items, die von fast keinem Probanden bzw. von fast allen gelöst werden konnten, wurden im Hinblick auf mögliche Konstruktionsmängel analysiert und gegebenenfalls modifiziert.

6.4.2 Deutscher Fachwortschatztest

Wie im vorangegangenen Abschnitt dargestellt, soll der deutsche Fachwortschatztests in der Hauptuntersuchung nur solche Fachbegriffe umfassen, die im Rahmen des Chemieunterrichts der Klasse 7 noch nicht eingeführt wurden. Bereits eingeführte Fachbegriffe wurden deshalb aus dem Test gestrichen. Für Items mit einer hohen bzw. sehr geringen Lösungshäufigkeit wurde geprüft, ob die Werte auf Mängel in der Itemformulierung zurückzuführen sind. Betrachtet man die Pretestergebnisse der monolingual sowie der bilingual unterrichteten Gruppe, so fällt auf, dass im rezeptiven Teil, in welchem die Bedeutung der Begriffe durch Multiple-Choice-Aufgaben abgefragt wurde, acht der 13 neu eingeführten Begriffe einer deutlich größeren Zahl von Schüler bekannt waren als es der Ratewahrscheinlichkeit von 25 Prozent entspricht. Auch im produktiven Teil fielen drei Items in beiden Gruppen durch besonders hohe Lösungshäufigkeiten auf (LH zwischen 42 und 84 Prozent), obwohl diese Begriffe noch nicht im Unterricht eingeführt worden waren. Somit

konnten 11 der insgesamt 23 Items bereits im Pretest von einer großen Zahl von Schülern gelöst werden. Bei den restlichen 12 Items entsprach die Lösungshäufigkeiten den Erwartungen. Sie lag bei den Multiple-Choice-Items nahe der Ratewahrscheinlichkeit (mittlere LH = 22,8 Prozent) bzw. deutlich darunter im produktiven Teil (mittlere LH = 13,4 Prozent). Betrachtet man die besonders häufig richtig gelösten Items etwas genauer, so fällt auf, dass es in 10 der insgesamt 11 Fällen um Bezeichnungen für Alkali- und Erdalkalimetalle und ihrer Verbindungen geht. Nach Durchsicht der Aufzeichnungen zur Durchführung der Pretests, fiel auf, dass in der monolingual sowie in der bilingual unterrichteten Gruppe der Sachfachtest zum Modul vor dem Test zum deutschen Fachwortschatz durchgeführt worden war. Im Informationstext zum zweiten Teil des Sachfachtests werden die Stoffnamen der Alkali- und Erdalkalimetalle, inklusive Elementsymbolen genannt. Um Pannen dieser Art im Rahmen der Hauptuntersuchung zu verhindern, wurde ein Ablaufplan, welcher die zeitliche Abfolge aller Tests enthält, erstellt. Für die Überarbeitung der anderen Items waren die Gründe unterschiedlich. Die folgende Abbildung zeigt das Vorgehen anhand von Item 25.

25	_____ sind Metalle. (Metalle der 1. Hauptgruppe des PSE)
	Die Formulierung wurde präzisiert und es wurde ein weiterer inhaltlicher Hinweis gegeben, da die Lösungshäufigkeit im Posttest, trotz der zentralen Bedeutung des Begriffs, zu den geringsten gehörte.
	Die _____ zählen zu den reaktivsten Metallen (Oberbegriff für die Metalle der 1. Hauptgruppe des PSE).

Abbildung 33 Überarbeitung von Item 25 des deutschen Fachwortschatztests

Von den ursprünglich 36 Items des DFW-Tests wurden 17 Items gestrichen, 12 Items modifiziert übernommen, sieben Items ohne Modifikation übernommen und eines wurde hinzugefügt. Der Test für die Hauptuntersuchung enthält 20 Items.

6.4.3 Englischer Fachwortschatztest

Der Test zum englischen Fachwortschatz wurde so überarbeitet, dass er in der Auswahl der Vokabeln sowie im Format der Items dem deutschen Fachwortschatz entsprach. Dazu wurden Items gestrichen oder sie wechselten in den Test zum zusätzlich erworbenen Wortschatz. Darüber hinaus waren die Gründe für die Überarbeitung der einzelnen Items wieder unterschiedlich. Die folgende Abbildung verdeutlicht das Vorgehen anhand von Item 4.

Die Pilotphase

4	They needed a ... for the experiment. Gesucht ist ein Löffel, mit dem Natrium unter Wasser fixiert werden kann. A sieve spoon B sodium spoon C sodium sieve D spoon sieve
	Das Itemformat wird an das Format des deutschen FWS-Tests angepasst, um die Vergleichbarkeit der Testergebnisse zu erhöhen. Es wird nicht mehr der korrekte Begriff aus vier Alternativen ausgewählt, sondern ein Merkmal bestimmt, das zeigt, ob der Begriff bekannt ist.
	They needed a sodium spoon. Der Begriff sodium spoon bezeichnet einen Löffel zum A Abfüllen von Natrium. B Fixieren von Natrium. C Verbrennen von Natrium. D Zerkleinern von Natrium.

Abbildung 34 Überarbeitung von Item 4 des englischen Fachwortschatztests

Zusätzlich wurden die Lösungshäufigkeiten der Items untersucht. Da kein Pretest durchgeführt worden war, wurden die Daten des Posttests ausgewertet⁶². Die insgesamt hohen Lösungshäufigkeiten (LH (EFWT) = 0,69) deuten darauf hin, dass die Items geeignet sind, den Wortschatzerwerb der Schüler abzubilden. Für Items mit einer geringen Lösungshäufigkeit wurde geprüft, ob die Werte auf Mängel in der Itemformulierung zurückzuführen sind. Hier fielen zwei Items auf, denn sie wurden von weniger als einem Drittel der Schüler gelöst. Die entsprechenden Begriffe waren im Rahmen des Moduls neu eingeführt worden, und den Schülern vor Beginn des Moduls unbekannt. Bei der Durchsicht der Unterrichtsmaterialien fiel auf, dass die entsprechenden Fachbegriffe auf den Arbeitsblättern nicht auftauchten. Die Arbeitsblätter und die Stundenentwürfe wurden entsprechend geändert. Von den ursprünglich 36 Items des EFW-Tests wurden elf Items in den Test zum zusätzlich erworbenen Wortschatz verschoben, fünf Items gestrichen, 14 Items modifiziert übernommen und sechs Items ohne Modifikation übernommen. Der Test für die Hauptuntersuchung enthält 20 Items.

6.4.4 Test zum zusätzlich erworbenen Fach- und Allgemeinwortschatz

Der Test zum zusätzlich erworbenen Wortschatz wurde für die Hauptuntersuchung so überarbeitet, so dass er nun aus einem Fach- (Test 3) und einem Allgemeinwortschatzteil (Test 4) besteht. Dabei wechselten Fachbegriffe aus dem EFW-Test in Test 3, wenn diese im Deutschen bereits bekannt waren. Begriffe der Allgemeinsprache, die - wie sich im Verlauf des Moduls zeigte - besonders hilfreich zur Erarbeitung der sachfachlichen Inhalte waren, erschienen nun in Test 4.

⁶² Die Entscheidung, auch den englischen Wortschatz zu testen, fiel erst im Verlauf des Moduls (siehe Abschnitt 6.4).

Wieder wurden die Lösungshäufigkeiten der Items des Post-Tests untersucht. Für Items mit einer geringen Lösungshäufigkeit wurde geprüft, ob die Werte auf Mängel in der Itemformulierung zurückzuführen sind. Darüber hinaus waren die Gründe für die Überarbeitung der einzelnen Items sehr unterschiedlich. In vielen Fällen konnte auf die Ergebnisse der Distraktorenanalyse zurückgegriffen werden, wie das folgende Beispiel zeigt.

6	Different compounds of carbon (Kohlenstoff) were analysed. Das Wort compound bedeutet: A Gemisch B Verbindung C Ausprägung D Einheit
	Die Distraktorenanalyse ergab, dass die Antworten C und D nicht gewählt wurden. Das lag vermutlich daran, dass sie den Schülern nicht als chemische Fachbegriffe bekannt sind. Sie wurden deshalb ersetzt und der Beispielsatz angepasst.
	Different compounds were analysed. Das Wort compound bedeutet: A Gemisch B Verbindung C Stoff D Probe

Abbildung 35 Überarbeitung von Item 6 des ZEFWT

Von den ursprünglich 36 Items des Tests zum zusätzlich erworbenen Fachwortschatz wurden sieben Items gestrichen. Elf Items wechselten aus dem Test zum englischen Fachwortschatz in diesen Test. Außerdem wurden acht Items ergänzt, da sie sich im Verlaufe der Unterrichtsreihe als besonders nützlich erwiesen hatten. Acht Items wurden modifiziert übernommen und 19 Items ohne Modifikation übernommen. Der Test für die Hauptuntersuchung enthält 48 Items.

6.5 Fragebögen

Im Rahmen der Pilotierung werden die Fragebögen selbst, aber auch die Durchführung und Auswertung der Datenerhebung mittels Fragebogen im Hinblick auf mögliche Mängel untersucht. Es geht darum, Schwächen bei der Formulierung der Items, der Arbeitsanweisungen oder der Instruktionen zum Ablauf der Testung aufzudecken. Außerdem wird die Codierung der Ergebnisse der Rating-Items (Likert-Skalen) untersucht. Die Bewertung der Testgüte orientiert sich auch hier an den klassischen Gütemerkmalen Objektivität, Validität und Reliabilität. In diesem Zusammenhang wird insbesondere geprüft, ob die vorgeschlagenen Subskalen durch eine Analyse der inneren Konsistenz bestätigt werden können. Dazu wurden sowohl Cronbachs Alpha als auch die Trennschärfe der Items bzw. der Skala berechnet⁶³. Die Pilotphase dient aber jedoch dazu, einen

⁶³ Die theoriegeleitete Subskalen-Bildung wurde dem faktorenanalytischen Verfahren an dieser Stelle vorgezogen. Rumlich (2016, S. 280) hebt in diesem Zusammenhang hervor, dass die Faktorenanalyse nur bei hinreichend

ersten Einblick in das Forschungsfeld zu erhalten, etwa um Fragestellungen weiter zu präzisieren oder den Einfluss weiterer Störvariablen abzubilden. Die Fragebögen der Pilotphase sind deshalb etwas anders aufgebaut als die Fragebögen der Hauptuntersuchung. Der folgende Abschnitt gibt hierzu einen Überblick.

6.5.1 Aufbau der Fragebögen

Die Fragebögen wurden im Rahmen der Pilotierung überarbeitet. Dabei kam es vor, dass bestimmte Items gestrichen bzw. neue Items ergänzt wurden.

Prefragebogen UG (FB1)	
Deckblatt	Erzeugen des Code-Namens
Teil 1	Einstellung der Schüler
	Die Fächer
	Fachinteresse Chemie
	Fachinteresse Englisch
Teil 2	Der bilinguale Unterricht
	Interesse (Erwartung)
	Nutzen (Erwartung)
	Kompetenzerleben (Erwartung)
	Mehrarbeit (Erwartung)
	Demografische Angaben
	Geschlecht
	Alter
	Chemienote
	Englischnote
	Externe Hilfe

Abbildung 36: Aufbau des Prefragebogens der CLIL-Gruppe

großen Stichproben zuverlässige Werte liefert, wobei Field (2009, S. 647) Stichprobengrößen von mehr als 300 Probanden vorschlägt. Hinzu kommt, dass die rein mathematische Ermittlung einzelner Faktoren auf Zufallsstichproben beruht, während im vorliegenden Fall einzelne Klassen untersucht wurden. Die Erhebung klassenbezogener Daten stellt jedoch keine Zufalls- sondern eine sogenannte *Ad-hoc-Stichprobe* dar.

Postfragebogen UG (FB2)	
Deckblatt	Erzeugen des Code-Namens
	Einstellung der Schüler
	Der bilinguale Unterricht
	Interesse
	Nutzen
	Kompetenzerleben
	Mehrarbeit
	Schülerurteil
	Fazit zum bilingualen Unterricht

Abbildung 37: Aufbau des Postfragebogens der CLIL-Gruppe

Die Pre- und Postfragebögen für die Kontrollgruppe sind weitgehend identisch aufgebaut. Es gibt insgesamt drei Unterschiede:

1. Items zum Interesse im Fach Englisch entfallen in der Kontrollgruppe.
2. Im Pre- und Postfragebogen der Kontrollgruppe beziehen sich die Fragen zur intrinsischen Motivation nicht auf den bilingualen Unterricht, sondern auf den deutschsprachigen Chemieunterricht bzw. das auf Deutsch durchgeführte Modul.
3. Beim Fazit wird anstelle der Frage danach, ob man den bilingualen Modulunterricht im Fach Chemie wiederholen solle, gefragt, inwieweit sich das Modul vom ansonsten erteilten Chemieunterricht unterschied (Schülerurteil).

Eine weitere Änderung betrifft die zur Erfassung der Motivation verwendete Subskala *Druck*. In seiner ursprünglichen Fassung sollte an dieser Stelle die mögliche *Mehrarbeit* erfasst werden, die Schüler leisten müssen, wenn sie in der Fremdsprache unterrichtet werden. Es wurde vermutet, dass sich die Mehrarbeit negativ auf die Motivation der Schüler auswirkt. Mehrarbeit an sich muss sich jedoch nicht zwangsläufig negativ auswirken, sie kann, je nach Schüler, auch motivierend wirken. Die Daten zur Mehrarbeit ließen sich deshalb nicht zweifelsfrei im Hinblick auf ihre motivationalen Auswirkungen auswerten. Für die Hauptuntersuchung wurde diese Subskala durch die Frage nach dem subjektiv empfundenen Druck ersetzt. Diese wirkt sich negativ auf die Motivation der Schüler aus, wie Untersuchungen von Deci und Ryan zeigen. Damit wird neben dem *Interesse*, dem *Kompetenzerleben* und dem *Nutzen* ein weiterer Faktor berücksichtigt, der auch im Rahmen der Selbstbestimmungstheorie von Deci und Ryan die Motivation der Schüler beeinflusst, nämlich der Faktor *Druck / Anspannung*. Die Items wurden in Anlehnung an die Items der KIM zum empfundenen Druck umformuliert. Auf eine Analyse der Skala Mehrarbeit wurde im Rahmen der

Pilotierung verzichtet. Die Reliabilitätsprüfung erfolgt über die Berechnung von Cronbachs Alpha und der Trennschärfe im Rahmen der Hauptuntersuchung. Die Werte waren akzeptabel (vgl. Abschnitt 7.5.1). Folgende Abbildung zeigt das Vorgehen am Beispiel des Postfragebogens der Untersuchungsgruppe.

Mehrarbeit durch CAE

Für CAE habe ich auch nicht mehr lernen müssen als vorher.

In CAE musste ich viel mehr arbeiten.

Für CAE brauchte ich nicht mehr tun als vorher.

CAE bedeutet, dass ich mehr üben musste als vorher.

Druck/Anspannung durch CAE

CAE war nicht so stressig.

CAE war relativ locker.

CAE war anstrengend.

Ich fand CAE anstrengend.

Abbildung 38: Ersetzten der Skala Mehrarbeit durch die Skala Druck

Es zeigte sich, dass das fachliche Interesse der Schüler einen Einfluss auf ihre Motivation im bilingualen Unterricht haben kann (vgl. Abschnitt 4.3.2). Deshalb sollen die Daten in der Hauptuntersuchung auch nach Interessensgruppen differenziert untersucht werden. Es wurde ein entsprechendes Item auf dem Prefragebogen ergänzt. Dabei werden vier Gruppen von Schülern betrachtet: Fremdsprachlich orientierte Schüler, naturwissenschaftlich orientierte Schüler, sowohl fremdsprachlich als auch naturwissenschaftlich interessierte Schüler sowie weder fremdsprachlich noch naturwissenschaftlich interessierte Schüler⁶⁴. Weil auch das fachspezifische Selbstkonzept für den Lernerfolg und die Motivation der Schüler eine wichtige Rolle spielt, wird es im Rahmen der Hauptuntersuchung durch zusätzliche Items auf den Fragebögen beider Gruppen erfasst (siehe Abschnitt 5.4.5). Außerdem soll in der Hauptuntersuchung das Bildungsniveau des Elternhauses berücksichtigt werden. In zahlreichen Untersuchungen konnte ein Zusammenhang zwischen dem Bildungsniveau und dem Lernerfolg nachgewiesen werden (vgl. Abschnitt 5.4.6). Ein entsprechendes Item wurde auf dem Postfragebogen ergänzt. Mit Einführung internationaler Klassen an der Schule hat außerdem der Anteil von Schülern zugenommen, deren Muttersprache

64 Zum Verfahren der Einteilung in Interessensgruppen siehe Abschnitt 5.2.4.

nicht Deutsch ist. Dabei handelt es sich überwiegend um Kinder, deren Eltern in einem benachbarten Militärstützpunkt der NATO Dienst tun. Die Schüler sind in der Regel für drei bis vier Jahre in Deutschland und verfügen, je nach Dauer ihres Aufenthaltes, nur über geringe Deutschkenntnisse. Solche, sich aus dem Migrationshintergrund ergebenden L1-Defizite, können sich jedoch sowohl auf den Lernerfolg im monolingual auf Deutsch als auch im bilingual geführten Unterricht auswirken, weil er auch Phasen enthält, in denen auf die Muttersprache zurückgegriffen wird. Um den Einfluss der Deutschkenntnisse zu kontrollieren, werden diese bzw. ein möglicher Migrationshintergrund über den Postfragebogen erfasst. Die Änderungen der Fragebögen der Kontrollgruppe für die Hauptuntersuchung erfolgte analog. Ein Überblick über die Endfassung der Fragebögen für die Hauptuntersuchung findet sich in Abschnitt 5.3.3.3.

6.5.2 Objektivität, Validität und Reliabilität

Die Fragebogenerhebung wurden nicht vom Untersuchungsleiter, der gleichzeitig der Chemielehrer in den Klassen war, durchgeführt, sondern von den jeweiligen Klassenlehrern. Damit soll vermieden werden, dass Schüler insbesondere bei Fragen zum Chemieunterricht durch die Anwesenheit des Fachlehrers beeinflusst werden. Um ein gewisses Maß an Objektivität bei der Durchführung zu gewährleisten, erhielten die Klassenlehrer vom Untersuchungsleiter persönlich Instruktionen zur Durchführung der Befragung. Sie wurden gebeten, die Informationen auf dem Deckblatt sowie die Einleitung inklusive Beispiel vor Beginn der Bearbeitung vorzulesen. Die Fragebogenitems sind bis auf die Frage nach dem Fazit zum bilingualen Unterricht geschlossene Items, sodass die Auswertungsobjektivität als hoch eingestuft werden muss. Für die Beschreibung der Skalen und die numerische Codierung der Rating-Aufgaben gibt es ebenso wie für die mathematische Auswertung (z.B. Mittelwertbildung) eindeutige Vorgaben (Interpretationsobjektivität)⁶⁵. Die Objektivität der Erhebung durch Schülerfragebögen kann demnach als insgesamt hoch eingestuft werden. Zur Prüfung auf Reliabilität wurden die innere Konsistenz der Subskalen durch Berechnung von Cronbachs Alpha sowie die Trennschärfen der einzelnen Items ermittelt. Dabei galten für Cronbachs Alpha Werte über 0,7 und für die Trennschärfe oberhalb von 0,3 als akzeptabel (vgl. Dörnyei, 2001, S. 204; Prüfer, 2012, S. 145). Die Berechnung von Cronbachs Alpha ergab, dass alle fünf Skalen des Prefragebogens akzeptable Reliabilitäten (Alpha zwischen 0,703 und 0,816) aufwiesen. Auch die Trennschärfen der Items dieser fünf Subskalen lagen im akzeptablen Bereich (r zwischen 0,387 und 0,778).

⁶⁵ Siehe hierzu Abschnitt 5.3.3.5.

Nr.	Subskala	Cronbachs Alpha	Trennschärfen
1	Interesse am Fach Chemie (FICH)	0,703	0,387-0,665
2	Interesse am Fach Englisch (FIEN)	0,765	0,397-0,682
3	Interesse BU (Erwartung) (EIBU)	0,741	0,425-0,684
4	Nutzen BU (Erwartung) (ENBU)	0,816	0,558-0,778
5	Kompetenzerleben BU (Erwartung) (EKBU)	0,774	0,520-0,669

Abb. Cronbachs Alpha und Trennschärfen der Skalen des Prefragebogens (N=31)

Es zeigte sich außerdem, dass die Streichung einzelner Items keine nennenswerte Verbesserung der inneren Konsistenz zur Folge hat. Für den Postfragebogen ergab die Berechnung von Cronbachs Alpha, dass nur die Skala Nutzen BU eine akzeptable Reliabilität ($\alpha = 0,706$) aufwies und die Werte bei den Skalen Interesse BU und Kompetenzerleben BU unterhalb der Grenze von 0.7 blieben ($\alpha = 0,548$ und $\alpha = 0,693$). Die Trennschärfen der Items lagen für die Skalen Nutzen BU und Kompetenzerleben BU im akzeptablen Bereich (r zwischen 0,338 und 0,674), während einzelne Items der Skala Interesse BU durch mangelnde Trennschärfe auffielen.

Nr.	Subskala	Cronbachs Alpha	Trennschärfen
1	Interesse bilingualer Unterricht (IBU)	0,548	0,033-0,628
2	Nutzen bilingualer Unterricht (NBU)	0,706	0,444-0,541
3	Kompetenzerleben bilingualer Unterricht (KEBU)	0,693	0,338-0,642

Abb. Cronbachs Alpha und Trennschärfen der drei Subskalen des Postfragebogens (N=29)

Die problematischen Skalen Interesse BU und Kompetenzerleben KBU wurden deshalb überarbeitet. Auf eine Streichung der betreffenden Items zur Verbesserung der inneren Konsistenz einer Subskala wurde verzichtet, da die Reliabilität durch die Reduzierung der Items leidet. Laut Bühner (2006, S. 192) sollte eine Subskala nämlich aus mindestens drei, besser, aber vier Items bestehen. Die Kennwerte der Items bzw. der Skalen müssen deshalb in der Hauptuntersuchung erneut überprüft werden.

6.5.3 Das Schülerfazit

Neben der Optimierung der Fragebögen selbst, sowie des Ablaufs der Datenerhebung und ihrer Auswertung, dient die Pilotphase auch dazu, erste Einblicke in das Forschungsfeld zu erhalten.

Hier lieferte das Fazit am Ende des Postfragebogens interessante Erkenntnisse. Die Schüler sollten hier angeben, ob sie eine solche bilinguale Unterrichtsreihe gerne wiederholen würden. Das offene Format der Frage am Ende des Postfragebogens ermöglichte es ihnen, sich ohne Lenkung mit den Vor- und Nachteilen bilingualen Unterrichts auseinanderzusetzen. Es zeigte sich, dass die Antworten eine sinnvolle Ergänzung zu den über Rating-Skalen erhobenen Daten zu Nutzen und Kompetenzerleben darstellten. Außerdem lieferten sie Rückmeldungen über mögliche Mängel in der methodischen Durchführung des Unterrichts. Dies soll anhand einiger Beispiele verdeutlicht werden. Auf die Frage, ob man eine solche Unterrichtsreihe auf Englisch häufiger machen sollte, antwortete die große Mehrheit der Schüler (86 %) mit „ja“. Sie begründeten dies damit, dass sie mehr lernen, nämlich in Chemie und Englisch. Gerade Englischkenntnisse und die Möglichkeit, diese zum Wissenserwerb einzusetzen, wurden von einigen Schülern als wichtige Voraussetzung für späteren Erfolg im Studium bewertet. Mehrere Schüler gaben an, dass sich durch den bilingualen Unterricht auch ihre Sprechfertigkeiten verbesserten und sie neue Vokabeln lernten. Neunzig Prozent der Schüler gaben an, keine Verständnisschwierigkeiten gehabt zu haben und ein Schüler war der Meinung, sich in dieser Form des Unterrichts intensiver mit den Inhalten auseinanderzusetzen als im regulären Chemieunterricht. Auch die stärkere Veranschaulichung und die Einübung durch Arbeitsblätter wurden hervorgehoben. Lediglich zwei Schüler (7 %) lehnten eine Wiederholung ab, weil sie vermuteten, dass schwächere Schüler Verständnisschwierigkeiten haben könnten. Auch die offene Frage am Ende des Postfragebogens der Kontrollgruppe zu Unterschieden zwischen dem regulären Chemieunterricht und der Reihe erwies sich als nützlich, da sie Hinweise darauf gibt, ob eine höhere Motivation im bilingualen Modul weniger auf den Einsatz der Fremdsprache als auf die besonders interessante Gestaltung des Unterrichts z.B. durch zahlreiche chemische Experimente zurückzuführen ist. Auf die Frage, ob der Unterricht im Modul anders war als die Schüler ihn sonst gewohnt waren, antworteten dazu 79 % mit "nein". Einige Schüler hoben jedoch die größere Zahl an Experimenten, und insbesondere das Schülerexperiment positiv hervor, waren sich aber uneinig, ob dies in vorangegangenen Reihen anders war.

7 Hauptuntersuchung: Ergebnisse und Auswertung

Im folgenden Abschnitt werden die Ergebnisse zu den drei Forschungsschwerpunkten der Studie vorgestellt. Diese betreffen den Bereich des sachfachlichen Lernens (Abschnitt 7.3), des (fremd-)sprachlichen Lernens (Abschnitt 7.4) sowie der Motivation (Abschnitt 7.5). Weil es sich bei den untersuchten Stichproben um Klumpenstichproben handelt, welche nicht vollständig randomisiert sind, müssen die Stichproben zunächst miteinander verglichen werden (Abschnitt 7.1). Nur wenn

sie in Bezug auf bestimmte Kontrollvariablen ähnlich sind, können die hier gewonnenen Daten verglichen werden. Gleiches gilt für die einzelnen Interessensgruppen, die in Abschnitt 7.2 beschrieben werden.

7.1 Die Stichproben

An der Hauptuntersuchung, die zwischen 2018 und 2021 an einem Gymnasium in Nordrhein-Westfalen stattfand, nahmen insgesamt 149 Schüler aus sechs achten Klassen teil. Es wurden solche Klassen ausgewählt, in denen der Untersuchungsleiter zu diesem Zeitpunkt unterrichtete. Zwar wurde der Klassenverband im Rahmen der Untersuchung beibehalten, bei der Verteilung der Klassen auf Untersuchungs- und Kontrollgruppe jedoch darauf geachtet, dass beide Gruppen gleich groß sind und der Anteil von Mädchen und Jungen ebenfalls etwa gleich groß ist. Weil eine solche *Klumpenstichprobe* nicht vollständig randomisiert ist, wurde außerdem der Versuch einer Parallelisierung beider Gruppen unternommen (Bortz & Döring, 2002, S. 438). Dazu wurde bei der Verteilung der Klassen auf die Gruppen zusätzlich darauf geachtet, dass sich diese im Hinblick auf mögliche Störvariablen wenig unterscheiden. Nach Bortz & Döring (2002, S. 527) gelten zwei Gruppen dann als parallel, wenn sie „bezüglich der Störvariablen annähernd gleiche Mittelwerte und Streuungen aufweisen“. Die potenziellen Störvariablen, die im Rahmen der Untersuchung kontrolliert werden, lassen sich vier Bereichen zuordnen: Demografische Merkmale, Leistungsvermögen, Interesse und Selbstkonzept.

7.1.1 Merkmale der Stichproben

Neben dem Alter und dem Geschlecht hat sich insbesondere für den naturwissenschaftlichen Unterricht gezeigt, dass sowohl das Bildungsniveau des Elternhauses sowie die Beherrschung der Unterrichtssprache einen Einfluss auf den Lernerfolg haben (vgl. Abschnitt 5.4.6). Weil immer mehr Schüler auch externe Hilfe in Form von Nachhilfeunterricht in Anspruch nehmen, wurde auch dieser Faktor erhoben.

7.1.2 Demografische Merkmale

Gruppe	N	Alter M (SD,SE)	Geschlecht		Akademiker- haushalt	Migrationshintergrund		Externe Hilfe
			Mädchen	Jungen		gesamt	wenig Deutsch	
CLIL	74	13,38 (0,54;0,06)	57,53%	42,47%	41,43%	30,43%	0,00%	0,00%
nicht-CLIL	75	13,20 (1,56;0,19)	57,35%	42,65%	28,57%	37,31%	6,70%	1,45%
gesamt	149	13,29 (1,16;0,10)	57,45%	42,55%	35,00%	33,82%	3,36%	0,71%

Tabelle 23: Demografische Angaben zu Untersuchungs- und Kontrollgruppe⁶⁶

Die CLIL-Schüler sind im Mittel 13,38 Jahre alt und damit geringfügig älter als die nicht-CLIL-Schüler. Der Unterschied ist jedoch minimal ($\Delta M = 0,18$) und statistisch nicht signifikant, wie der t-Test ($t(142) = 0,960$, $p = 0,339$) zeigt⁶⁷. In beiden Gruppen ist der Anteil der Mädchen etwa 15 % größer als der Anteil der Jungen, wobei der Mädchenanteil in der CLIL-Gruppe minimal größer ist als in der nicht-CLIL-Gruppe ($\Delta = 0,18$ %). Der Unterschied ist jedoch auch hier gemäß Chi-Quadrat-Test ($\chi^2 = 0,000473$, $p = 0,983$) statistisch nicht signifikant. Beide Gruppen spiegeln damit den leichten Mädchenüberhang an der Schule wider, der sich aus der Geschichte der Schule als ursprünglich reine Mädchenschule erklärt. Der Mädchenanteil ist um 4 % größer als an den Gymnasien des Landes Nordrhein-Westfalen im gleichen Zeitraum. Ihr Anteil betrug dort im Mittel 53,0 % (MSW-NRW, 2021). In beiden Gruppen überwiegt der Anteil der Schüler, die nicht aus einem Akademikerhaushalt kommen. In der CLIL-Gruppe sind das 41,43 % und damit 12,86 % mehr als in der nicht-CLIL-Gruppe. Der Chi-Quadrat-Test zeigt jedoch, dass dieser Unterschied statistisch nicht signifikant ist ($\chi^2 = 2,543$; $p = 0,111$). Der Anteil der Schüler mit Migrationshintergrund macht insgesamt etwa ein Drittel aller Schüler aus und ist in der nicht-CLIL-Gruppe um etwa 7 % größer als in der CLIL-Gruppe ($\Delta = 6,88$ %). Auch hier erweist sich der Unterschied aufgrund des Chi-Quadrat-Tests jedoch statistisch als nicht signifikant ($\chi^2 = 0,719$; $p = 0,397$). Die Gruppe der Schüler mit Migrationshintergrund ist in Bezug auf die Deutschkenntnisse jedoch sehr heterogen. Für die Mehrheit der Schüler ist Deutsch ihre Muttersprache. Der Anteil der

66 Sämtliche Daten wurden über die Fragebögen erhoben. Der Anteil fehlender Werte war insgesamt gering und lag zwischen 3 % (Alter) und 9 % (Migrationshintergrund). Neben der Probandenzahl (N) und dem arithmetischen Mittel (M) sind in den folgenden Tabellen in der Regel die Standardabweichung (SD) und der Standardfehler (SE) angegeben.

67 Die Varianzhomogenität wurde mithilfe des Levene-Tests überprüft ($F = 0,417$, $p = 0,519$). Der Kolmogorov-Smirnov-Test ($Z(144) = 0,373$, $p < 0,01$) lies zwar keine Normalverteilung der Daten erkennen, Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend der t-Test der nonparametrischen Alternative vorgezogen.

Schüler, die über mangelnde Deutschkenntnisse verfügen, ist deutlich geringer. Er wurde über die Zahl der Schüler ermittelt, die zum Zeitpunkt der Untersuchung an einer internationalen Klasse teilnehmen. Diese Klassen sind an der Schule für Schüler mit Migrationshintergrund eingerichtet, die über mangelnde Deutschkenntnisse verfügen. Der Anteil dieser Schüler ist insgesamt gering und liegt im Mittel bei etwa 3 %. Die Schüler befinden sich alle in der nicht-CLIL-Gruppe, hier machen sie knapp 7 % der Schüler aus. Der Anteil der Schüler, die im Fach Chemie nach eigener Auskunft auf externe Hilfe zurückgreifen, liegt in beiden Fällen nahe null. Ein statistisch signifikanter Unterschied liegt auch hier nicht vor ($\chi^2 = 1,051$; $p = 0,305$).

7.1.3 Leistungsvermögen

Der kognitive Fähigkeitstest (Heller & Perleth, 2000) wurde von insgesamt 146 Schülern bearbeitet. Von diesen erhielten 55,5 % Testform A und 44,5 % Testform B. Die Werte für Cronbachs Alpha betragen für die insgesamt 70 Items in Test A $\alpha = 0,858$ und in Test B $\alpha = 0,888$. Die durchschnittliche Itemschwierigkeit beträgt für Testform A $p = 0,64$ und für Testform B $p = 0,57$ und liegt damit im mittleren Bereich zwischen 0,2 und 0,8 (vgl. Bortz & Döring, 2002, S. 218). Etwa ein Drittel der Items der Testform A (35,7 %) liegen mit $p_{\min} = 0,07$ und $p_{\max} = 0,93$ außerhalb dieses Bereiches. Bei Testform B sind es 12,9 % der Items mit $p_{\min} = 0,15$ und $p_{\max} = 0,94$. Damit war Test B insgesamt etwas schwieriger als Test A. Besonders schwierige bzw. leichte Items haben zwar geringe Trennschärfen zur Folge, sind aber dennoch hier erwünscht, da sie dazu beitragen, zwischen besonders leistungsstarken bzw. besonders schwachen Probanden zu differenzieren. Sie werden deshalb im Test belassen. Lediglich Items mit negativen Trennschärfen werden entfernt (vgl. Rumlich 2016, S. 273). Die mittleren Trennschärfen betragen $r = 0,27$ für Testform A und $r = 0,30$ für Testform B und liegen damit im mittelmäßigen Bereich (vgl. Bortz und Döring, 2002, S. 219). Drei Items des verbalen Subtests haben jedoch negative Trennschärfen (V18A, V28B und V29A) und werden deshalb aus dem Test entfernt. Im Mittel haben die Probanden, die Testform A bearbeitet haben, mehr Punkte erzielt ($M = 43,2$, $SD = 9,3$) als diejenigen, die Testform B bearbeitet haben ($M = 38,5$, $SD = 10,8$). Das Ergebnis stimmt mit der insgesamt etwas größeren Schwierigkeit von Testform A überein. Der Mittelwertsunterschied ist statistisch signifikant ($t(144) = 2,806$, $p =$

0,006)⁶⁸. Er stellt jedoch nach Cohen (1988) nur einen kleinen Effekt dar (Cohen's $d = 0,47$). Daher werden beide Testformen bei der weiteren Auswertung als ein Gesamttest betrachtet⁶⁹.

Gruppe	N	V3 M (SD; SE)	Q1 M (SD; SE)	N1 M (SD; SE)	KFT M (SD; SE)
CLIL	72	10,4 (3,2;0,4)	14,1 (4,4;0,5)	17,0 (6,4;0,8)	41,5 (10,1;1,2)
nicht-CLIL	74	10,3 (3,4;0,4)	14,1 (4,0;0,5)	16,3 (7,0;0,8)	40,7 (10,4;1,2)
gesamt	146	10,3 (3,3;0,3)	14,1 (4,2;0,3)	16,6 (6,7;0,6)	41,1 (10,2;0,8)

Tabelle 24: Deskriptive Statistik zum kognitiven Fähigkeitstest

Die CLIL-Schüler schnitten im Gesamttest (KFT) insgesamt etwas besser ab als die nicht-CLIL Schüler ($\Delta M = 0,8$). Diese Überlegenheit war auch in den Subtests zum verbalen Denken (V3: $\Delta M = 0,1$) und insbesondere im Subtest zum non-verbalen Denken (N1: $\Delta M = 0,7$) zu beobachten. Die durchgeführten t-Tests zeigten jedoch, dass keiner der beobachteten Mittelwertsunterschiede statistisch signifikant war (KFT: $t(144) = 0,486$, $p = 0,630$; V3: $t(144) = 0,241$, $p = 0,810$; Q1: $t(144) = -0,113$, $p = 0,910$; N1: $t(144) = 0,696$; $p = 0,488$)⁷⁰. Der Lesegeschwindigkeits- und -verständnistest für die Klassen 5 – 12 (LGVT 5-12+) wurde von insgesamt 146 Schülern bearbeitet. Von diesen erhielten 52,4 % Testform A und 47,6 % Testform B. Eine Reliabilitätsprüfung über Cronbachs Alpha und die Trennschärfen der Items ist nicht sinnvoll, da es sich um einen Speed-Test handelt. Die Differenzierung zwischen den Schülern erfolgt dabei nicht über die Aufgabenschwierigkeit, sondern über die Anzahl der gelösten Aufgaben. Die von den Autoren ermittelten Retest- und Paralleltestreliabilität deuten auf eine hohe Reliabilität hin (vgl. Schneider et

68 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests ($F = 2,546$, $p = 0,113$), das Vorliegen der Normalverteilung der KFT-Daten anhand des Kolmogorov-Smirnov-Tests ($Z(146) = 0,069$, $p = 0,088$) bestätigt.

69 Für den Vergleich der Ergebnisse in CLIL- und nicht-CLIL-Gruppe ist der Unterschied insofern unerheblich, da der Anteil beider Testformen in den Gruppen nahezu identisch ist.

70 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (KFT: $F = 0,38$, $p = 0,846$; V3: $F = 0,269$, $p = 0,605$; Q1: $F = 1,036$, $p = 0,311$; N1: $F = 3,450$, $p = 0,065$), das Vorliegen der Normalverteilung der Daten anhand des Kolmogorov-Smirnov-Tests für den KF-Test ($Z(146) = 0,069$, $p = 0,088$) sowie für den quantitativen Subtest (Q1: $Z(146) = 0,71$, $p = 0,069$) bestätigt. Der Kolmogorov-Smirnov-Test zeigte jedoch auch, dass sowohl für den verbalen Subtest (V3: $Z(146) = 0,096$, $p = 0,002$) als auch für den non-verbalen Subtest (N1: $Z(146) = 0,152$, $p < 0,001$) keine Normalverteilung der Daten vorlag. Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Ziegler (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

al., 2017, S. 26f.). Die beiden Testformen lassen sich außerdem hinsichtlich ihrer Aufgabenschwierigkeit vergleichen, wenn man die ersten zehn Aufgaben betrachtet, die von 99,3 % aller Schüler bearbeitet wurden. Man ermittelt dabei eine mittlere Aufgabenschwierigkeit von $p = 0,88$ für Testform A und $p = 0,89$ für Testform B. Die nur minimale Abweichung von $\Delta p = 0,1 \%$, weist auf eine hohe Parallelität der beiden Testformen hin.

Gruppe	N	LV M (SD,SE)	LGS M (SD,SE)	LGN M (SD,SE)
CLIL	73	50,3 (9,06;1,06)	47,7 (9,28;1,09)	58,9 (11,98;1,40)
nicht-CLIL	73	47,3 (9,04;1,06)	47,9 (8,49;0,99)	53,4 (12,72;1,49)
gesamt	146	48,8 (9,14;0,76)	47,8 (8,86;0,73)	56,14 (8,86;1,04)

Tabelle 25: Deskriptive Statistik zum Lesegeschwindigkeits- und -verständnistest

Schneider et al. (2017) empfehlen zum Vergleich der Testleistungen die Umrechnung der Rohdaten in T-Werte. Diese wurden anhand einer Normierungsstichprobe von 7142 Schülern aus 13 Bundesländern ermittelt und beziehen sich im vorliegenden Fall auf Schüler der Klasse 8 am Gymnasium (vgl. Schneider et al., 2017, S. 15ff.). In den Tests zum Leseverständnis (LV) und zur Lesegenauigkeit (LGN) schnitten die CLIL-Schüler etwas besser ab als die nicht-CLIL-Schüler (LV: $\Delta M = 3,0$; LGN: $\Delta M = 5,5$). Bei der Lesegeschwindigkeit (LGS) hingegen war der Unterschied deutlich kleiner ($\Delta M = 0,2$). Hier schnitten die nicht-CLIL-Schüler etwas besser ab. Die Mittelwertsunterschied im LGN-Test war statistisch signifikant (LGN: $t(144) = 2,712$, $p = 0,007$)⁷¹. Er repräsentiert aber nach Cohen (1988) nur einen kleinen Effekt (LV: Cohen's $d = 0,327$; LGN: Cohen's $d = 0,449$). Die Unterschiede im LGS-Test ($t(144) = -0,093$, $p = 0,926$) und im Leseverständnistest (LV: $t(144) = 1,974$, $p = 0,050$) waren statistisch nicht signifikant. Die C-Tests wurden von insgesamt 145 Schülern bearbeitet⁷². Der Wert für Cronbachs Alpha beträgt für die insgesamt 97 Items $\alpha = 0,922$ und liegt damit im sehr guten Bereich. Die durchschnittliche

71 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (LV: $F = 0,050$, $p = 0,823$; LGN: $F = 0,040$, $p = 0,841$; LGS: $F = 0,112$, $p = 0,738$), das Vorliegen der Normalverteilung der Daten wurde für den LV-Test und den LGS-Test anhand des Kolmogorov-Smirnov-Tests (LV: $Z(146) = 0,072$, $p = 0,062$; LGS: $Z(146) = 0,067$, $p = 0,2$) bestätigt. Die Daten des LGN-Tests waren jedoch laut Kolmogorov-Smirnov-Test ($Z(146) = 0,142$, $p < 0,001$) nicht normalverteilt. Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Ziegler (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

72 Die Testformen A und B unterschieden sich nur in der Reihenfolge der vier Texte. Auf eine Differenzierung nach der Testform kann deshalb bei der Reliabilitätsanalyse verzichtet werden.

Itemschwierigkeit beträgt $p = 0,65$, was einem mittleren Schwierigkeitsgrad entspricht (vgl. Bortz & Döring, 2002, S. 218). Siebenunddreißig Prozent der Items befinden sich mit $p_{\min} = 0,02$ und $p_{\max} = 0,99$ außerhalb dieses Bereiches. Sie werden jedoch im Test belassen, da sie dazu beitragen, zwischen besonders guten bzw. besonders schwachen Probanden zu differenzieren. Einbußen bei der Trennschärfe einzelner Items werden in Kauf genommen, weil der Tests mit 97 Items recht umfangreich ist. Die Trennschärfen liegen überwiegend im mittleren Bereich, wobei die mittlere Trennschärfe mit $r = 0,31$ ebenfalls in diesem Bereich liegt (vgl. Bortz & Döring, 2002, S. 219). Items mit negativen Trennschärfen gibt es nicht.

Gruppe	N	C-Test M (SD; SE)	N	Englischnote MD
CLIL	73	63,8 (13,2; 1,56)	73	9,00
nicht-CLIL	72	61,3 (14,5; 1,71)	67	9,00
Gesamt	145	62,5 (13,9; 1,15)	140	9,00

Tabelle 26: Deskriptive Statistik zum C-Test und zur Englischnote

Vergleicht man die Mittelwerte der Ergebnisse im C-Test beider Gruppen, so fällt auf, dass die CLIL-Gruppe etwas besser abschneidet ($\Delta M = 2,5$) als die nicht-CLIL-Gruppe. Der t-Test bestätigt, dass der Unterschied statistisch nicht signifikant ist ($t(143) = 1,093$, $p = 0,276$)⁷³. Die Schulnoten wurden mit Tendenz erfasst und in der für die Oberstufe üblichen Skala von 0 (ungenügend) bis 15 (sehr gut plus) abgebildet. Die Mediane der Schulnote im Fach Englisch sind identisch. Der Wert 9 entspricht dabei der Note „befriedigend plus“. Der Mann-Whitney-U-Test belegt, dass der Unterschied in den Schulnoten ebenfalls statistisch nicht signifikant ist ($U = 2470,500,5$, $z = -0,048$, $p = 0,962$). Der ACL-Test wurde von insgesamt 138 Schülern bearbeitet. Die Berechnung von Cronbachs Alpha für die beiden Testformen zeigt, dass die innere Konsistenz des ACLT nach der Überarbeitung der Items im Rahmen der Pilotierung nochmals verbessert werden konnte. Die Werte für den gesamten Test liegen jetzt für beide Testformen bei über 0,8 und damit im guten Bereich (Testform A: $\alpha = 0,86$; Testform B: $\alpha = 0,83$). Die durchschnittliche Aufgabenschwierigkeit lag mit 0,53 wieder im angestrebten mittleren Bereich (Testform A: $p = 0,53$; Testform B: $p = 0,54$). Damit konnte der Schwierigkeitsunterschied zwischen beiden Testformen weiter verringert werden.

⁷³ Der Levene-Test zeigt Varianzhomogenität an ($F = 0,327$, $p = 0,568$), der Kolmogorov-Smirnov-Test ($Z(145) = 0,087$, $p = 0,009$) hingegen keine Normalverteilung. Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Ziegler (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

Insgesamt gelang es, mehr Aufgaben von mittlerer Schwierigkeit zu erzeugen und den Anteil der leichten (LH 0,8 – 1,0) von 26 % auf 14 % zu reduzieren. Der Anteil der schweren Aufgaben (LH 0,0 – 0,2) betrug 7 %. Dabei handelte es sich um Aufgaben zum Anforderungsbereich III. Damit waren diese Aufgaben wieder deutlich schwerer als die Aufgaben zu den Anforderungsbereichen I und II, die mit einer mittleren Schwierigkeit von 0,55 bzw. 0,57 ähnlich schwer waren. Die Analyse der Trennschärfen ergab eine mittlere Trennschärfe von etwa 0,3 für beide Tests (Test A: TS = 0,31, Test B: TS = 0,28), was laut Bortz & Döring (2002, S. 219) ein mittelmäßiger Wert ist. Damit haben sich die Trennschärfen im Mittel gegenüber der Pilotierung zwar kaum verändert, betrachtet man jedoch die Items, die in der Pilotierung durch sehr geringe Trennschärfen ($TS < 0.1$) auffielen, so hat sich deren Zahl nach der Überarbeitung von 9 auf 2 verringert. Von den insgesamt 84 Items der Testformen A und B weisen lediglich drei eine negative Trennschärfe auf und werden deshalb für die Auswertung aus dem Test gestrichen. Dies sind die Items 34 (Test A) und die Items 2 und 18 (Test B).

Gruppe	N	AFB 1 M (SD, SE)	AFB 2/3 M (SD, SE)	ACLT M (SD, SE)	N	Chemienote MD
CLIL	69	7,3 (2,9;0,4)	16,3 (6,8;0,8)	23,6 (9,1;1,1)	73	8,00
nicht-CLIL	69	7,2 (2,9;0,4)	16,7 (6,7;0,8)	23,8 (8,7;1,1)	67	8,00
Gesamt	138	7,2 (2,9;0,3)	16,5 (6,7;0,6)	23,7 (8,9;0,8)	140	8,00

Tabelle 27: Deskriptive Statistik zum allgemeinen Chemieleistungstest und zur Chemienote

Die Ergebnisse des ACL-Tests zeigen, dass beide Gruppen über sehr ähnliche Voraussetzungen in Bezug auf den Umgang mit chemischem Fachwissen verfügen. Während die CLIL-Schüler im ersten Teil des Allgemeinen Leistungstest Chemie (AFB I) etwas besser abschneiden als die nicht-CLIL-Schüler, erzielen letztere im zweiten Teil (AFB II/III) ein besseres Ergebnis. Insgesamt schneiden die nicht-CLIL-Schüler deshalb im ACL-Test geringfügig besser ab als die CLIL-Schüler. Die Unterschiede sind jedoch nur sehr klein und in keinem Fall statistisch signifikant, wie t-Tests zeigen (AFBI: $t(136) = 0,290$, $p = 0,772$; AFBII/III: $t(136) = -0,314$, $p = 0,753$; ; ACLT: $t(136) = -0,143$, $p = 0,886$)⁷⁴. Auch die Schulnoten im Fach Chemie deuten darauf hin, dass beide

⁷⁴ Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (AFBI: $F = 0,025$, $p = 0,874$; AFBII/III: $F = 0,034$, $p = 0,855$; ACLT: $F = 0,034$, $p = 0,854$), das Vorliegen der Normalverteilung der Daten für den ACL-Test anhand des Kolmogorov-Smirnov-Tests (ACLT: $Z(138) = 0,074$, $p = 0,06$) bestätigt. Die Daten des AFBI-Teils sowie des AFBII/III-Teils waren jedoch laut Kolmogorov-Smirnov-Test nicht normalverteilt (AFBI: $Z(138) = 0,077$, $p = 0,043$; AFBII/III: $Z(138) = 0,109$, $p < 0,001$). Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung von der Normalverteilung hin. Deshalb wurde der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

Gruppen in Bezug auf ihre Leistungen im Fach Chemie vergleichbar sind. Die Mediane (MD) der Schulnoten im Fach Chemie sind identisch, sie entsprechen der Note „befriedigend“. Der Mann-Whitney-U-Test belegt zudem, dass die Unterschiede in den Schulnoten statistisch nicht signifikant sind ($U=2311,500,5$, $z = -0,891$, $p = 0,373$). Beide Gruppen sind damit im Fach Chemie um eine Notenstufe schlechter als im Fach Englisch.

7.1.4 Interesse und Selbstkonzept

Die Überprüfung der Reliabilität im Rahmen der Hauptuntersuchung ergab sowohl für das Interesse am Fach Chemie wie auch für das Interesse am Fach Englisch gute Kennwerte für Cronbachs Alpha. Sie waren mit $FICH = 0.85$ und $FIEN = 0.88$ besser als in der Pilotuntersuchung. Auch die Werte für das Selbstkonzept lagen mit $SKCH = 0.80$ und $SKEN = 0,67$ im akzeptablen Bereich.

Gruppe	N	Chemie		Englisch	
		Interesse M (SD; SE)	Selbstkonzept M (SD; SE)	Interesse M (SD; SE)	Selbstkonzept M (SD; SE)
CLIL	71	3,17 (0,76;0,09)	3,38 (0,70;0,08)	3,51 (0,60;0,07)	3,57 (0,55;0,07)
nicht-CLIL	70	2,91 (0,71;0,08)	3,11 (0,70;0,08)	3,53 (0,73;0,09)	3,72 (0,45;0,05)
Gesamt	141	3,04 (0,74;0,06)	3,25 (0,71;0,06)	3,52 (0,67;0,06)	3,65 (0,51;0,04)

Tabelle 28: Deskriptive Statistik zu Interesse und Selbstkonzept in Englisch und Chemie

Es zeigt sich, dass in beiden Gruppen ein deutliches Fachinteresse sowie ein insgesamt positives Selbstkonzept in Bezug auf die Fächer Chemie und Englisch vorliegen. Die Mittelwerte liegen zwischen 3 und 4, beziehungsweise knapp unterhalb von 3 ($FICH_{\text{nicht-CLIL}} = 2,91$), was bedeutet, dass die Probanden im Mittel den positiv formulierten Items „eher“, beziehungsweise „völlig“ zustimmen und den negativ formulierten Items „weniger“ bis „gar nicht“.

Erwartungskonform sind die Werte für das Fach Englisch in beiden Gruppen höher als für das Fach Chemie, das bei Schülern häufig als schwierig und unbeliebt gilt (Reiss, Sälzer, Schiepe-Tiska, & Klieme, 2016). Beim Vergleich der Gruppen zeigt sich aber, dass sowohl Interesse als auch Selbstkonzept in Bezug auf das Fach Chemie in der CLIL-Gruppe ausgeprägter sind als in der nicht-CLIL-Gruppe, während für das Fach Englisch die Werte ähnlich sind, beziehungsweise etwas höher für die nicht-CLIL-Gruppe (vgl. Tabelle oben). t-Tests belegen, dass die Unterschiede für das Fach Chemie zudem statistisch signifikant sind ($FICH: t(139) = 2,046$, $p = 0,043$; $SKCH: t(139) = 2,304$, $p = 0,023$). Die Berechnung der Effektstärken zeigt zudem, dass jeweils ein mittlerer bis großer Effekt vorliegt ($FICH: \text{Cohen's } d = 0,74$; $SKCH: \text{Cohen's } d = 0,70$). Die Unterschiede für das Fach Englisch sind statistisch nicht signifikant ($FIEN: t(139) = -0,180$, $p = 0,857$; $SKEN: t(139)$

= -1,726. $p = 0,087$)⁷⁵. Die Voraussetzung der CLIL-Gruppe sind damit in Bezug auf Chemieinteresse und Chemieselbstkonzept besser als in der nicht-CLIL-Gruppe, während Englischinteresse und Englischselbstkonzept vergleichbar sind. Zusammenfassend lässt sich festhalten, dass die Parallelisierung der beiden Gruppen im Hinblick auf demografische Merkmale, Leistungsvermögen, Interesse und Selbstkonzept gelungen ist. In fast allen Bereichen ergeben sich für die Untersuchungs- und Kontrollgruppe vergleichbare Ergebnisse sowohl für die Gruppenmittelwerte als auch für die Standardabweichung. Statistisch signifikante Unterschiede lassen sich nur für die Lesegenauigkeit sowie für das Interesse am Fach Chemie und das Chemieselbstkonzept feststellen. In all diesen Fällen schneiden die CLIL-Schüler besser ab als die nicht-CLIL-Schüler.

7.2 Die Interessensgruppen

Von den 140 Schülern, die den Prefragebogen bearbeitet haben, können lediglich sechs keiner Interessensgruppe zugeordnet werden, weil die entsprechende Angabe auf dem Fragebogen fehlte. Von den verbleibenden 134 Schülern zählt die Gruppe der *beidseitig wenig orientierten Schüler* (bswo-Schüler) drei Mitglieder. Weil zuverlässige Daten in der bswo-Gruppe aufgrund der kleinen Gruppengröße ($N=3$) nicht zu erwarten sind, wird diese Gruppe von den weiteren Untersuchungen ausgeschlossen. Die Gruppe der *fremdsprachlich orientierten Schüler* (fso-Gruppe) und der *naturwissenschaftlich orientierten Schüler* (nwo-Gruppe) sind vergleichbar groß und machen knapp Dreiviertel aller Schüler aus, während fast ein Viertel der Schüler sich als *beidseitig orientiert* bezeichnet (bso-Gruppe). Die folgende Tabelle zeigt wichtige Kontrollvariablen, die einen Einfluss auf die sachfachlichen und sprachlichen Leistungen sowie auf die Motivation im bilingualen Chemieunterricht haben können.

75 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (FICH: $F = 0,384$, $p = 0,537$; FIEN: $F = 1,172$, $p = 0,281$; SKCH: $F = 0,260$, $p = 0,611$; SKEN: $F = 3,238$, $p = 0,074$), das Vorliegen der Normalverteilung der Daten anhand des Kolmogorov-Smirnov-Tests überprüft. Es zeigte sich, dass die Daten in keinem der vier Fälle normalverteilt sind (FICH: $Z(140) = 0,148$, $p < 0,001$; FIEN $Z(140) = 0,249$, $p < 0,001$; SKCH: $Z(140) = 0,146$, $p < 0,001$; SKEN: $Z(140) = 0,270$, $p < 0,001$). Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

Hauptuntersuchung: Ergebnisse und Auswertung

Merkmal	Wert	Interessensgruppe					
		fso	nwo	bso	bswo	kA	Gesamt
		Demographie					
Schülerzahl	N	53	46	32	3	6	140
Alter	M	13,15	13,43	13,30	13,67	13,33	13,29
Anteil Jungen	%	21,20	63,00	45,50	33,30	66,70	42,90
KFT (V3;Q1;N1)	M	42,41 (41,10;42,35;44,08)	41,11 (39,30;42,33;41,80)	43,64 (39,75;45,22;45,94)	34,56 (23,00;27,00;53,67)	40,03 (36,00;41,33;43,00)	42,00 (39,59;42,63;43,93)
LGVT 5-12+ (LVS;LGS;LGN)	M	50,81;49,87;58,04	46,15;44,28;54,38	51,26;49,44;56,41	41,67;41,00;47,33	48,17;48,33;53,67	49,07;47,70;56,05
		Chemie					
ACLT (AFB1;AFB2/3)	M	20,46 (6,61;13,85)	23,17(6,45;16,72)	22,18 (7,21;14,97)	16,67 (4,67;12,00)	23,67 (6,83;16,83)	21,80 (6,66;15,14)
FICH	M	2,64	3,40	3,28	2,50	2,86	3,04
SKCH	M	2,90	3,55	3,43	1,92	3,56	3,24
		Englisch					
C-Test	M	66,83	55,43	62,39	48,00	55,67	61,20
FIEN	M	3,76	3,12	3,67	3,83	3,46	3,52
SKEN	M	3,82	3,35	3,81	3,50	3,54	3,64

Tabelle 29: Ausprägung verschiedener Kontrollvariablen in den Interessensgruppen

Man erkennt, dass die fso- und die nwo-Gruppe die größten Interessensgruppen darstellen. Etwa 80 Prozent der Schüler bezeichnen sich entweder als fremdsprachlich- oder als naturwissenschaftlich orientiert. Die Gruppe der fso-Schüler ist geringfügig größer als die nwo-Gruppe, wobei sie sich im Altersdurchschnitt nur wenig unterscheiden. Die einfaktorielle ANOVA bestätigt, dass hier kein statistisch signifikanter Unterschied vorliegt ($F(4;137) = 0,463$; $p = 0,763$)⁷⁶. Der Jungenanteil ist in der nwo-Gruppe mit 63 % dreimal so hoch wie in der fso-Gruppe. Dies entspricht der Annahme, dass Jungen sich mehr für Naturwissenschaften interessieren und Mädchen mehr für Sprachen (vgl. Abschnitt 4.3.2.). Der Anteil von Jungen in der bso-Gruppe entspricht hingegen mit 45,5 % in etwa dem Jungenanteil der Gesamtgruppe, der bei 42,2 % liegt. Der Chi-Quadrat-Test bestätigt, dass ein statistisch signifikanter Unterschied in der Verteilung des Geschlechts auf die Interessensgruppen vorliegt ($\chi^2 = 19,246$; $p < 0,001$). Die im Anschluss durchgeführten Post-hoc-Tests zeigen, dass die Verteilung von Jungen und Mädchen in der nwo- und in der fso-Gruppe statistisch signifikant von

⁷⁶ Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests ($F = 0,312$; $p = 0,870$) nachgewiesen. Die Daten waren laut Kolmogorov-Smirnov-Test nicht normalverteilt ($Z(142) = 0,319$, $p < 0,001$). Bühner & Ziegler (2009, S. 372) empfehlen dennoch, nicht auf ein nonparametrisches Testverfahren auszuweichen, da „die Varianzanalyse stabil gegenüber Verletzungen der Normalverteilung“ ist. Der zusätzlich durchgeführte Kruskal-Wallis-Test liefert ein ähnliches Ergebnis ($H(3) = 2,166$; $p = 0,539$).

der erwarteten Verteilung abweichen ($p_{\text{nwo-Jungen}} = 0,0007$; $p_{\text{nwo-Mädchen}} = 0,0007$; $p_{\text{fso-Mädchen}} = 0,00007$; $p_{\text{fso-Jungen}} = 0,00007$)⁷⁷. Verriere (2014, S. 242) und Lipski-Buchholz (2017, S. 112) kommen für das Fach Mathematik zu ähnlichen Ergebnissen und folgern, dass Mathematik eher eine Domäne der Jungen und Englisch eher eine Domäne der Mädchen ist. Die Unterschiede in den kognitiven Fähigkeiten der Schüler sind hingegen gering und statistisch nicht signifikant, wie eine einfaktorielle ANOVA belegt (KFT: $F(4;134) = 1,031$; $p = 0,394$; V3: $F(4;132) = 1,611$, $p = 0,175$; Q1: $F(4;133) = 1,648$; $p = 0,166$; N1: $F(4;134) = 1,242$; $p = 0,394$)⁷⁸. Im LGV-Test schneiden die nwo-Schüler erkennbar schlechter ab als die fso- und die bso-Schüler, die vergleichbare Werte erzielen. Die einfaktorielle ANOVA zeigt, dass sowohl für das Leseverständnis, als auch für die Lesegeschwindigkeit ein statistisch signifikanter Unterschied vorliegt (LVS: $F(4;135) = 2,899$; $p = 0,024$; LGS: $F(4;136) = 3,614$; $p = 0,008$)⁷⁹. Aufgrund der Post-hoc-Tests nach Bonferroni zeigt sich, dass lediglich der Unterschied zwischen der fso-Gruppe und der nwo-Gruppe bei der Lesegeschwindigkeit statistisch signifikant ist (LGS: $p = 0,012$). Dabei sind die Effektgrößen in beiden Fällen nach Cohen (1988, S. 284) minimal bis klein (LGS: $f = 0,096$). Die Ergebnisse aus den C-Tests zeigen, dass die fso-Schüler im Mittel leistungsstärker im Fach Englisch sind als die nwo-Schüler ($\Delta M = 11,4$); die bso-Schüler nehmen eine Zwischenposition ein. Es zeigte sich, dass ein statistisch signifikanter Unterschied zwischen den Gruppen vorliegt ($H(3) = 19,906$; $p < 0,001$)⁸⁰. Die paarweisen Vergleiche lassen erkennen, dass nur der Unterschied zwischen der fso- und der nwo-Gruppe ($Z = -4,466$; $p < 0,001$) statistisch signifikant ist. Die Effektstärke ist mit $r =$

77 Um die Irrtumswahrscheinlichkeit für falsche Signifikanzen zu verringern, wurde die Bonferroni-Korrektur durchgeführt (vgl. Bühner & Ziegler, 2009, S. 551).

78 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (KFT: $F = 0,825$; $p = 0,511$; V3: $F = 1,145$, $p = 0,338$; Q1: $F = 2,204$; $p = 0,072$; N1: $F = 0,144$; $p = 0,965$) nachgewiesen. Die Daten waren laut Kolmogorov-Smirnov-Test für KFT ($Z(139) = 0,053$, $p = 0,200$) sowie für KFTN1 ($Z(139) = 0,073$; $p = 0,069$) normalverteilt, nicht aber für KFTV3 ($Z(139) = 0,109$; $p < 0,001$) und KFTQ1 ($Z(138) = 0,100$; $p = 0,002$). Der zusätzlich durchgeführte Kruskal-Wallis-Test lieferte für KFTV3 ($H(3) = 4,855$; $p = 0,183$) und KFTQ1 ($H(3) = 3,348$; $p = 0,341$) ein ähnliches Ergebnis.

79 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (LVS: $F = 0,196$; $p = 0,940$; LGS: $F = 0,188$, $p = 0,945$; LGN: $F = 0,781$; $p = 0,540$) nachgewiesen. Die Daten waren laut Kolmogorov-Smirnov-Test für LVS ($Z(140) = 0,072$, $p = 0,072$) und für LGS ($Z(141) = 0,065$; $p = 0,200$) normalverteilt, für LGN ($Z(142) = 0,116$; $p < 0,001$) jedoch nicht. Der zusätzlich durchgeführte Kruskal-Wallis-Test lieferte für LGN ($H(3) = 5,089$; $p = 0,165$) allerdings ein ähnliches Ergebnis.

80 Das Vorliegen von Varianzhomogenität konnte mit Hilfe des Levene-Tests ($F = 3,495$; $p = 0,009$) nicht nachgewiesen werden. Die Daten waren laut Kolmogorov-Smirnov-Test ($Z(143) = 0,112$; $p < 0,001$) auch nicht normalverteilt. Deshalb wurde der Kruskal-Wallis-Test durchgeführt.

0,4 mittel bis stark⁸¹. Für das Fach Chemie ergeben sich aus den Ergebnissen zum ACL-Test geringfügige Vorteile für die nwo- gegenüber der fso-Gruppe, wobei die Werte der bso-Gruppe wieder dazwischen liegen. Die Unterschiede sind jedoch statistisch nicht signifikant, wie eine einfaktorielle ANOVA ergibt (ACLT: $F(4;138) = 0,657$; $p = 0,623$; ACLT1: $F(4;138) = 0,532$; $p = 0,712$; ACLT2/3: $F(4;138) = 1,080$; $p = 0,369$)⁸². Der Kruskal-Wallis-Test zeigt, dass sowohl für das Interesse als auch für das Selbstkonzept in beiden Fächern signifikante Unterschiede zwischen den Gruppen bestehen (FICH: $H(3) = 32,586$; $p < 0,001$; FIEN: $H(3) = 29,386$; $p < 0,001$; SKCH: $H(3) = 27,837$; $p < 0,001$; SKEN: $H(3) = 20,327$; $p < 0,001$)⁸³. Dabei sind erwartungskonform die Werte für das Interesse und das Selbstkonzept bei den nwo-Schülern im Fach Chemie höher als bei den fso-Schülern. Genau umgekehrt verhält es sich im Fach Englisch. Hier sind die Werte für Interesse und Selbstkonzept der fso-Schüler höher als die Werte der nwo-Schüler. Die paarweisen Vergleiche belegen, dass alle Unterschiede statistisch hoch signifikant sind (FICH fso-nwo: $Z = 5,271$; $p < 0,001$; SKCH fso-nwo: $Z = 4,591$; $p < 0,001$; FIEN nwo-fso: $Z = -5,088$; $p < 0,001$; SKEN nwo-fso: $Z = -3,993$; $p < 0,001$)⁸⁴. Die Effektstärken liegen mit $0,4 < r < 0,5$ im mittleren bis starken Bereich. Dies kann auch als Hinweis auf die Validität des hier gewählten Verfahrens zur Einteilung der Schüler in Interessensgruppen verstanden werden. Für die Gruppe der bso-Schüler finden sich, ebenfalls erwartungskonform, hohe Werte für Interesse und Selbstkonzept in beiden Fächern. Die Werte sind für das Fach Chemie signifikant höher als die Werte der fso-Schüler und

81 Die zusätzlich durchgeführte einfaktorielle ANOVA bestätigt, dass statistisch signifikante Unterschiede zwischen den Gruppen bestehen ($F(4;138) = 4,113$; $p = 0,004$), wobei auch der Post-hoc-Test nach Bonferroni lediglich den Mittelwertsunterschied zwischen der fso- und der nwo-Gruppe als signifikant ausweist ($p = 0,004$).

82 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (ACLT1: $F = 0,920$; $p = 0,454$; ACLT2/3: $F = 0,827$; $p = 0,510$; ACLT: $F = 0,963$; $p = 0,430$) nachgewiesen. Die Daten waren laut Kolmogorov-Smirnov-Test für ACLT1 ($Z(143) = 0,071$, $p = 0,077$) sowie für ACLT2/3 ($Z(143) = 0,066$; $p = 0,200$) und ACLT ($Z(143) = 0,053$; $p = 0,200$) normalverteilt.

83 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests für FICH ($F = 0,583$; $p = 0,676$), nicht aber für FIEN ($F = 5,537$; $p < 0,001$), SKCH ($F = 3,136$; $p = 0,017$) und SKEN ($F = 9,635$; $p < 0,001$) nachgewiesen. Die Daten waren laut Kolmogorov-Smirnov-Test auch weder für FICH ($Z(140) = 0,110$, $p < 0,001$) noch für FIEN ($Z(140) = 0,212$; $p < 0,001$), SKCH ($Z(140) = 0,124$; $p < 0,001$) oder SKEN ($Z(140) = 0,220$; $p < 0,001$) normalverteilt. Deshalb wurde ein nonparametrisches Testverfahren gewählt.

84 Die zusätzlich durchgeführte einfaktorielle ANOVA liefert ähnliche Ergebnisse wie der Kruskal-Wallis-Test und erweist sich als robust gegen die Verletzung der Voraussetzungen (FICH: $F(4;135) = 9,519$; $p < 0,001$; FIEN: $F(4;135) = 7,675$; $p < 0,001$; SKCH: $F(4;135) = 11,390$, $p < 0,001$; SKEN: $F(4;135) = 7,555$; $p < 0,001$). Auch die Post-hoc-Tests nach Bonferroni zeigen, dass die oben beschriebenen Unterschiede statistisch höchst signifikant sind ($p < 0,001$).

für das Fach Englisch signifikant höher als die Werte der nwo-Schüler, wie paarweise Vergleiche zeigen (FICH fso-bso: $Z = -3,905$; $p = 0,001$; SKCH fso-bso: $Z = -3,055$; $p = 0,002$; FIEN nwo-bso: $Z = -3,953$; $p < 0,001$; SKEN nwo-bso: $Z = -3,327$; $p < 0,001$). Die Effektstärken liegen mit $0,2 < r < 0,5$ im mittleren bis starken Bereich. Zusammenfassend kann festgehalten werden, dass fso-Schüler signifikant höhere Werte in Englischinteresse und Englischselbstkonzept aufweisen als nwo-Schüler. Letztere zeigen jedoch signifikant höhere Werte in naturwissenschaftlichem Interesse und naturwissenschaftlichem Selbstkonzept auf als die fso-Schüler. Bei den fso-Schülern kommt zudem ein signifikant größeres Leistungsvermögen in der Fremdsprache, im Leseverständnis und der Lesegeschwindigkeit dazu. Außerdem sind die Anteile der Mädchen in der fso-Gruppe und die Anteile der Jungen in der nwo-Gruppe signifikant erhöht.

7.3 Der sachfachliche Lernerfolg

Zur Überprüfung der Reliabilität des Tests wurden Cronbachs Alpha sowie die Itemschwierigkeiten und Trennschärfen für die Items beider Testformen berechnet. Dabei wurden die Daten der Posttests zugrunde gelegt. Der Posttest wurde im Rahmen der Hauptuntersuchung von insgesamt 144 Schülern bearbeitet. Die Werte für Cronbachs Alpha liegen im guten Bereich (Testform A: $\alpha = 0,84$; Testform B: $\alpha = 0,89$). Die durchschnittliche Aufgabenschwierigkeit lag mit 0,61 im angestrebten mittleren Bereich (Testform A: $p = 0,64$; Testform B: 0,59). Damit konnte das Schwierigkeitsniveau im Vergleich zum Pilottest, wie geplant, etwas angehoben werden ($p_{\text{Pilot}} = 0,73$). Der Anteil der leichten Aufgaben wurde von 38 % auf 17 % reduziert und der Anteil der schweren Aufgaben von 0 % auf 2 % erhöht. Die Aufgaben zum Anforderungsbereich III ($p = 0,35$) waren dabei wieder deutlich schwerer als die Aufgaben zu den Anforderungsbereichen I und II, die mit einer mittleren Schwierigkeit von 0,69 bzw. 0,61 ähnlich schwer waren. Die Analyse der Trennschärfen ergab eine mittlere Trennschärfe von etwa 0,3 für beide Tests (Test A: $TS = 0,33$, Test B: $TS = 0,36$), was laut Bortz (2002, S. 219) ein mittelmäßiger Wert ist. Damit haben sich die Trennschärfen im Mittel gegenüber der Pilotierung zwar kaum verändert, betrachtet man jedoch die Items, die in der Pilotierung durch sehr geringe Trennschärfen ($TS < 0,1$) auffielen, so hat sich deren Zahl nach der Überarbeitung von 28 auf 7, also um 75 %, verringert. Von den insgesamt 84 Items der Testformen A und B weisen jedoch drei eine negative Trennschärfe auf und werden deshalb aus dem Test gestrichen. Dies sind die Items 3 (Test A) und 10 (Test A) sowie das Item 8 (Test B).

7.3.1 Vergleich der Testergebnisse in Untersuchungs- und Kontrollgruppe

Die folgende Tabelle zeigt die deskriptive Statistik zu den CLT-Ergebnissen in beiden Gruppen. Dabei wurden nur solche Probanden berücksichtigt, die zu allen drei Zeitpunkten am Test teilgenommen hatten. Fehlten Testergebnisse, so wurden die betreffenden Datensätze von der Berechnung ausgeschlossen (balanciertes Design).

Zeit	Gruppe	N	CLT M (SD; SE)
Pre	CLIL	67	11,49 (4,6;0,6)
	nicht-CLIL	57	10,01 (4,6;0,6)
	gesamt	124	10,85 (4,7)
Post	CLIL	67	29,78 (11,0;1,3)
	nicht-CLIL	57	28,12 (10,1;1,4)
	gesamt	124	29,02 (10,6)
Follow-Up	CLIL	67	25,66 (10,1;1,2)
	nicht-CLIL	57	24,23 (9,9;1,3)
	gesamt	124	25,00 (10,0)

Tabelle 30: Deskriptive Statistiken zu den Chemieleistungstests

Die Analyse des Profilplots bestätigt für beide Gruppen einen sehr deutlichen Anstieg der mittleren Testergebnisse vom Pre- zum Posttest (Zugewinn 1) sowie ein ebenfalls deutliches Absinken der Werte vom Post- zum Follow-Up-Test (Vergessen). Die Testergebnisse der nicht-CLIL-Gruppe liegen zu allen drei Zeitpunkten unter denen der CLIL-Gruppe. Der Unterschied bleibt dabei vom Pre- bis zum Follow-Up-Test nahezu konstant und beträgt im Mittel 1,5 Punkte.

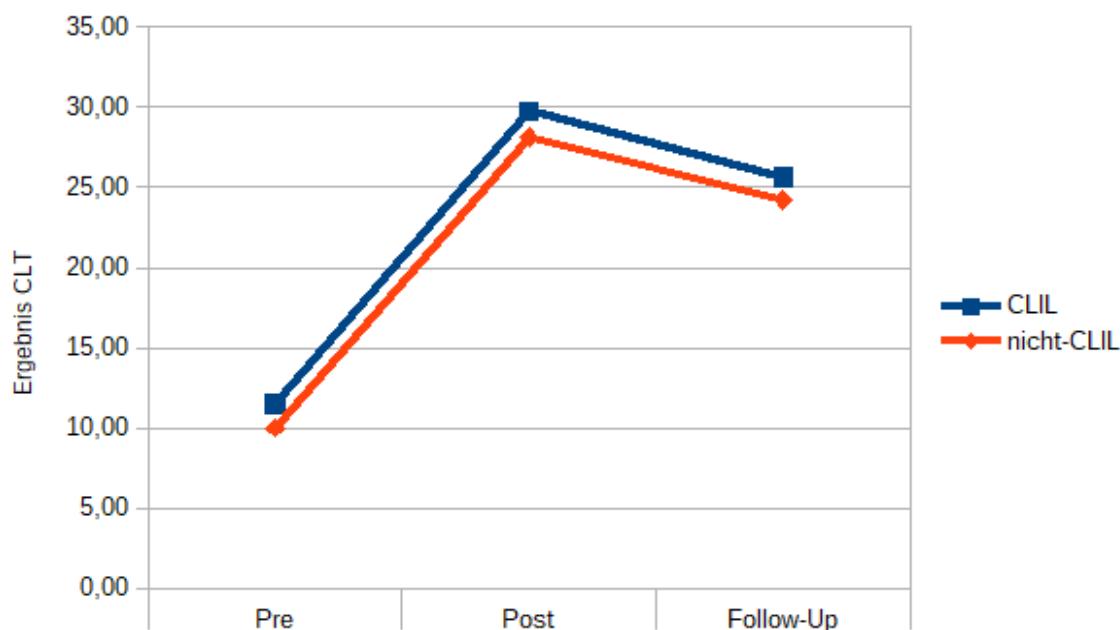


Abbildung 39: Ergebnisse im Chemieleistungstest

Um zu prüfen, inwieweit die Unterschiede statistisch signifikant sind, wurde eine zweifaktorielle ANOVA mit Messwiederholung (gemischtes Design) berechnet. Dabei wurde zwischen den Faktoren Zeit (Faktorstufen: Pretest, Posttest und Follow-Up-Test) und Gruppe (Faktorstufen: CLIL und nicht-CLIL) unterschieden. Die Bezeichnung "gemischtes Design" macht deutlich, dass im Rahmen des Modells sowohl "Zwischensubjekt Effekte" (Faktor: Gruppe) als auch "Innersubjekt Effekte" (Faktor: Zeit) betrachtet werden. Während die Bedingungen der Varianzhomogenität nach dem Levene-Test ($p_{CLTPR} = 0,57$, $p_{CLTPO} = 0,34$, $p_{CLTFU} = 1,0$) und der Sphärizität nach dem Mauchly-Test erfüllt sind ($p = 0,703$), erwiesen sich die Daten als nicht normalverteilt, wodurch eine Voraussetzung des gewählten Berechnungsverfahrens nicht erfüllt ist. Rost (2007, S. 191) weist jedoch darauf hin, dass Varianzanalysen als robuste Verfahren selbst bei "massiver Abweichung von der Normalverteilung nur unwesentlich beeinflusst [werden]". Auch nach Bühner & Ziegler (2009, S. 518) ist das Verfahren relativ robust gegenüber einer Verletzung der Normalverteilungsannahme, wenn die Stichproben, wie bei der vorliegenden Untersuchung der Fall, etwa gleich groß sind. Die Auswertung mittels ANOVA ergab, dass für den Faktor Zeit ein signifikanter Unterschied der Pre-, Post- und Follow-Up-Testergebnisse vorliegt ($F(2,244) = 264,6$; $p < 0,01$), für den Faktor Gruppe jedoch nicht ($F(1,122) = 1,386$; $p = 0,241$). Auch die Ergebnisse für die Interaktion von Zeit und Gruppe sind nicht signifikant ($F(2,244) = 0,14$; $p = 0,986$). Für den Faktor Zeit belegen die paarweisen Vergleiche, dass die Unterschiede zwischen den Pre- und den Posttestmittelwerten, zwischen den Post- und den Follow-Up-Testmittelwerten sowie zwischen den Pre- und den Follow-Up-Testwerten in beiden Gruppen statistisch hoch signifikant sind ($p < 0,001$). Damit scheint das Modul in beiden Gruppen zu einem vergleichbar hohen unmittelbaren Lernerfolg geführt zu haben von im Mittel 18,2 Punkten (Zugewinn 1). Auch die Leistungseinbußen nach sechs Wochen sind in beiden Gruppen nahezu identisch und betragen im Mittel 4,0 Punkte (Vergessen). Ermittelt man den Wert für den mittelfristigen Lernerfolg als Differenz zwischen den Follow-Up-Test- und den Pretestwerten (Zugewinn 2), so zeigt sich, dass auch hier nahezu identische Werte von im Mittel 14,2 Punkten vorliegen. Anschließend wurden die Zugewinne und das Vergessen einer weiteren Untersuchung unterzogen. Dazu wurde für jeden einzelnen Probanden Werte für den individuellen Zugewinn und das Vergessen, als Differenzwerte zwischen Post- und Pretest (Zugewinn 1) bzw. zwischen Follow-Up-Test und Posttest (Vergessen) sowie zwischen Follow-Up-Test und Pretest (Zugewinn 2) berechnet. Diese wurden dann zusätzlich mittels t-Test auf Signifikanz der Mittelwertsunterschiede untersucht⁸⁵. Es zeigt sich, dass weder die Unterschiede im unmittelbaren

85 Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (Zugewinn 1: $F = 0,133$; $p = 0,716$; Vergessen: $F = 5,996$; $p = 0,160$; Zugewinn 2: $F = 0,863$, $p = 0,355$) und das Vorliegen der Normalverteilung

Zugewinn (Zugewinn 1: $T = 0,148$, $p = 0,883$) noch die Unterschiede im Vergessen (Vergessen: $T = -,132$; $p = 0,895$) sowie die Unterschiede im mittelfristigen Zugewinn (Zugewinn 2: $T = 0,015$; $p = 0,988$) statistisch signifikant sind.

Zeit	Gruppe	N	M (SD; SE)
Zugewinn 1 CLT	CLIL	67	18,28 (9,1;1,1)
	nicht-CLIL	57	18,04 (9,6;1,3)
Vergessen CLT	CLIL	67	-4,12 (10,9;1,3)
	nicht-CLIL	57	-3,89 (7,3;01,0)
Zugewinn 2 CLT	CLIL	67	14,16 (9,2;1,1)
	nicht-CLIL	57	14,14 (8,4;1,1)

Tabelle 31: Deskriptive Statistik zu Zugewinn und Vergessen im Chemieleistungstest

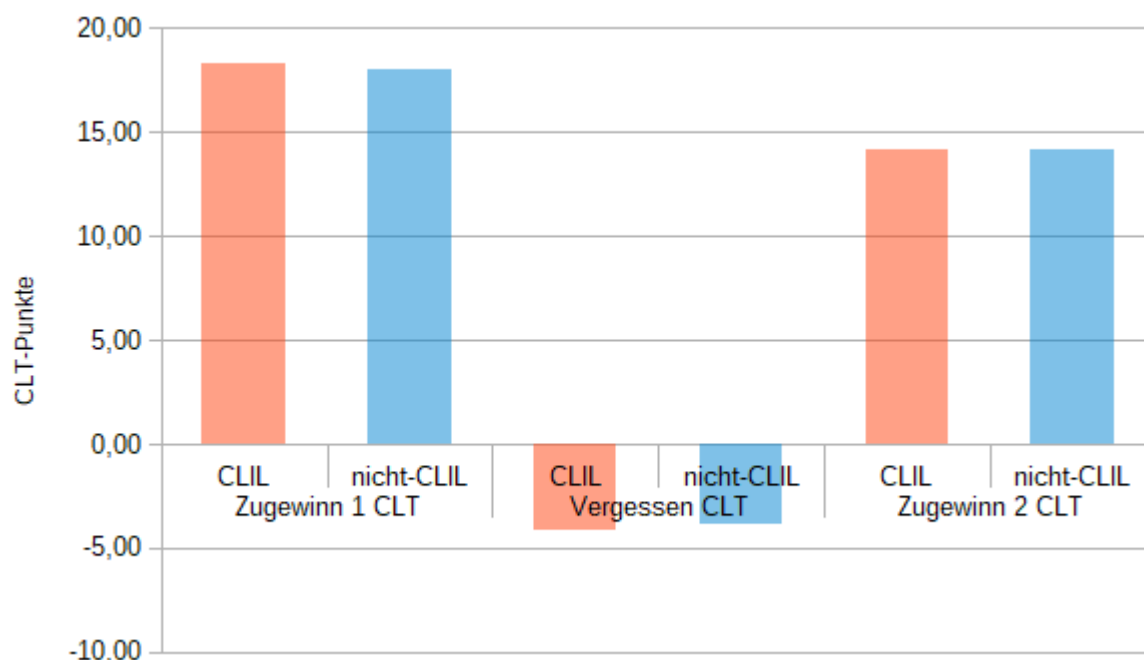


Abbildung 40: Zugewinn und Vergessen im Chemieleistungstest

Wie in Abschnitt 7.1 gesehen, unterscheiden sich Untersuchungs- und Kontrollgruppe statistisch signifikant im chemischen Fachinteresse, im chemischen Selbstkonzept und in der Lesegenauigkeit. In all diesen drei Variablen erzielten die CLIL-Schüler höhere Werte. Im Folgenden wird deshalb

anhand des Kolmogorov-Smirnov-Tests (Zugewinn 1: $Z(124) = 0,076$, $p = 0,076$; Zugewinn 2: $Z(124) = 0,065$, $p = 0,200$) bestätigt. Die Daten zum Vergessen waren laut Kolmogorov-Smirnov-Test nicht normalverteilt (Vergessen: $Z(124) = 0,087$, $p = 0,021$). Der Empfehlung Bühners & Ziegler (2009, S. 260f.) folgend, wurde ein t-Test durchgeführt.

überprüft, ob diese Variablen mit dem unmittelbaren Zugewinn korrelieren und die guten Ergebnisse der CLIL-Gruppe im CL-Test darauf zurückzuführen sind.

	ZU1CLT- CLIL	ZU1CLT- nicht-CLIL
FICH	-0,082	0,056
SKCH	0,055	-0,033
LGN	0,181*	0,029
* $p < 0,05$		

Tabelle 32: Korrelationen zum Lernerfolg⁸⁶

Es findet sich nur eine statistisch signifikante Korrelation, und zwar zwischen der Lesegenauigkeit und dem unmittelbaren Zugewinn der CLIL-Schüler. Chemisches Fachinteresse und Selbstkonzept scheinen keinen Einfluss auf den Lernerfolg im Modul zu haben. Das Ausmaß des Einflusses der Lesegenauigkeit wird mit Hilfe der multiplen linearen Regression weiter unten bestimmt. In einem Modell, das weitere Prädiktoren berücksichtigt, zeigt sich, dass die Lesegenauigkeit ebenso wie das chemische Fachinteresse und das chemische Selbstkonzept keinen statistisch signifikanten Einfluss auf den unmittelbaren Zugewinn durch das Modul haben (vgl. Seite 195).

7.3.2 Der Einfluss verschiedener Variablen auf den sachfachlichen Zugewinn

Im Folgenden wird untersucht, welche Faktoren für den sachfachlichen Lernerfolg der CLIL-Schüler verantwortlich sind. Dabei werden sowohl leistungsbezogene Variablen wie die allgemeine Leistungsfähigkeit in Chemie (ACLT) bzw. in Englisch (CT) als auch motivational-affektive Faktoren, wie fachbezogenes Interesse (FIEN, FICH) und Selbstkonzept (SKEN, SKCH) betrachtet. Es geht also um die Frage, inwieweit ein großes Interesse, ein positives Selbstkonzept aber auch ein hohes Leistungsvermögen den sachfachlichen Lernerfolg im Modul beeinflussen. Auch kognitive Faktoren (KFT) werden dabei berücksichtigt. Die entsprechenden Korrelationsberechnungen ergaben statistisch hoch signifikante Korrelationen zwischen den leistungsbezogenen Variablen ACLT und CT und dem Zugewinn, sowie eine statistisch signifikante Korrelation zwischen der kognitiven Leistungsfähigkeit (KFT) und dem Lernerfolg. Offensichtlich führen eine hohe Leistungsfähigkeit in Chemie und in Englisch, sowie ausgeprägte kognitive Fähigkeiten zu einem größeren sachfachlichen Lernerfolg im bilingualen Modul. Während das Fachinteresse in keinem der beiden Fächer mit dem Zugewinn korreliert, wirkt sich ein positives Selbstkonzept im Fach

⁸⁶ Es wurde der Korrelationskoeffizient Kendall-Tau berechnet, da in der nicht-CLIL-Gruppe kein linearer Zusammenhang zwischen dem Zugewinn und den drei Variablen zu erkennen war.

Englisch ebenfalls statistisch hochsignifikant auf den Lernerfolg (ZU1CLT-CLIL) aus. Nach Cohen (1988) entsprechen die Korrelationen ab $r = 0,30$ einem mittleren Effekt (Bühner & Ziegler, 2009, S. 607).

		ZU1CLT- CLIL
Chemie	FICH	-0,117
	SKCH	0,026
	ACLT	0,428**
Englisch	FIEN	-0,012
	SKEN	0,404**
	CT	0,393**
Kognitiv	KFT	0,316*
* $p < 0,05$; ** $p < 0,01$		

Tabelle 33: Korrelationen zum unmittelbaren Zugewinn in der CLIL-Gruppe⁸⁷

Untersucht man die Frage, wie die Zugewinne mit der Leistungsstärke im Fach Chemie zusammenhängen, so stellt man mit Blick auf die entsprechenden Streudiagramme für die CLIL und die nicht-CLIL-Schüler fest, dass die Werte in der nicht-CLIL-Gruppe stärker streuen als in der CLIL-Gruppe. Im unteren Leistungsbereich (0 – 19 Punkte im ACLT) finden sich, anders als in der CLIL-Gruppe, auch größere Zugewinne von 30 und mehr Punkten. Auf der anderen Seite sind bei den leistungsstarken Schülern (ACLT zwischen 30 und 57 Punkten) in der nicht-CLIL-Gruppe auch Schüler mit sehr geringen Zugewinnen von weniger als 10 Punkten zu beobachten.⁸⁸

87 Für die intervallskalierten Variablen wurde der Pearson-Korrelationskoeffizient berechnet, da keine Ausreißer vorlagen und die Inspektion der Streudiagramme jeweils einen linearen Zusammenhang erkennen ließ. Die Werte von Schiefe und Kurtosis waren in allen Fällen nahe bei null, so dass zumindest von einer annähernd univariaten Normalverteilung ausgegangen werden kann. Die bivariate Normalverteilung konnte statistisch nicht überprüft werden. Wegen der Gruppengrößen ($N > 30$) wird jedoch von einer bivariaten Normalverteilung ausgegangen (zentraler Grenzwertsatz). Die für FICH und SKCH ermittelten Koeffizienten unterscheiden sich von den oben berechneten Werten, da sie hier nach Pearson und nicht, wie weiter oben geschehen, nach Kendall-Tau berechnet wurden.

88 Die Punktintervalle der Leistungsbereiche wurden so gewählt, dass sich in jeder Leistungsgruppe ein Drittel der CLIL-Schüler befinden (ACLT-Terzile).

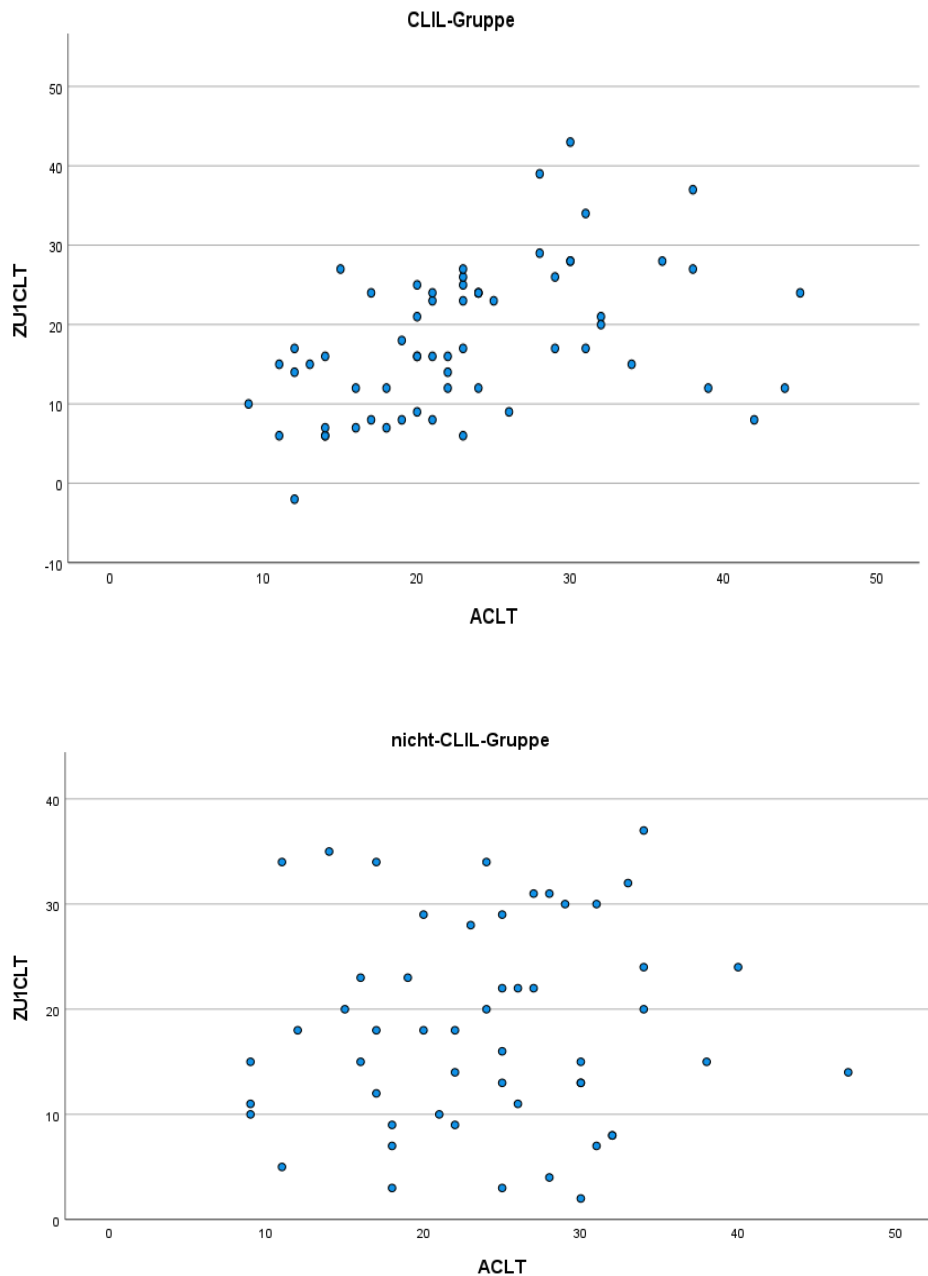


Abbildung 41: Abhängigkeit des Zugewinns von der Leistungsfähigkeit in Chemie

Berechnet man die mittleren Zugewinne für diese drei Leistungsgruppen so zeigt sich, dass diese in der nicht-CLIL-Gruppe relativ gleich sind und zwischen 17,0 und 20,0 Punkten betragen. In der CLIL-Gruppe sind die Zugewinne in der unteren Leistungsgruppe jedoch deutlich geringer und mit 11,7 Punkten nur etwa halb so groß wie in der oberen Leistungsgruppe.

	ACLT 0–19 M(ZU1CLT) (N;M(ACLT))	ACLT 20–24 M(ZU1CLT) (N;M(ACLT))	ACLT 25–57 M(ZU1CLT) (N;M(ACLT))
nicht-CLIL	17,2 (17;14,5)	20,0 (9;22,0)	18,4 (27;30,4)
CLIL	11,7 (20;14,6)	18,6 (22;22,0)	23,7 (21;33,2)

Tabelle 34: Zugewinn im CL-Test in den drei ACLT-Leistungsgruppen

Die Zugewinne in der unteren Leistungsgruppe sind in der nicht-CLIL-Gruppe um gut 30 % größer als in der CLIL-Gruppe. Offensichtlich profitieren leistungsschwächere Schüler eher vom Regelunterricht. Während die Unterschiede in der mittleren Leistungsgruppe gering sind, schneiden die leistungsstarken Schüler im CLIL-Unterricht um gut 20 % besser ab als im Regelunterricht. Es scheint, als ob die leistungsstarken Schüler vom CLIL-Unterricht profitieren⁸⁹. Inwieweit diese Unterschiede statistisch signifikant sind, soll mit einer zweifaktoriellen ANOVA mit den Faktoren Gruppe (Faktorstufen: CLIL und nicht-CLIL) und Leistungsgruppe-ACLT (Faktorstufen: ACLT0-19, ACLT20-24, ACLT25-57)) überprüft werden⁹⁰. Die Analyse ergibt für den Faktor Gruppe keinen signifikanten Unterschied ($F(1;110) = 0,105$; $p = 0,746$), wohl aber für den Faktor Leistungsgruppe-ACLT ($F(2;110) = 6,031$; $p = 0,003$) sowie den Interaktionseffekt ($F(2,110) = 4,065$; $p = 0,020$). Der anschließend durchgeführte Post-hoc-Test nach Bonferroni zeigt, dass ein statistisch signifikanter Unterschied dabei nur zwischen der ersten (ACLT0-19) und der dritten Leistungsgruppe (ACLT25-57) vorliegt ($p = 0,003$). Damit zeigt sich aufgrund der ermittelten Daten, dass die leistungsschwachen Schüler im Regelunterricht einen statistisch signifikant größeren Lernerfolg haben als im CLIL-Unterricht, während die leistungsstarken Schüler im CLIL-Unterricht statistisch signifikant besser abschneiden. Untersucht man die Frage, wie die Zugewinne mit der Leistungsstärke im Fach Englisch zusammenhängen, so stellt man mit Blick auf die entsprechenden Streudiagramme für die CLIL und die nicht-CLIL-Schüler fest, dass die Werte in beiden Gruppen stark streuen. Im unteren Leistungsbereich (0 – 60 Punkte im CT) finden sich

⁸⁹ Da die CLIL-Gruppe sowohl in der oberen als auch in der unteren Leistungsgruppe etwas leistungstärker ist (ACLT = 33,2 bzw. ACLT = 14,6) als die entsprechende nicht-CLIL-Gruppe (ACLT = 30,4 bzw. ACLT = 14,5), wurde, nach Prüfung der Voraussetzungen, ein t-Test durchgeführt. Es zeigte sich, dass die Unterschiede weder in der unteren noch in der oberen Leistungsgruppe statistisch signifikant waren (ACLT25-57: $T = 1,728$; $p = 0,091$; ACLT0-19: $T = 0,075$; $p = 0,941$).

⁹⁰ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 0,265$; $p = 0,931$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten normalverteilt waren ($Z(124) = 0,076$; $p = 0,076$).

jedoch erkennbar größere Zugewinne bei den nicht-CLIL-Schülern, während im oberen Leistungsbereich die CLIL-Schüler bessere Ergebnisse erzielen⁹¹.

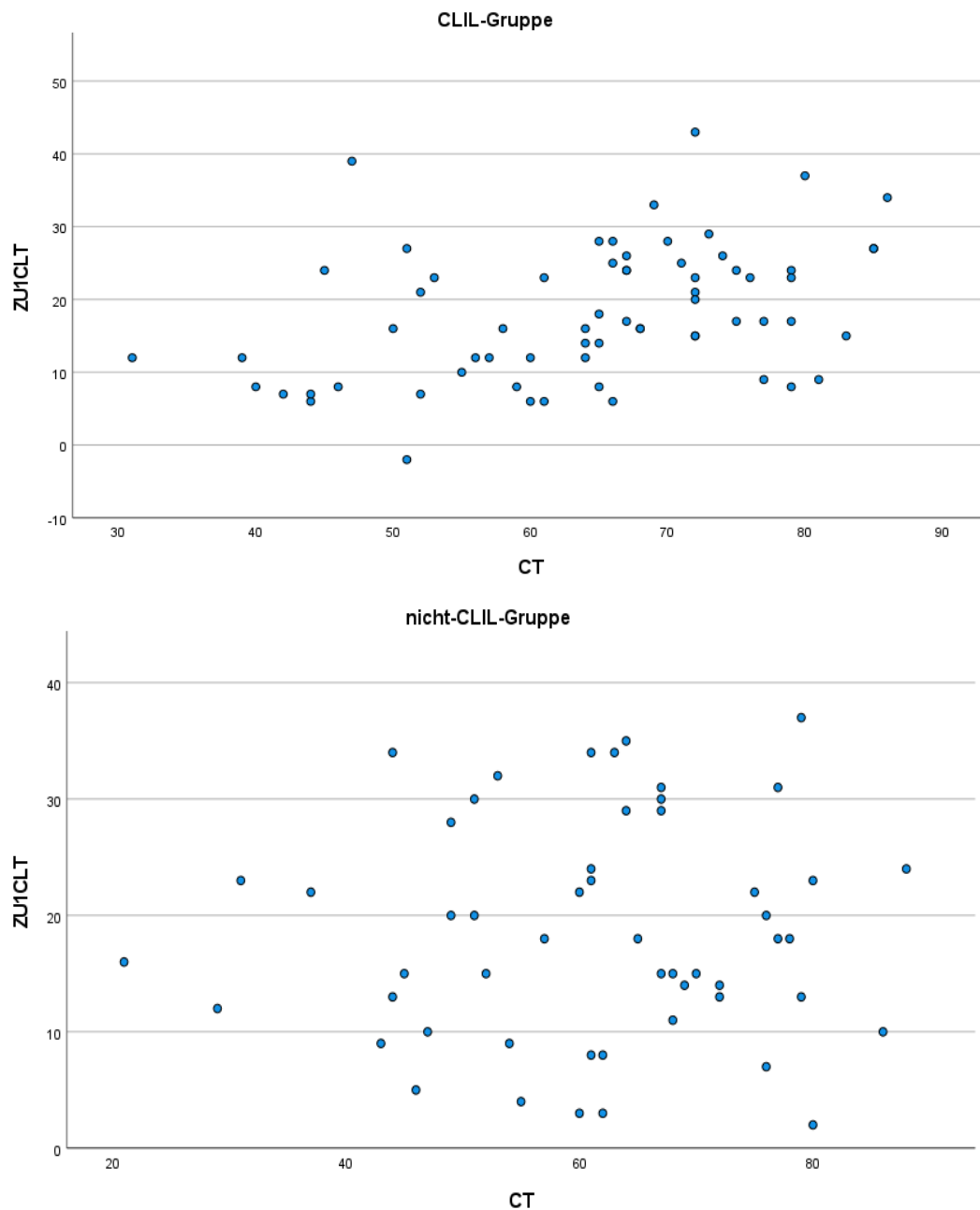


Abbildung 42: Abhängigkeit des Zugewinns von der Leistungsfähigkeit in Englisch

91 Die Punktintervalle der Leistungsbereiche wurden so gewählt, dass sich in jeder Leistungsgruppe ein Drittel der CLIL-Schüler befinden (CT-Terzile).

Die Berechnung der mittleren Zugewinne für diese drei Leistungsgruppen belegt dies. Die Unterschiede liegen bei 23 % im unteren bei und 18 % im oberen Leistungsbereich⁹².

	CT 0-60	CT 61-71	CT 72-100
	M(ZU1CLT) (N;M(CT);M(ACLT))	M(ZU1CLT) (N;M(CT);M(ACLT))	M(ZU1CLT) (N;M(CT);M(ACLT))
nicht-CLIL	17,2 (21;46,6;22,5)	20,9 (18;64,8;23,7)	18,0 (14;78,2;28,2)
CLIL	13,2 (22;49,6;20,5)	19,4 (21;66,0;23,2)	21,9 (23;77,2;26,4)

Tabelle 35: Zugewinn im CL-Test in den drei Leistungsgruppen des C-Tests

Inwieweit diese Unterschiede statistisch signifikant sind, soll wieder mit Hilfe einer zweifaktoriellen ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Leistungsgruppe-CT (Faktorenstufen: CT0-60; CT61-71; CT72-100) überprüft werden⁹³. Die Analyse ergibt für den Faktor Gruppe ($F(1;113) = 0,096$; $p = 0,757$) sowie für den Interaktionsfaktor ($Z(2;113) = 1,846$; $p = 0,163$) keinen signifikanten Effekt, wohl aber für den Faktor Leistungsgruppe-CT ($F(2;113) = 3,954$; $p = 0,022$). Das bedeutet, dass die Leistungsstärke in Englisch zwar einen signifikanten Einfluss auf den Lernerfolg im Modul hat, sich hieraus jedoch keine Unterschiede zwischen den CLIL- und den Regelschülern ergeben.⁹⁴ Die oben berichteten Unterschiede zwischen den leistungsschwachen und den leistungsstarken Schüler CLIL- bzw. nicht-CLIL-Schüler sind somit statistisch nicht signifikant. Differenziert man nun in den Gruppen der leistungsstarken bzw. leistungsschwachen Englischlernern ($CT_{\text{niedrig}} = CT\ 0-60$; $CT_{\text{hoch}} = CT\ 72-100$) nach der Leistungsstärke im Fach Chemie ($ACLT_{\text{niedrig}} = ACLT\ 0-19$; $ACLT_{\text{mittel}} = ACLT\ 20-24$; $ACLT_{\text{hoch}} = ACLT\ 25-57$), so lassen sich jeweils drei weitere Untergruppen bilden.

92 Dabei ist zu berücksichtigen, dass die mittlere Leistungsstärke der Leistungsgruppen sowohl in Bezug auf das Fach Englisch als auch in Bezug auf das Fach Chemie jeweils voneinander abweichen.

93 Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 0,594$; $p = 0,705$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten normalverteilt waren ($Z(124) = 0,076$; $p = 0,076$).

94 Dies widerspricht dem Ergebnis der multiplen linearen Regression, wonach die Englischkompetenz keinen statistisch signifikanten Einfluss auf den Lernerfolg im Modul hat (siehe Tabelle 38). Als nächstes wurde deshalb ein Post-hoc-Test nach Bonferroni durchgeführt. Dieser zeigte, dass ein statistisch signifikanter Unterschied nur zwischen der ersten CT-Gruppe und der dritten vorliegt ($p = 0,31$). Diese CT-Gruppen unterscheiden sich jedoch ebenfalls deutlich in ihren ACLT-Werten, welche hochsignifikant mit dem unmittelbaren Zugewinn im CL-Test korrelieren (siehe Tabelle 38). Anschließend wurde deshalb, nach Prüfung der Voraussetzungen, ein t-Test durchgeführt, um festzustellen, ob die Unterschiede statistisch signifikant sind. Es zeigte sich, dass dies der Fall ist ($T = -3,042$; $p = 0,003$). Der anhand der ANOVA ermittelte signifikante Unterschied der CT-Gruppen ist somit auf die unterschiedlichen ACLT-Werte zurückzuführen.

Die folgenden beiden Tabellen zeigen den mittleren Zugewinn im Chemieleistungstest (ZU1CLT), die CT- und die ACLT-Gruppenmittelwerte sowie die Standardabweichungen der leistungsschwachen Englischlerner in der CLIL und der nicht-CLIL-Gruppe.

			nicht-CLIL			CLIL			ΔM
			N	M	SD	N	M	SD	
CT niedrig	ACLT niedrig	ZU1CLT	6	15,50	10,90	11	8,27	4,54	7,23
		CT	6	40,00	7,90	11	52,18	7,13	-12,18
		ACLT	6	13,67	3,78	11	14,91	3,21	-1,24
	ACLT mittel	ZU1CLT	4	18,75	7,80	7	18,43	7,46	0,32
		CT	4	52,75	3,50	7	46,57	7,68	6,18
		ACLT	4	22,75	1,00	7	22,14	1,57	0,61
	ACLT hoch	ZU1CLT	9	17,44	10,20	3	21,00	15,59	-3,56
		CT	9	47,33	12,50	3	48,67	10,60	-1,34
		ACLT	9	28,22	3,00	3	37,00	8,19	-8,78

Tabelle 36: Zugewinn weniger sprachbegabter Schüler nach Leistungsstärke in Chemie

Es zeigt sich, dass bei Schülern, die sowohl in der Fremdsprache als auch im Sachfach leistungsschwach sind, der Unterschied im Lernerfolg durch das Modul ($\Delta M_{ZUG1CLT}$) zwischen CLIL- und nicht-CLIL-Schülern größer wird. Die Differenz ist im Vergleich zum Mittelwert über alle drei ACLT-Leistungsgruppen ($\Delta M_{ZU1CLT} = 4,0$) mit 7,23 fast doppelt so hoch. Diese Schüler schneiden damit im Mittel besser im Regelunterricht ab. Der Einsatz der Fremdsprache hat in dieser Untergruppe offensichtlich negative Auswirkungen auf die Leistungen im Sachfach und dies, obwohl die CLIL-Schüler im Mittel sowohl in der Fremdsprache ($\Delta M_{CT} = -12,18$) und als auch im Fach Chemie ($\Delta M_{ACLT} = -1,24$) etwas leistungstärker sind. Der Mann-Whitney-U-Test zeigt jedoch, dass der Unterschied im Zugewinn zwischen CLIL- und nicht-CLIL-Schülern auch in dieser Untergruppe statistisch nicht signifikant ist ($U = 46,50$; $p = 0,172$)⁹⁵. Betrachtet man zuletzt den sachfachlichen Lernerfolg im Modul (ZU1CLT) der leistungstarken Englischlerner differenziert nach Leistungsstärke im Fach Chemie, so erkennt man, dass in der Gruppe der Schüler, die sowohl in der Fremdsprache als auch im Fach Chemie leistungstark sind, der Unterschied (ΔM_{ZU1CLT}) weiter ansteigt.

⁹⁵ Aufgrund der geringen Stichprobengröße wurde auf einen nonparametrischen Test zurückgegriffen.

Hauptuntersuchung: Ergebnisse und Auswertung

			nicht-CLIL			CLIL			ΔM
			N	M	SD	N	M	SD	
CT hoch	ACLT niedrig	ZU1CLT	2	20,50	3,54	4	18,50	5,75	2,00
		CT	2	79,00	1,41	4	77,00	6,27	2,00
		ACLT	2	18,00	1,41	4	12,75	1,71	5,25
	ACLT mittel	ZU1CLT	2	14,00	5,66	4	18,50	7,14	-4,50
		CT	2	81,50	6,36	4	77,00	1,63	4,50
		ACLT	2	20,50	0,71	4	22,00	1,83	-1,50
	ACLT hoch	ZU1CLT	10	18,30	10,69	13	23,77	10,43	-5,47
		CT	10	77,40	4,62	13	77,54	5,09	-0,14
		ACLT	10	31,80	6,29	13	32,00	4,93	-0,20

Tabelle 37: Zugewinn sprachbegabter Schüler differenziert nach Leistungsstärke in Chemie

Verglichen mit dem Mittelwert über alle drei ACLT-Leistungsgruppen ($M_{ZU1CLT} = 3,9$) liegt der Unterschied jetzt bei 5,47. Diese Schüler schneiden damit im Mittel besser im CLIL-Unterricht ab. Der Einsatz der Fremdsprache hat in dieser Untergruppe offensichtlich positive Auswirkungen auf die Leistungen im Sachfach. Dabei sind die Unterschiede im mittleren Leistungsvermögen in Englisch ($\Delta M_{CT} = -0,14$) und in Chemie ($\Delta M_{ACLT} = -0,20$) nur sehr gering. Der Mann-Whitney-U-Test zeigt jedoch, dass der Unterschied im Zugewinn zwischen CLIL- und nicht-CLIL-Schülern auch in dieser Untergruppe statistisch nicht signifikant ist ($U = 45,00$; $p = 0,215$)⁹⁶.

Zusammenfassend lässt sich feststellen, dass die Leistungsfähigkeit in der Fremdsprache zwar statistisch signifikant mit der Leistung im CL-Test der CLIL-Schüler korreliert, sich jedoch keine statistisch signifikanten Belege dafür finden, dass Schüler mit geringen Fremdsprachenkenntnissen bessere Leistungen im Regelunterricht erzielen und Schüler mit besonders guten Fremdsprachenkenntnissen bessere Ergebnisse im CLIL-Unterricht erlangen. Dennoch lassen sich aus den hier untersuchten Daten tendenzielle Vorteile für leistungsschwache Englischler im Regelunterricht erkennen, insbesondere, wenn diese auch im Fach Chemie leistungsschwach sind. Außerdem findet man tendenzielle Vorteile für leistungsstarke Englischler im CLIL-Unterricht, insbesondere, wenn diese auch im Fach Chemie leistungsstark sind. Die Korrelationsanalysen haben gezeigt, dass neben den Englischleistungen (CT) und den Chemieleistungen (ACLT) auch die Lesegenauigkeit (LGN), die kognitiven Fähigkeiten (KFT) und das Englischselbstkonzept (SKEN) mit dem sachfachlichen Lernerfolg (ZU1CLT-CLIL) korrelieren (vgl. Tabellen Seite 187 und Seite

⁹⁶ Aufgrund der geringen Stichprobengröße wurde auf einen nonparametrischen Test zurückgegriffen.

188). Mit Hilfe einer multiplen linearen Regression soll nun der Einfluss dieser Variablen auf den Lernerfolg weiter untersucht werden. Das Modell berücksichtigt alle fünf Variablen⁹⁷.

Einfluss auf den Lernerfolg im Modul			
Variable	Unstandardisiert (B)	Standardisiert (β)	Standardfehler (SE)
Konstante	-18,461*		9,1400
SKEN	4,776*	0,276*	2,0640
KFT	-0,0050	-0,0050	0,1520
LGN	0,0310	0,0380	0,0980
CT	0,1410	0,1970	0,0910
ACLT	0,353*	0,336*	0,1530
R ²	0,3320		
korrr. R ²	0,2680		
F (df = 5;52)	5,175***		
* p < 0,05; **p < 0,01; ***p < 0,001			

Tabelle 38: β -Koeffizienten für die Prädiktoren des Lernerfolgs im Modul

Es zeigt sich, dass in diesem Modell nur noch die Leistungsfähigkeit in Chemie (ACLT) sowie das Englischselbstkonzept (SKEN) einen signifikanten Einfluss auf den sachfachlichen Lernerfolg (ZU1CLT-CLIL) haben. Die multiple lineare Regression bestätigt somit das mittels ANOVA und Mann-Whitney-U-Test ermittelte Fehlen eines statistisch signifikanten Einflusses der Englischleistung (CT) (siehe oben). Außerdem zeigt sich, dass auch die kognitiven Fähigkeiten (KFT) sowie die Lesegenauigkeit keinen signifikanten Einfluss auf den Lernerfolg im Modul haben. Die allgemeine Chemieleistungsfähigkeit erweist sich dabei als der beste Prädiktor für den Lernerfolg im bilingualen Chemieunterricht. Durch das Modell können 26,8 % der Varianz aufgeklärt werden.

97 Wie bereits gesehen, korrelieren die unabhängigen Variablen SKEN, KFT, LGN, CT und ACLT mit der abhängigen Variable ZU1CLT-CLIL und bilden einen linearen Zusammenhang ab (siehe Seite 193). Die Korrelationen der unabhängigen Variablen untereinander sind gering und liegen zwischen $r = 0,264$ und $r = 0,600$. Sie sind damit kleiner als $r = 0,6$, so dass keine Multikollinearität vorliegt (Meindl, 2011, S. 242). Die Inspektion der Histogramme ergab eine Normalverteilung der Residuen, die Inspektion des Streudiagramms der standardisierten Residuen lieferte keinen Hinweis auf das Vorliegen von Heteroskedastizität.

7.3.3 Sachfachliche Leistungen im Bereich der *higher order thinking skills*

Die vierte Untersuchungshypothese betrifft die sogenannten *higher order thinking skills*. Es wird vermutet, dass die bilingual unterrichteten Schüler im kognitiv anspruchsvolleren Bereich bessere Leistungen erzielen als die auf Deutsch unterrichteten Schüler (siehe Abschnitt 5.3.1.1). Dazu wird geprüft, ob erstere in den Aufgaben des Anforderungsbereichs 2 und 3 des Sachfachtests signifikant besser abschnitten als die Schüler des Regelunterrichts. Die folgende Tabelle zeigt die deskriptive Statistik zum CL-Test. Dabei wurden wieder nur solche Probanden berücksichtigt, die zu allen drei Zeitpunkten am Test teilgenommen hatten. Fehlten Testergebnisse, so wurden die betreffenden Datensätze von der Berechnung ausgeschlossen (balanciertes Design).

Zeit	Gruppe	N	CLT AFB 2/3 M (SD; SE)
Pre	CLIL	67	9,45 (4,2;0,5)
	nicht-CLIL	57	8,16 (4,2;1,0,6)
	gesamt	124	8,85 (4,3)
Post	CLIL	67	22,34 (9,1;1,1)
	nicht-CLIL	57	20,51 (8,4;1,2)
	gesamt	124	21,50 (8,8)
Follow-Up	CLIL	67	18,78 (9,0;1,1)
	nicht-CLIL	57	17,37 (8,2;1,1)
	gesamt	124	18,13 (8,6)

Tabelle 39: Deskriptive Statistik zum Anforderungsbereich II/III des Chemieleistungstests

Die Analyse des Profilplots bestätigt für beide Gruppen, wie schon im CL-Test, einen sehr deutlichen Anstieg der mittleren Testergebnisse vom Pre- zum Posttest (Zugewinn 1) sowie ein ebenfalls deutliches Absinken der Werte vom Post- zum Follow-Up Test (Vergessen). Die Testergebnisse der nicht-CLIL-Gruppe liegen zu allen drei Zeitpunkten unter denen der CLIL-Gruppe. Der Unterschied bleibt dabei vom Pre- bis zum Follow-Up-Test nahezu konstant und beträgt, wie schon im CL-Test, im Mittel 1,5 Punkte. Die Kurvenverläufe zum Lernerfolg und Vergessen in den Anforderungsbereichen 2 und 3 ähneln damit stark den Verläufen, die für den Gesamttest beobachtet wurden (vgl. Abschnitt 7.3.1).

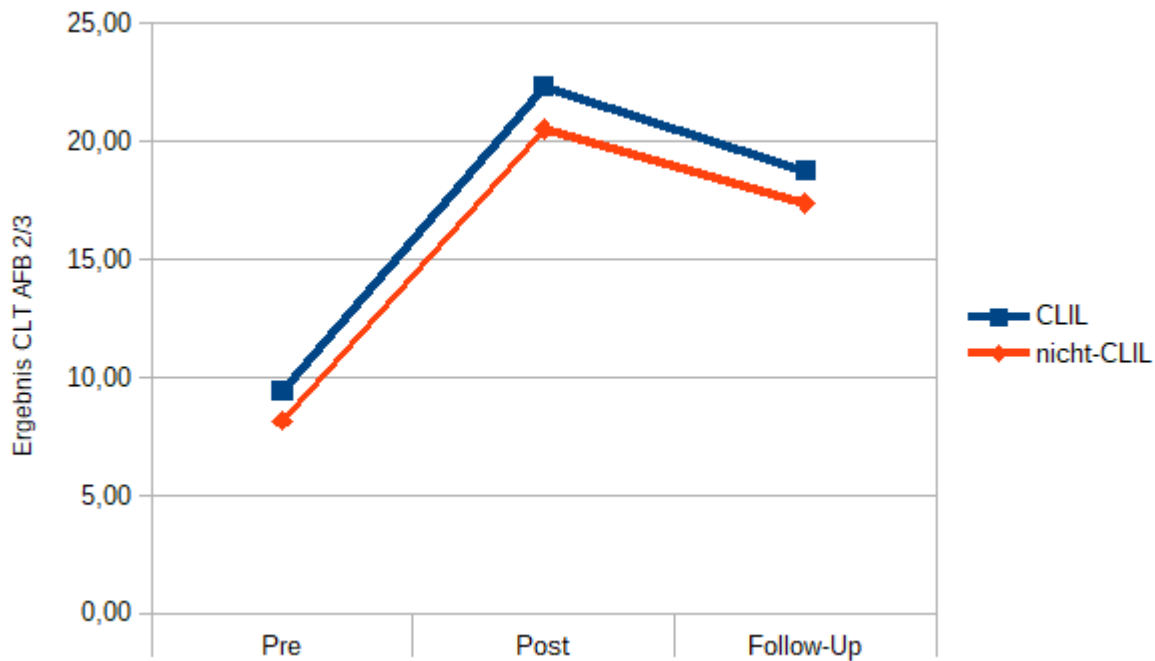


Abbildung 43: Ergebnisse im Anforderungsbereich II/III des Chemieleistungstest

Um zu prüfen, inwieweit die Unterschiede statistisch signifikant sind, wurde eine zweifaktorielle ANOVA mit Messwiederholung (gemischtes Design) berechnet. Dabei wurde zwischen den Faktoren Zeit (Faktorstufen: Pretest, Posttest und Follow-Up-Test) und Gruppe (Faktorstufen: CLIL und nicht-CLIL) unterschieden⁹⁸. Die Auswertung mittels ANOVA ergab, dass für den Faktor Zeit ein signifikanter Unterschied der Pre-, Post- und Follow-Up-Testergebnisse vorliegt ($F(2,244) = 164,5$; $p < 0,01$), für den Faktor Gruppe jedoch nicht ($F(1,122) = 1,993$; $p = 0,161$). Auch die Ergebnisse für die Interaktion von Zeit und Gruppe sind nicht signifikant ($F(2,244) = 0,079$; $p = 0,924$). Für den Faktor Zeit belegen die paarweisen Vergleiche, dass die Unterschiede zwischen den Pre- und den Posttestmittelwerten, zwischen den Post- und den Follow-Up-Testmittelwerten sowie zwischen den Pre- und den Follow-Up-Testwerten in beiden Gruppen statistisch hoch signifikant sind ($p < 0,001$). Wie schon bei der Betrachtung des gesamten Tests, ergeben sich damit auch für den Testteil mit Aufgaben zu den Anforderungsbereichen 2 und 3 keine statistisch signifikanten Unterschiede zwischen den Ergebnissen der CLIL- und den der nicht-CLIL-Schüler. Anschließend

⁹⁸ Während die Bedingungen der Varianzhomogenität nach dem Levene-Test ($p_{CLTAFB2/3PR} = 0,277$, $p_{CLTAFB2/3PO} = 0,0374$, $p_{CLTAFB2/3FU} = 0,737$) und der Sphärizität nach dem Mauchly-Test erfüllt sind ($p = 0,260$), erwiesen sich die Daten nach Inspektion des Histogramms als nicht normalverteilt, wodurch eine Voraussetzung des gewählten Berechnungsverfahrens nicht erfüllt ist. Auch hier wurde, den Hinweisen Bühners & Ziegler (2009, S. 518) zur Robustheit des Verfahrens gegenüber einer Verletzung der Normalverteilungsannahme folgend, eine ANOVA durchgeführt.

wurden - wie schon bei der Analyse der CLT-Ergebnisse - die Zugewinne und das Vergessen einer weiteren Untersuchung unterzogen. Dazu wurden für jeden einzelnen Probanden Werte für den individuellen Zugewinn sowie das Vergessen als Differenzwerte zwischen den Ergebnissen des Post- und des Pretests (Zugewinn 1) bzw. zwischen den Ergebnissen des Follow-Up- und des Posttests (Vergessen) sowie zwischen den Ergebnissen des Follow-Up- und des Pretests (Zugewinn 2) berechnet.

Zeit	Gruppe	N	M (SD; SE)
Zugewinn 1 CLT AFB 2/3	CLIL	67	12,90 (7,4;0,9)
	nicht-CLIL	57	12,35 (7,9;1,0)
Vergessen CLT AFB 2/3	CLIL	67	-3,57 (10,0;1,2)
	nicht-CLIL	57	-3,14 (6,4;0,9)
Zugewinn 2 CLT AFB 2/3	CLIL	67	9,33 (8,5;1,0)
	nicht-CLIL	57	9,21 (6,9;0,9)

Tabelle 40: Zugewinn und Vergessen im Anforderungsbereich II/III des Chemieleistungstest

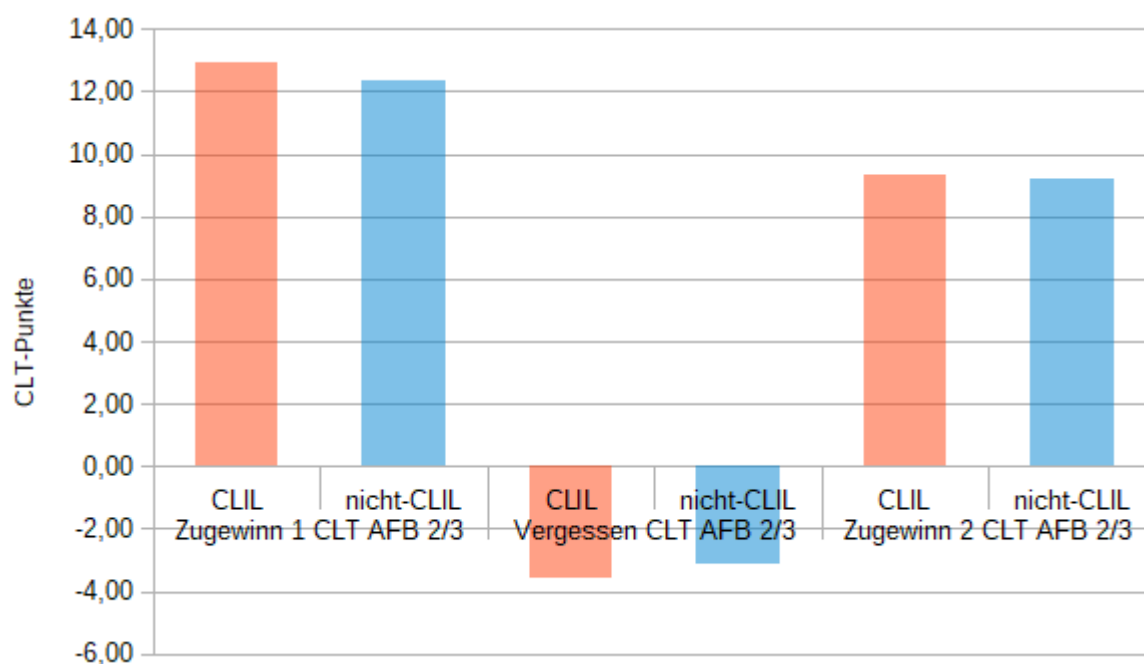


Abbildung 44: Zugewinn und Vergessen im Anforderungsbereich II/III des CL-Tests

Diese wurden dann zusätzlich mittels t-Test auf Signifikanz der Mittelwertsunterschiede untersucht⁹⁹. Es zeigt sich, dass weder die Unterschiede im unmittelbaren Zugewinn (Zugewinn 1:

⁹⁹ Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (Zugewinn 1: $F = 0,368$; $p = 0,545$; Vergessen: $F = 6,118$; $p = 0,150$; Zugewinn 2: $F = 1,828$, $p = 0,179$), das Vorliegen der Normalverteilung anhand

$T = 0,396$, $p = 0,693$) noch die Unterschiede im Vergessen (Vergessen: $T = -0,276$; $p = 0,783$) sowie die Unterschiede im mittelfristigen Zugewinn (Zugewinn 2: $T = 0,084$; $p = 0,933$) statistisch signifikant sind. Damit finden sich in den vorliegenden Daten keine Hinweise darauf, dass die bilingual unterrichteten Kinder den auf Deutsch unterrichteten Kindern in den kognitiv anspruchsvolleren Aufgaben überlegen wären (Hypothese 4).

7.4 Der sprachliche Lernerfolg

7.4.1 Zum Lernerfolg im deutschen Fachwortschatz

Zur Überprüfung der Reliabilität des Tests wurden die Itemschwierigkeiten für Pre- und Posttest sowie Cronbachs Alpha bestimmt. Dabei zeigt sich, dass die Lösungshäufigkeiten (LH) der Multiple-Choice-Items im Pretest erwartungskonform überwiegend im Bereich der Ratewahrscheinlichkeit ($LH = 0.25$) liegen. Anders als im Pilottest, bei dem noch gut die Hälfte dieser Items Lösungshäufigkeiten von mehr als $LH=0.5$ aufwies ist das im Test zur Hauptuntersuchung nur noch bei Item 1 der Fall. Lediglich die Items 13, 14 und 16 zum produktiven Wortschatz weisen hohe Lösungshäufigkeiten im Pretest auf¹⁰⁰. Die Daten zum Posttest zeigen erwartungsgemäß einen deutlichen Anstieg der Lösungshäufigkeiten. Dabei erweisen sich die Items 19 und 20, wie schon im Pilottest, als die schwersten¹⁰¹. Die Werte für Cronbachs Alpha (CA) liegen für den Pretest bei $\alpha = 0,488$ ($N=138$), für den Posttest bei $\alpha = 0,713$ ($N=144$) und für den Follow-Up-Test bei $\alpha = 0,755$ ($N=136$). Die geringe Reliabilität des Pretests ergibt sich dabei aus der Tatsache, dass die Schüler über keinerlei Vorkenntnisse verfügen und eine zuverlässige Messung deshalb nicht möglich ist. Die Reliabilitäten des Posttests und des Follow-Up-Tests sind hingegen akzeptabel. Es zeigt sich zudem, dass die Streichung einzelner Items keine nennenswerte Verbesserung der Reliabilität zur Folge hat. Deshalb können alle 20 Items bei der Auswertung berücksichtigt werden.

des Kolmogorov-Smirnov-Tests (Zugewinn 1: $Z(124) = 0,071$, $p = 0,194$; Zugewinn 2: $Z(124) = 0,076$, $p = 0,079$) bestätigt. Die Daten zum Vergessen waren laut Kolmogorov-Smirnov-Test nicht normalverteilt (Vergessen: $Z(124) = 0,087$, $p < 0,001$). Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Ziegler (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

¹⁰⁰ Bei allen drei Items mussten die Schüler anhand des Elementsymbols den Namen des Alkalimetalls angeben.

Offensichtlich ist das Vorwissen der Schüler hier größer als erwartet.

¹⁰¹ Im Pilottest hatten diese Items die Nummern 27 und 34.

Rezeptiver Wortschatz											
Itemnummer	1	2	3	4	5	6	7	8	9	10	11
LH Pretest	0,53	0,37	0,20	0,39	0,27	0,32	0,29	0,23	0,35	0,41	0,32
LH Posttest	0,83	0,73	0,59	0,87	0,85	0,70	0,39	0,83	0,79	0,82	0,53
CA wenn Item gestrichen*	0,71	0,70	0,73	0,70	0,71	0,72	0,72	0,70	0,70	0,70	0,71
Produktiver Wortschatz											
Itemnummer	12	13	14	15	16	17	18	19	20		
LH Pretest	0,35	0,60	0,78	0,01	0,61	0,00	0,00	0,00	0,01		
LH Posttest	0,96	0,94	0,83	0,74	0,82	0,81	0,83	0,24	0,28		
CA wenn Item gestrichen*	0,71	0,71	0,71	0,68	0,70	0,68	0,68	0,69	0,70		

* Werte beziehen sich auf die Ergebnisse im Posttest

Tabelle 41: Lösungshäufigkeiten und Cronbachs Alpha zum deutschen Fachwortschatztest

Die folgende Tabelle zeigt die deskriptive Statistik zu den DFWT-Ergebnissen in beiden Gruppen. Dabei wurden nur solche Probanden berücksichtigt, die zu allen drei Zeitpunkten am Test teilgenommen hatten. Fehlten Testergebnisse, so wurden die betreffenden Datensätze von der Berechnung ausgeschlossen (balanciertes Design).

Zeit	Gruppe	N	DFWT M (SD; SE)
Pre	CLIL	63	5,89 (2,65;0,33)
	nicht-CLIL	61	6,20 (2,72;0,29)
	gesamt	124	6,04 (2,47;0,22)
Post	CLIL	63	14,25 (2,85;0,36)
	nicht-CLIL	61	14,90 (3,10;0,40)
	gesamt	124	14,57 (2,98;0,27)
Follow-Up	CLIL	63	13,71 (3,38;0,43)
	nicht-CLIL	61	14,05 (2,60;0,33)
	gesamt	124	13,88 (3,01;0,27)

Tabelle 42: Deskriptive Statistik zu den deutschen Fachwortschatztests

Die Analyse des Profilplots bestätigt für beide Gruppen einen starken Anstieg der mittleren Testergebnisse vom Pre- zum Posttest (Zugewinn 1) sowie ein ebenfalls deutliches Absinken der Werte vom Post- zum Follow-Up-Test (Vergessen). Die Testergebnisse der nicht-CLIL-Gruppe liegen dabei zu allen drei Zeitpunkten geringfügig über denen der CLIL-Gruppe. Der Unterschied steigt vom Pre- bis zum Posttest geringfügig von 0,31 Punkte auf 0,65 Punkte an und sinkt dann wieder auf 0,34 Punkte ab. Damit ist der Zugewinn vom Pre- zum Follow-Up-Test nahezu identisch.

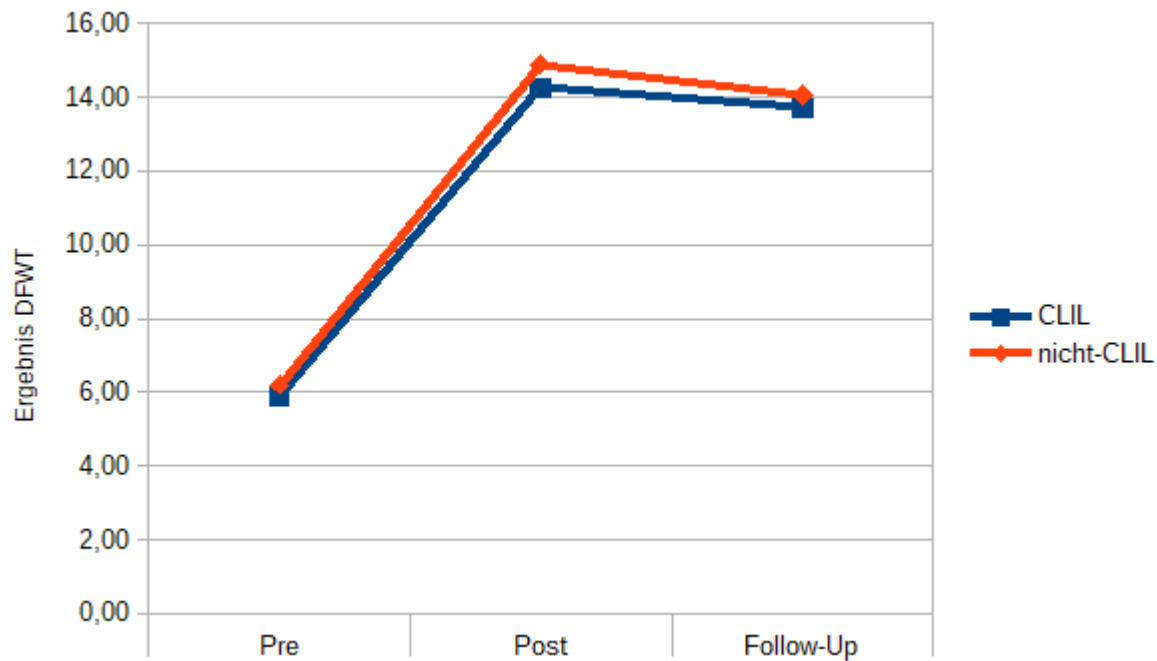


Abbildung 45: Ergebnisse im deutschen Fachwortschatztest

Um zu prüfen, inwieweit die Unterschiede statistisch signifikant sind, wurde eine zweifaktorielle ANOVA mit Messwiederholung (gemischtes Design) berechnet. Dabei wurde zwischen den Faktoren Zeit (Faktorstufen: Pretest, Posttest und Follow-Up Test) und Gruppe (Faktorstufen: CLIL und nicht-CLIL) unterschieden¹⁰². Die Auswertung mittels ANOVA ergab, dass für den Faktor Zeit ein signifikanter Unterschied der Pre-, Post- und Follow-Up-Testergebnisse vorliegt ($F(1,663;244) = 513,795$; $p < 0,01$), für den Faktor Gruppe jedoch nicht ($F(1,122) = 1,298$; $p = 0,257$). Auch die Ergebnisse für die Interaktion von Zeit und Gruppe sind nicht signifikant ($F(1,663;244) = 0,204$; $p = 0,775$). Für den Faktor Zeit belegen die paarweisen Vergleiche, dass die Unterschiede zwischen den Pre- und den Posttestmittelwerten und zwischen den Pre- und den Follow-Up-Testwerten in beiden Gruppen statistisch hoch signifikant sind ($p < 0,001$). Der Unterschied zwischen den Post- und Follow-Up-Testwerten ist statistisch signifikant ($p = 0,007$). Anschließend wurden die Zugewinne und das Vergessen einer weiteren Untersuchung unterzogen. Dazu wurden für jeden einzelnen Probanden Werte für den individuellen Zugewinn und das Vergessen als Differenzwerte

¹⁰² Während die Bedingung der Varianzhomogenität nach dem Levene-Test ($p_{DFWPR} = 0,169$; $p_{DFWPO} = 0,761$; $p_{DFWFO} = 0,13$) erfüllt war, konnte mithilfe des Mauchly-Tests keine Sphärizität festgestellt werden ($p < 0,001$), weshalb eine Korrektur nach Greenhouse-Geisser durchgeführt wurde. Zwar erwiesen sich die Daten nach dem Kolmogorov-Smirnov-Test als nicht normalverteilt, die Werte für Schiefe und Kurtosis wichen jedoch nur geringfügig von null ab. Die Varianzanalyse ist nach Bühner & Ziegler (2009) relativ robust gegenüber einer Verletzung der Normalverteilungsannahme, wenn die Stichproben, wie bei der vorliegenden Untersuchung der Fall, etwa gleich groß sind.

zwischen den Ergebnissen des Post- und Pretests (Zugewinn 1) bzw. zwischen den Ergebnissen des Follow-Up- und Posttests (Vergessen) sowie zwischen den Ergebnissen des Follow-Up- und Pretests (Zugewinn 2) berechnet. Diese wurden schließlich zusätzlich mittels t-Test auf Signifikanz der Mittelwertsunterschiede untersucht¹⁰³. Es zeigt sich, dass weder die Unterschiede im unmittelbaren Zugewinn (Zugewinn 1: $T = -0,499$, $p = 0,619$) noch die Unterschiede im Vergessen (Vergessen: $T = 0,702$; $p = 0,0484$) sowie die Unterschiede im mittelfristigen Zugewinn (Zugewinn 2: $T = -0,044$; $p = 0,965$) statistisch signifikant sind.

Zeit	Gruppe	N	M (SD; SE)
Zugewinn 1 DFWT	CLIL	63	8,36 (3,63; 0,46)
	nicht-CLIL	61	8,7 (3,95; 0,51)
Vergessen DFWT	CLIL	63	-0,54 (2,31; 0,29)
	nicht-CLIL	61	-0,85 (2,64; 0,34)
Zugewinn 2 DFWT	CLIL	63	7,82 (3,68; 0,46)
	nicht-CLIL	61	7,85 (3,22; 0,41)

Abbildung 46: Zugewinn und Vergessen im deutschen Fachwortschatztest

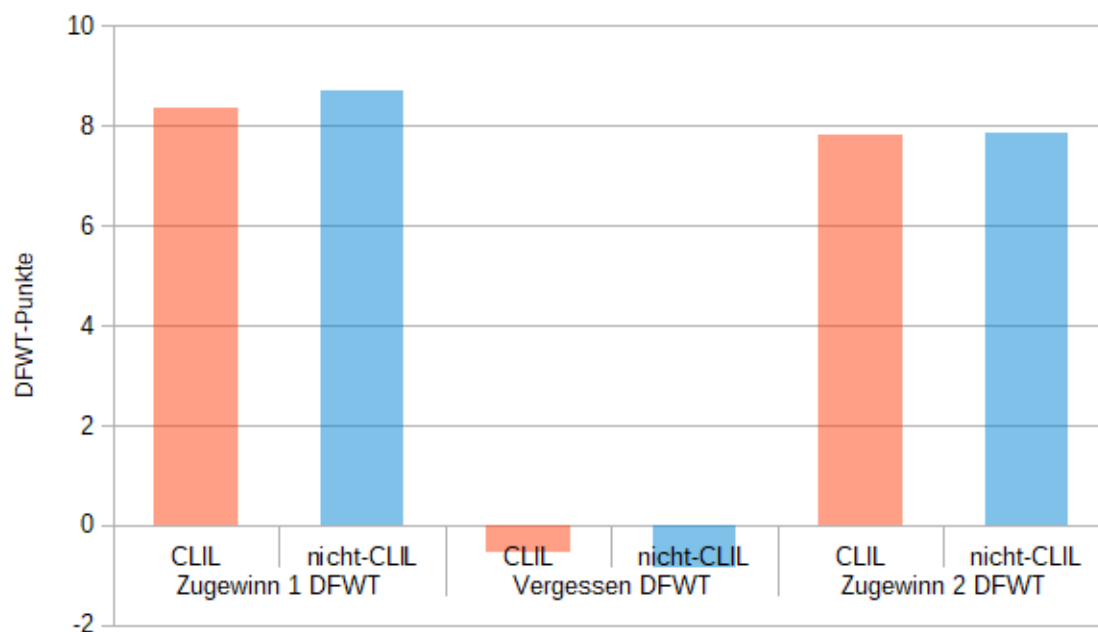


Abbildung 47: Zugewinn und Vergessen im deutschen Fachwortschatztest

¹⁰³ Das Vorliegen von Varianzhomogenität wurde mit Hilfe des Levene-Tests (Zugewinn 1: $F = 0,089$; $p = 0,766$; Vergessen: $F = 0,000$; $p = 0,987$; Zugewinn 2: $F = 1,148$, $p = 0,286$) bestätigt. Zwar erwiesen sich die Daten nach dem Kolmogorov-Smirnov-Test als nicht normalverteilt, die Werte für Schiefe und Kurtosis wichen jedoch nur geringfügig von null ab. Der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend, wurde deshalb ein t-Test durchgeführt.

Es konnten in Bezug auf die Verfügbarkeit des deutschen Fachwortschatzes somit keine signifikanten Leistungsunterschiede zwischen CLIL- und nicht-CLIL-Schülern festgestellt werden. Die bilingual unterrichteten Schüler erzielten vergleichbar gute Ergebnisse wie die Regelschüler - sowohl unmittelbar im Anschluss an das Modul als auch sechs Wochen später. Dies gilt für den rezeptiven wie auch für den produktiven Wortschatz. Die Ergebnisse in den Testteilen zum rezeptiven Wortschatz sind dabei in beiden Gruppen nahezu identisch. Die Differenz zwischen den Gruppenmittelwerten liegt hier zwischen 0,1 und 0,3 Punkten. Im produktiven Testteil liegen die Ergebnisse der CLIL-Gruppe zwar zu allen drei Testzeitpunkten etwa 0,5 Punkte unter den Ergebnissen der nicht-CLIL-Gruppe, Zugewinn und Vergessen sind jedoch nahezu identisch. Anschließend wurde überprüft, ob der mittelfristige Zugewinn im deutschen Fachwortschatztest mit dem Fachinteresse in Chemie (FICH), dem Chemieselbstkonzept (SKCH) und der Lesegenauigkeit korreliert. Es ist denkbar, dass die CLIL-Schüler nur deshalb vergleichbar gute Werte im deutschen Fachwortschatz erzielten wie die nicht-CLIL-Schüler, weil sie ein größeres Interesse an Chemie haben und über ein besseres Selbstkonzept verfügen. Der Vergleich der Stichproben hatte für diese Variablen signifikant bessere Werte in der CLIL-Gruppe ergeben (siehe Abschnitt 7.1). Dazu wurde jeweils der Korrelationskoeffizient nach Pearson berechnet¹⁰⁴.

	ZU2DFWST-CLIL
FICH	-0,099
SKCH	0,174
LGN	,337**
* $p < 0,05$; ** $p < 0,01$	

Tabelle 43 Korrelationen zum mittelfristigen Zugewinn im deutschen Fachwortschatz

Es zeigt sich, dass weder das Fachinteresse noch das Selbstkonzept, mit dem mittelfristigen Lernerfolg im deutschen Fachwortschatz korrelieren, wohl aber die Lesegenauigkeit.

Interessanterweise findet sich jedoch mit $r = 0,193$ keine statistisch signifikante Korrelation der Lesegenauigkeit mit dem mittelfristigen Zugewinn im englischen Fachwortschatztests

¹⁰⁴ Für die intervallskalierten Variablen wurde der Pearson-Korrelationskoeffizient berechnet, da keine Ausreißer vorlagen und die Inspektion der Streudiagramme jeweils einen linearen Zusammenhang erkennen ließ. Die Werte von Schiefe und Kurtosis waren in allen Fällen nahe bei null, so dass zumindest von einer annähernd univariaten Normalverteilung ausgegangen werden kann. Die bivariate Normalverteilung konnte statistisch nicht überprüft werden. Wegen der Gruppengrößen ($N > 30$) wird jedoch von einer bivariaten Normalverteilung ausgegangen (zentraler Grenzwertsatz).

(ZU2EFWST-CLIL). Das kann bedeuten, dass der Transfer der englischen Fachbegriffe zurück ins Deutsche nicht automatisch erfolgt, sondern genaues Lesen der deutschen Fachtexte erfordert.

7.4.2 Zum Lernerfolg im englischen Fachwortschatz

Zur Überprüfung der Reliabilität des Tests wurden – wie schon beim deutschen Fachwortschatztest – die Itemschwierigkeiten für Pre- und Posttest sowie Cronbachs Alpha bestimmt.

Rezeptiver Wortschatz											
Itemnummer	1	2	3	4	5	6	7	8	9	10	11
LH Pretest	0,46	0,27	0,11	0,39	0,33	0,27	0,26	0,33	0,27	0,39	0,20
LH Posttest	0,77	0,78	0,32	0,88	0,84	0,77	0,22	0,90	0,89	0,74	0,56
LH Posttest DFWT	0,83	0,73	0,59	0,87	0,85	0,70	0,39	0,83	0,79	0,82	0,53
Differenz LH (EFWT-DFWT)	-0,07	0,05	-0,28	0,01	-0,02	0,07	-0,17	0,07	0,10	-0,08	0,03
CA wenn Item gestrichen*	0,74	0,75	0,74	0,74	0,74	0,74	0,76	0,74	0,74	0,74	0,75
Produktiver Wortschatz											
Itemnummer	12	13	14	15	16	17	18	19	20		
LH Pretest	0,04	0,46	0,74	0,04	0,03	0,00	0,01	0,00	0,00		
LH Posttest	0,75	0,99	0,84	0,58	0,66	0,85	0,81	0,18	0,26		
LH Posttest DFWT	0,96	0,94	0,83	0,74	0,82	0,81	0,83	0,24	0,28		
Differenz LH (EFWT-DFWT)	-0,20	0,05	0,00	-0,17	-0,16	0,04	-0,02	-0,06	-0,02		
CA wenn Item gestrichen*	0,73	0,76	0,77	0,74	0,74	0,73	0,73	0,74	0,75		

* Werte beziehen sich auf die Ergebnisse im Posttest

Tabelle 44: Lösungshäufigkeiten und Cronbachs Alpha des englischen Fachwortschatztests

Dabei zeigt sich, dass die Lösungshäufigkeiten der Multiple-Choice-Items im Pretest erwartungskonform überwiegend im Bereich der Ratewahrscheinlichkeit ($LH = 0,25$) liegen¹⁰⁵.

Beim produktiven Wortschatz liegt die Lösungshäufigkeit, wie erwartet, nahe bei null¹⁰⁶. Die Daten zum Posttest zeigen wieder einen deutlichen Anstieg der Lösungshäufigkeiten. Weil die Ergebnisse des EFWT später mit denen des DFWT verglichen werden sollen, werden die Lösungshäufigkeiten beider Tests im Folgenden gegenübergestellt. Die mittlere Lösungshäufigkeit liegt im EFWT mit $LH = 0,68$ nur 4 % unter der mittleren Lösungshäufigkeit des DFWT mit $LH = 0,72$ ¹⁰⁷. Fünfzehn der 20 Items des EFWT waren völlig analog zu den Items des DFWT konzipiert und unterschieden

¹⁰⁵ Item 1 wies dabei - wie bereits im deutschen Fachwortschatztest - eine erhöhte Lösungshäufigkeit ($LH = 0,46$) auf.

¹⁰⁶ Lediglich die Items 13 und 14 hatten - wie im DFWT - hohe Lösungshäufigkeiten im Pretest. Bei beiden Items mussten die Schüler anhand des Elementsymbols den Namen des Alkalimetalls angeben. Da die englischen Namen im Fall von Item 13 und 14 den deutschen Namen entsprechen, konnten die Schüler hier - wie schon im deutschen Fachwortschatztest - die Stoffnamen aus den Elementsymbolen herleiten. Anders als im DFWT gelang das für Item 16 jedoch diesmal nicht, da das Elementsymbol (Na) im Englischen keinen Hinweis auf den Namen (*sodium*) gibt.

¹⁰⁷ Dabei erwiesen sich die Items 3,7,19 und 20, wie im DFWT, als die schwersten.

sich lediglich in der Sprache des Beispielsatzes und des gesuchten Fachbegriffs. Bei fünf Items zum rezeptiven Wortschatz (Items 1, 3, 6, 9 und 11) waren jedoch im EFWT andere Distraktoren gewählt worden (vgl. Abschnitt 5.3.2.4). Ein Vergleich der Lösungshäufigkeiten der Posttests zeigt, dass der Unterschied für vier der fünf Items weniger als 10 % beträgt. Lediglich für Item 3 betrug die Differenz 28 %. Die Werte für Cronbachs Alpha liegen für den Pretest bei $\alpha = 0,403$ ($N = 69$), für den Posttest bei $\alpha = 0,752$ ($N = 73$) und für den Follow-Up-Test bei $\alpha = 0,777$ ($N = 69$). Die geringe Reliabilität des Pretests ergibt sich dabei aus der Tatsache, dass die Schüler über keinerlei Vorkenntnisse verfügen und eine zuverlässige Messung deshalb nicht möglich ist. Die Reliabilitäten des Posttests und des Follow-Up-Tests sind hingegen akzeptabel. Es zeigt sich zudem, dass die Streichung einzelner Items keine nennenswerte Verbesserung der Reliabilität zur Folge hat. Deshalb können, wie schon beim deutschen Fachwortschatztest, alle 20 Items bei der Auswertung berücksichtigt werden. Die folgende Tabelle zeigt die deskriptive Statistik zu den EFWT- und den DFWT-Ergebnissen der CLIL-Schüler. Dabei wurden nur solche Probanden berücksichtigt, die zu allen drei Zeitpunkten an den Tests teilgenommen hatten. Fehlten Testergebnisse, wurden die betreffenden Datensätze von der Berechnung ausgeschlossen (balanciertes Design).

Zeit	Test	N	M (SD;SE)
Pre	DFWT	63	5,89 (2,65;0,33)
	EFWT	63	4,81 (2,11;0,27)
Post	DFWT	63	14,25 (2,86;0,36)
	EFWT	63	13,49 (3,37;0,43)
Follow-Up	DFWT	63	13,71 (3,38;0,43)
	EFWT	63	13,29 (3,32;0,42)

Tabelle 45: Ergebnisse der CLIL-Schüler im deutschen und im englischen Fachwortschatztest

Die Analyse des Profilplots bestätigt – wie schon für den DFWT - auch für den EFWT einen starken Anstieg der mittleren Testergebnisse vom Pre- zum Posttest (Zugewinn 1) sowie ein ebenfalls deutliches Absinken der Werte vom Post- zum Follow-Up-Test (Vergessen). Die Testergebnisse im DFWT liegen dabei zu allen drei Zeitpunkten geringfügig über denen im EFWT. Der Unterschied sinkt vom Pre- bis zum Posttest von 1,08 Punkte auf 0,76 Punkte und dann weiter auf 0,42 Punkte. Damit ist der Zugewinn vom Pre- zum Follow-Up-Test im EFWT um 0,7 Punkte größer.

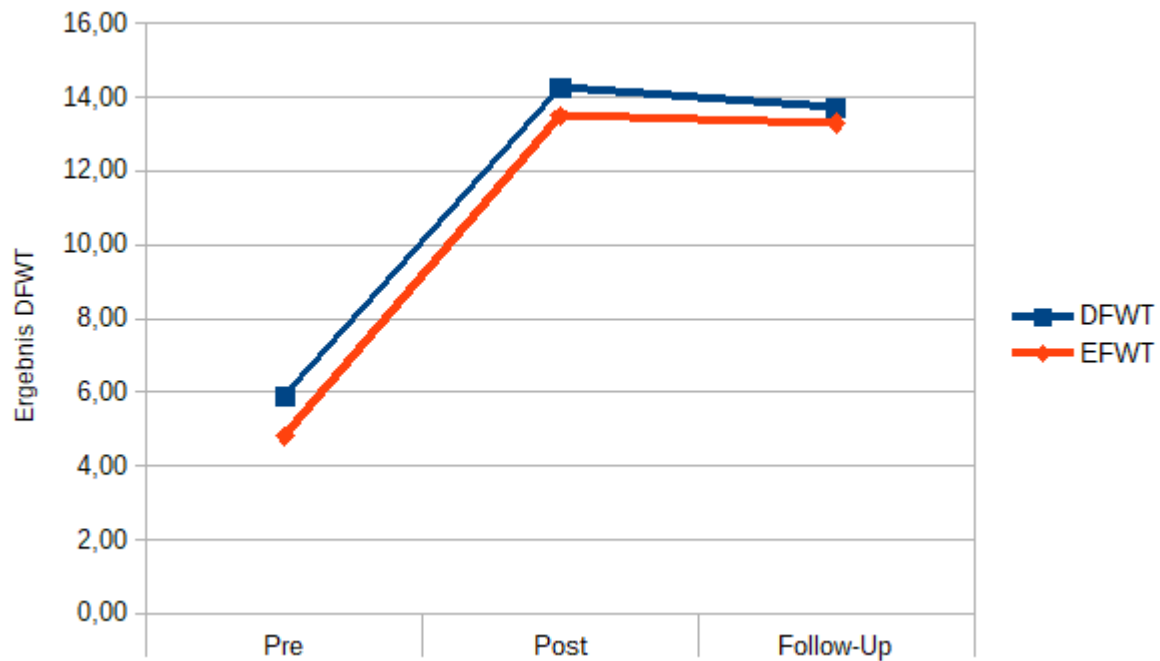


Abbildung 48: Ergebnisse der CLIL-Schüler im deutschen und englischen Fachwortschatztest

Um zu prüfen, inwieweit die Unterschiede statistisch signifikant sind, wurde sowohl für den DFWT als auch für den EFWT eine einfaktorielle ANOVA gerechnet¹⁰⁸. Die Auswertung mittels ANOVA ergab für den DFWT, dass für den Faktor Zeit ein signifikanter Unterschied zwischen den Pre-, Post- und Follow-Up-Testergebnissen vorliegt ($F(1,600;99,175) = 258,202$; $p < 0,01$). Die paarweisen Vergleiche belegen, dass die Unterschiede zwischen den Pre- und den Posttestmittelwerten und zwischen den Pre- und den Follow-Up-Testwerten statistisch hoch signifikant sind ($p < 0,001$). Der Unterschied zwischen den Post- und Follow-Up-Testwerten ist hingegen statistisch nicht signifikant ($p = 0,206$). Für den EFWT lieferte die ANOVA erwartungsgemäß ein ähnliches Ergebnis¹⁰⁹. Auch hier ergab sich für den Faktor Zeit ein signifikanter Unterschied zwischen den Pre-, Post- und Follow-Up-Testergebnissen

¹⁰⁸ Zwar erwiesen sich die Daten nach dem Kolmogorov-Smirnov-Test als nicht normalverteilt, die Werte für Schiefe und Kurtosis wichen jedoch nur geringfügig von null ab. Die Varianzanalyse ist nach Bühner & Ziegler (2009) relativ robust gegenüber einer Verletzung der Normalverteilungsannahme, wenn die Stichproben, wie bei der vorliegenden Untersuchung der Fall, etwa gleich groß sind. Da der Mauchly-Tests für den DFWT keine Sphärizität erkennen ließ ($p < 0,001$), wurde eine Korrektur nach Greenhouse-Geisser durchgeführt.

¹⁰⁹ Zwar erwiesen sich die Daten nach dem Kolmogorov-Smirnov-Test als nicht normalverteilt, die Werte für Schiefe und Kurtosis wichen jedoch nur geringfügig von null ab. Die Varianzanalyse ist zudem nach Bühner & Ziegler (2009) relativ robust gegenüber einer Verletzung der Normalverteilungsannahme, wenn die Stichproben, wie bei der vorliegenden Untersuchung der Fall, etwa gleich groß sind. Da der Mauchly-Tests für den DFWT keine Sphärizität erkennen ließ ($p < 0,001$), wurde eine Korrektur nach Greenhouse-Geisser durchgeführt.

($F(1,825;113,127) = 263,005$; $p < 0,01$). Die paarweisen Vergleiche belegen, dass die Unterschiede zwischen den Pre- und den Posttestmittelwerten und zwischen den Pre- und den Follow-Up-Testwerten statistisch hoch signifikant sind ($p < 0,001$). Der Unterschied zwischen den Post- und Follow-Up-Testwerten ist hingegen statistisch nicht signifikant ($p = 1,000$).

Anschließend wurden die Zugewinne und das Vergessen einer weiteren Untersuchung unterzogen. Dazu wurden für jeden einzelnen Probanden Werte für den individuellen Zugewinn und das Vergessen als Differenzwerte zwischen den Ergebnissen im Post- und Pretest (Zugewinn 1) bzw. zwischen den Ergebnissen im Follow-Up-Test und Posttest (Vergessen) sowie zwischen den Ergebnissen im Follow-Up-Test und Pretest (Zugewinn 2) berechnet. Diese wurden dann zusätzlich mittels t-Test auf Signifikanz der Mittelwertsunterschiede untersucht¹¹⁰. Es zeigt sich, dass weder der Unterschied im unmittelbaren Zugewinn (Zugewinn 1: $T = -0,842$, $p = 0,403$) noch der Unterschied im Vergessen (Vergessen: $T = -0,879$; $p = 0,383$) bzw. im mittelfristigen Zugewinn (Zugewinn 2: $T = -1,663$; $p = 0,101$) statistisch signifikant sind.

Zeit	Test	N	M (SD; SE)
Zugewinn 1	DFWT	63	8,36 (3,63;0,46)
	EFWT	63	8,68 (3,89;0,49)
Vergessen	DFWT	63	-0,54 (2,31;0,29)
	EFWT	63	-0,21 (2,47;0,31)
Zugewinn 2	DFWT	63	7,82 (3,68;0,46)
	EFWT	63	8,48 (3,81;0,48)

Tabelle 46: Zugewinn und Vergessen der CLIL-Schüler in den Fachwortschatztests

¹¹⁰ Zwar erwiesen sich die Daten nach dem Kolmogorov-Smirnov-Test als nicht normalverteilt, die Werte für Schiefe und Kurtosis wichen jedoch nur geringfügig von null ab. Der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend, wurde ein t-Test durchgeführt.

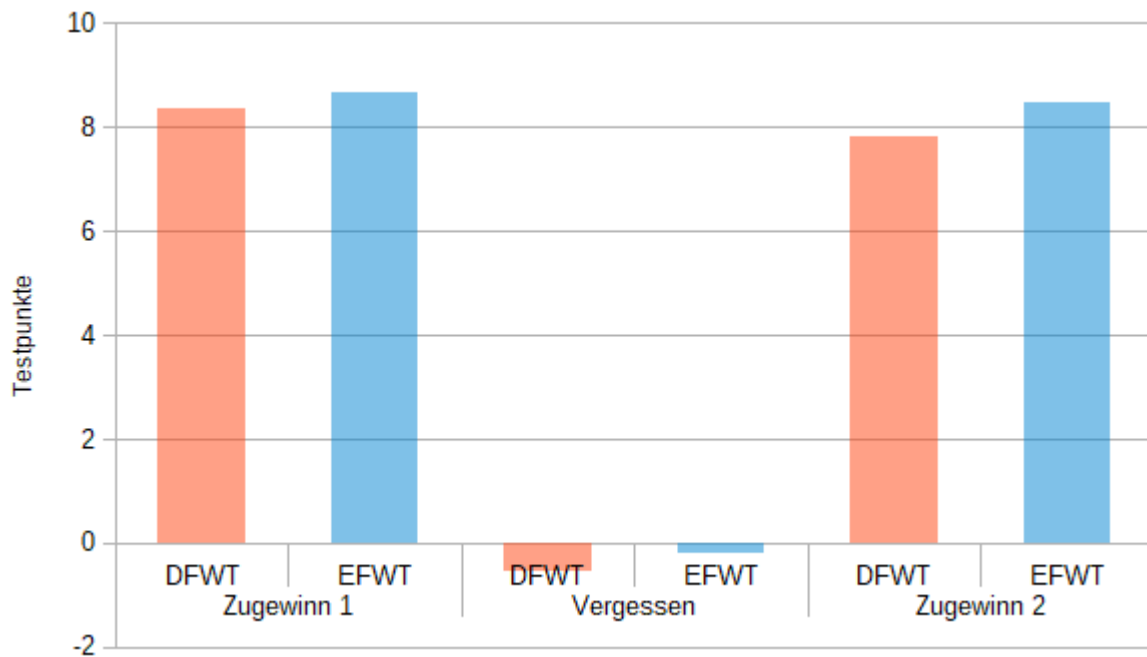


Abbildung 49: Zugewinn und Vergessen der CLIL-Schüler in den Fachwortschatztests

In Bezug auf die Verfügbarkeit des englischen und des deutschen Fachwortschatzes konnten somit keine signifikanten Unterschiede bei den CLIL-Schülern festgestellt werden. Die unmittelbaren und die mittelfristigen Zugewinne im EFWT sind zwar etwas größer und das Vergessen etwas geringer als im DFWT, diese Unterschiede sind jedoch statistisch nicht signifikant. Ganz ähnliche Ergebnisse findet man für den rezeptiven wie auch für den produktiven Fachwortschatz. Unterschiede sind lediglich tendenziell und in keinem Fall statistisch signifikant. Die Testmittelwerte im DFWT sind dabei zu allen drei Zeitpunkten etwas größer als im EFWT. Der mittelfristige Zugewinn ist aber in beiden Teilen im EFWT etwas größer als im DFWT. Es fällt außerdem auf, dass das Vergessen in beiden Tests im rezeptiven Teil geringer ausfällt als im produktiven Teil. Im EFWT schneiden die Schüler im rezeptiven Teil des Follow-Up-Tests sogar besser ab als im Posttest. Die Daten zeigen, dass die parallele Vermittlung der deutschen und der englischen Fachbegriffe im Mittel zu vergleichbaren Ergebnissen geführt hat.

7.4.3 Zum fremdsprachlichen Lernerfolg im Fach- und Allgemeinwortschatz

Zur Überprüfung der Reliabilität des Tests wurden die Itemschwierigkeiten für Pre- und Posttest sowie Cronbachs Alpha bestimmt.

Hauptuntersuchung: Ergebnisse und Auswertung

	Rezeptiver Wortschatz														
Itemnummer	1	2	3	4	5	6	7	8	9	10					
LH Pretest	0,32	0,51	0,70	0,41	0,38	0,29	0,72	0,32	0,33	0,57					
LH Posttest	0,93	0,96	0,90	0,88	0,92	0,78	0,88	0,64	0,84	0,93					
CA wenn Item gestrichen*	0,87	0,87	0,88	0,87	0,87	0,87	0,87	0,87	0,86	0,88					
	Produktiver Wortschatz														
Itemnummer	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
LH Pretest	0,06	0,01	0,00	0,01	0,20	0,03	0,74	0,01	0,03	0,32	0,04	0,12	0,06	0,36	0,00
LH Posttest	0,70	0,88	0,73	0,55	0,78	0,82	0,93	0,95	0,95	0,95	0,93	0,82	0,93	0,89	0,71
CA wenn Item gestrichen*	0,87	0,87	0,86	0,87	0,87	0,86	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,86
* Werte beziehen sich auf die Ergebnisse im Posttest															

Tabelle 47: Lösungshäufigkeiten und Cronbachs Alpha im ZEFWT

Dabei zeigt sich für den zusätzlich erworbenen Fachwortschatz, dass die Lösungshäufigkeiten der Multiple-Choice-Items zum rezeptiven Wortschatz im Pretest erwartungskonform überwiegend im Bereich der Ratewahrscheinlichkeit ($LH = 0,25$) liegen. Erhöhte Werte zwischen 0,5 und 0,7 finden sich für die Items 2, 3, 7 und 10. Im produktiven Teil sind die Lösungshäufigkeiten deutlich niedriger und erwartungsgemäß in fast allen Fällen nahe bei null. Hier weisen die Items 15, 17, 20 und 24 höhere Werte auf¹¹¹. Die Daten zum Posttest zeigen wieder einen deutlichen Anstieg der Lösungshäufigkeiten. Während die Lösungshäufigkeiten im Pretest im Mittel bei 0,26 liegen, findet sich für den Posttest ein Mittelwert von 0,80. Dies entspricht einem mittleren Anstieg von 0,54. Auch für die Items, die bereits im Pretest höhere Lösungshäufigkeiten hatten, findet sich im Posttest ein weiterer Anstieg, der zwischen 0,16 (Item 7) und 0,63 (Item 20) liegt. Die Werte für Cronbachs Alpha betragen im Pretest $\alpha = 0,312$ ($N = 69$), im Posttest $\alpha = 0,874$ ($N = 73$) und im Follow-Up-Test $\alpha = 0,758$ ($N = 68$). Die geringe Reliabilität des Pretests ergibt sich dabei aus der Tatsache, dass die Schüler über wenig Vorkenntnisse verfügen und eine zuverlässige Messung deshalb nicht möglich ist. Die Reliabilitäten des Posttests und des Follow-Up-Tests sind hingegen akzeptabel bis gut. Man erkennt zudem, dass die Streichung einzelner Items keine nennenswerte Verbesserung der Reliabilität zur Folge hat. Deshalb können alle 25 Items bei der Auswertung berücksichtigt werden. Für den zusätzlich erworbenen Allgemeinwortschatz zeigt sich, dass die Lösungshäufigkeiten der Multiple-Choice-Items im Pretest im Mittel bei $LH = 0,22$ liegen.

¹¹¹ Die erhöhten Werte müssen nicht auf eine Schwäche des Items hindeuten, sondern können auch bedeuten, dass die entsprechenden Begriffe einer größeren Zahl von Schülern bereits vor Modulstart bekannt waren. Es liegen jedoch keine Vergleichswerte aus der Pilotierung vor, da dort aus organisatorischen Gründen kein Pretest durchgeführt wurde.

Hauptuntersuchung: Ergebnisse und Auswertung

Rezeptiver Wortschatz													
Itemnummer	1	2	3	4	5	6	7	8	9	10			
LH Pretest	0,43	0,17	0,20	0,06	0,52	0,74	0,43	0,74	0,32	0,52			
LH Posttest	0,82	0,74	0,78	0,77	0,82	0,96	0,81	0,97	0,77	0,86			
CA wenn Item gestrichen*	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92			
Produktiver Wortschatz													
Itemnummer	11	12	13	14	15	16	17	18	19	20	21	22	23
LH Pretest	0,01	0,01	0,06	0,07	0,13	0,13	0,06	0,00	0,22	0,25	0,00	0,00	0,01
LH Posttest	0,73	0,47	0,64	0,66	0,74	0,90	0,34	0,51	0,82	0,85	0,55	0,53	0,66
CA wenn Item gestrichen*	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92	0,92

* Werte beziehen sich auf die Ergebnisse im Posttest

Tabelle 48: Lösungshäufigkeiten der Items und Cronbachs Alpha des ZEAWT

Während im produktiven Teil alle Lösungshäufigkeiten im Pretest nahe bei null liegen, weichen die Werte der Items 1, 5, 6, 7, 8 und 10 im rezeptiven Teil deutlich nach oben ab. Auch diese Items erreichten jedoch im Posttest höhere Werte, wobei Item 6 den geringsten Zuwachs von 0,22 erfährt. Die mittlere Lösungshäufigkeit im Posttest liegt bei 0,73, der mittlere Zuwachs beträgt somit 0,51. Die Werte für Cronbachs Alpha betragen im Pretest $\alpha = 0,604$ ($N = 69$), im Posttest $\alpha = 0,924$ ($N = 73$) und im Follow-Up-Test $\alpha = 0,855$ ($N = 68$). Die geringe Reliabilität des Pretests ergibt sich dabei wieder aus der Tatsache, dass die Schüler über wenig Vorkenntnisse verfügen und eine zuverlässige Messung deshalb nicht möglich ist. Die Reliabilitäten des Posttests und des Follow-Up-Tests sind hingegen gut. Es zeigt sich zudem, dass die Streichung einzelner Items keine nennenswerte Verbesserung der Reliabilität zur Folge hat. Deshalb können alle 23 Items bei der Auswertung berücksichtigt werden. Die folgende Tabelle zeigt die deskriptive Statistik zu den Ergebnissen der beiden Tests zum zusätzlich erworbenen Wortschatz. Dabei wurden nur solche Probanden berücksichtigt, die zu allen drei Zeitpunkten am Test teilgenommen hatten. Fehlten Testergebnisse, so wurden die betreffenden Datensätze von der Berechnung ausgeschlossen (balanciertes Design).

Zeit	Test	N	M (SD; SE)
Pre	ZEFWT	63	6,46 (2,56; 0,28)
	ZEAWT	63	5,03 (2,45; 0,31)
Post	ZEFWT	63	21,05 (4,39; 0,55)
	ZEAWT	63	16,68 (6,01; 0,76)
Follow-Up	ZEFWT	63	18,73 (3,86; 0,49)
	ZEAWT	63	13,67 (4,87; 0,61)

Tabelle 49: Deskriptive Statistik zum ZEFWT und zum ZEAWT

Die Analyse des Profilplots bestätigt für beide Tests einen starken Anstieg der mittleren Testergebnisse vom Pre- zum Posttest (Zugewinn 1) sowie ein ebenfalls deutliches Absinken der Werte vom Post- zum Follow-Up-Test (Vergessen). Dabei ist zu berücksichtigen, dass der ZEFWT 25 Items und der ZEAWT 23 Items umfasst und die Ergebnisse im ZEFWT auch deshalb etwas höher ausfallen.

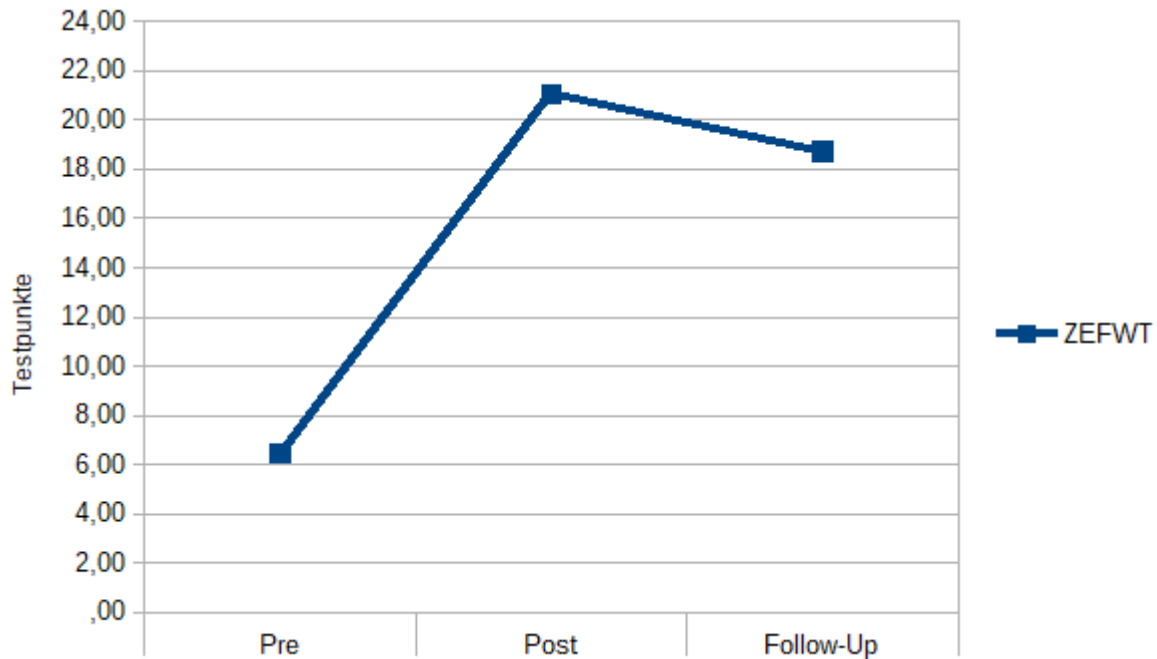


Abbildung 50: Ergebnisse im Test zum zusätzlich erworbenen Fachwortschatz

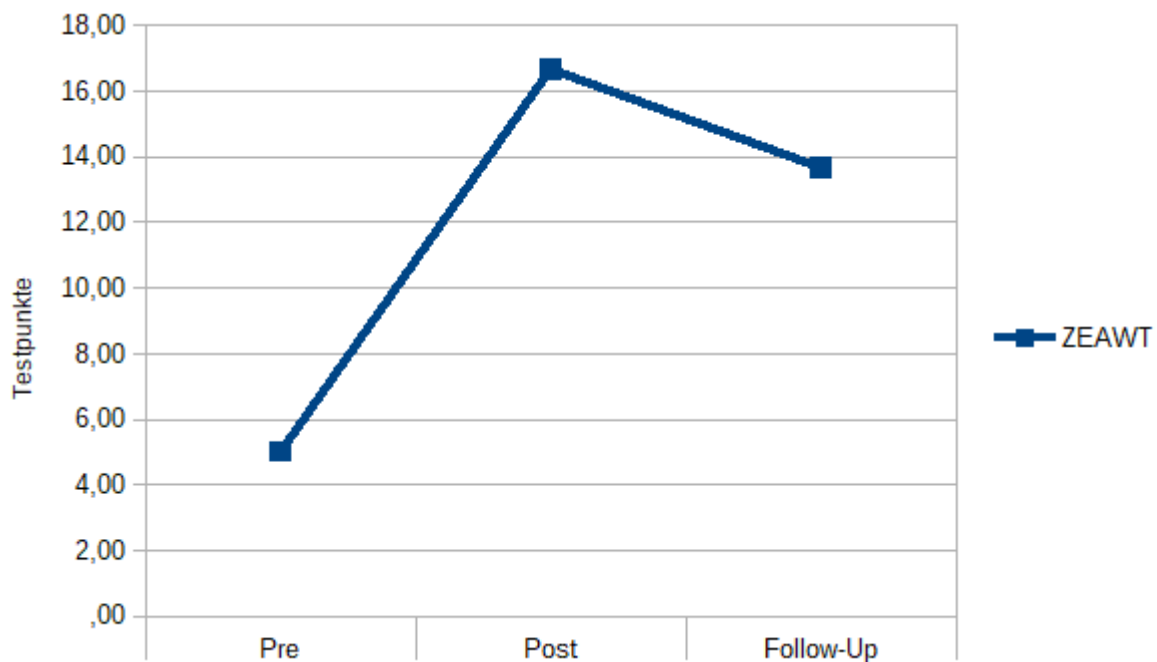


Abbildung 51: Ergebnisse im Test zum zusätzlich erworbenen Allgemeinwortschatz

Für beide Tests bestätigt die einfaktorielle ANOVA jeweils, dass für den Faktor Zeit ein statistisch signifikanter Effekt vorliegt (ZEFWT: $F(1,580;97,975) = 502,972$; $p < 0,001$; ZEAWT: $F(1,425;88,339) = 199,207$; $p < 0,001$)¹¹². Die anschließend durchgeführten paarweisen Vergleiche zeigen außerdem, dass die Unterschiede zwischen den Pre- und den Posttestmittelwerten, zwischen den Post- und Follow-Up-Testwerten und zwischen den Pre- und den Follow-Up-Testwerten in beiden Tests statistisch hoch signifikant sind ($p < 0,001$). Damit haben sowohl der Modulunterricht (Zugewinn1) als auch die anschließenden sechs Wochen (Vergessen) zu einer statistisch hochsignifikanten Zu- bzw. Abnahme des englischen Fach- und Allgemeinwortschatzes geführt. Um zu überprüfen, in welchem Umfang englische Fach- bzw. Allgemeinsprache erworben wurde, wird zunächst der insgesamt erworbene Fachwortschatz ermittelt. Da es sich bei dem Test zum zusätzlich erworbenen Fachwortschatz (ZEFWT) um einen Stichprobentest handelte, bei dem von den insgesamt 45 zusätzlich zu erwerbenden Fachwörtern nur 25 im Test auftauchten, werden die Testergebnisse zuerst auf 45 Wörter hochgerechnet (ZEFW45). Anschließend werden dazu die Ergebnisse des Tests zur englischen Fachsprache (EFTW) addiert. Damit ergibt sich der im Mittel insgesamt beherrschte englische Fachwortschatz (EFTW65) zu den drei Testzeitpunkten.

Zeit	Test	M
Pre	ZEFWT	6,46
	ZEFW45	11,63
	EFTW	4,81
	EFTW65	16,44
Post	ZEFWT	21,05
	ZEFW45	37,89
	EFTW	13,49
	EFTW65	51,38
Follow-Up	ZEFWT	18,73
	ZEFW45	33,71
	EFTW	13,29
	EFTW65	47,00

Tabelle 50: Ergebnisse in Bezug auf den insgesamt zu erwerbenden englischen Fachwortschatz

¹¹² Zwar erwiesen sich die Daten nach dem Kolmogorov-Smirnov-Test als nicht normalverteilt, die Werte für Schiefe und Kurtosis wichen jedoch nur geringfügig von null ab. Die Varianzanalyse ist nach Bühner & Ziegler (2009) relativ robust gegenüber einer Verletzung der Normalverteilungsannahme, wenn die Stichproben, wie bei der vorliegenden Untersuchung der Fall, etwa gleich groß sind. Da der Mauchly-Tests für den DFWT keine Sphärizität erkennen ließ ($p < 0,001$), wurde eine Korrektur nach Greenhouse-Geisser durchgeführt.

Auch beim Test zum zusätzlich erworbenen Allgemeinwortschatz handelt es sich um einen Stichprobentest, bei dem 23 der insgesamt 40 Begriffe der Allgemeinsprache im Test auftauchten. Auch hier werden die Ergebnisse zu den drei Testzeitpunkten auf die Gesamtwortzahl hochgerechnet (ZEA40).

Zeit	Test	M
Pre	ZEA40	5,03
	ZEA40	8,75
Post	ZEA40	16,68
	ZEA40	29,01
Follow-Up	ZEA40	13,67
	ZEA40	23,77

Tabelle 51: Ergebnisse in Bezug auf den insgesamt zu erwerbenden Allgemeinwortschatz

Berechnet man anschließend wieder die Differenzen aus Pre- und Posttestergebnis (Zugewinn 1), aus Follow-Up- und Posttestergebnis (Vergessen) sowie aus Follow-Up- und Pretestergebnis (Zugewinn 2), so geben die Mittelwerte (M) an, wie viele Wörter des Allgemein- bzw. des Fachwortschatzes insgesamt im Mittel dazugelernt bzw. vergessen wurden.

Zeit	Wortschatz	M	M _{max}	M%
Zugewinn 1	EFW65	34,94	48,56	71,95
	EAW40	20,26	31,25	64,83
Vergessen	EFW65	-4,38	51,38	-8,52
	EAW40	-5,23	29,01	-18,03
Zugewinn 2	EFW65	30,57	48,56	62,95
	EAW40	15,03	31,25	48,09

Tabelle 52: Zugewinn und Vergessen in Bezug auf den insgesamt zu erwerbenden Wortschatz

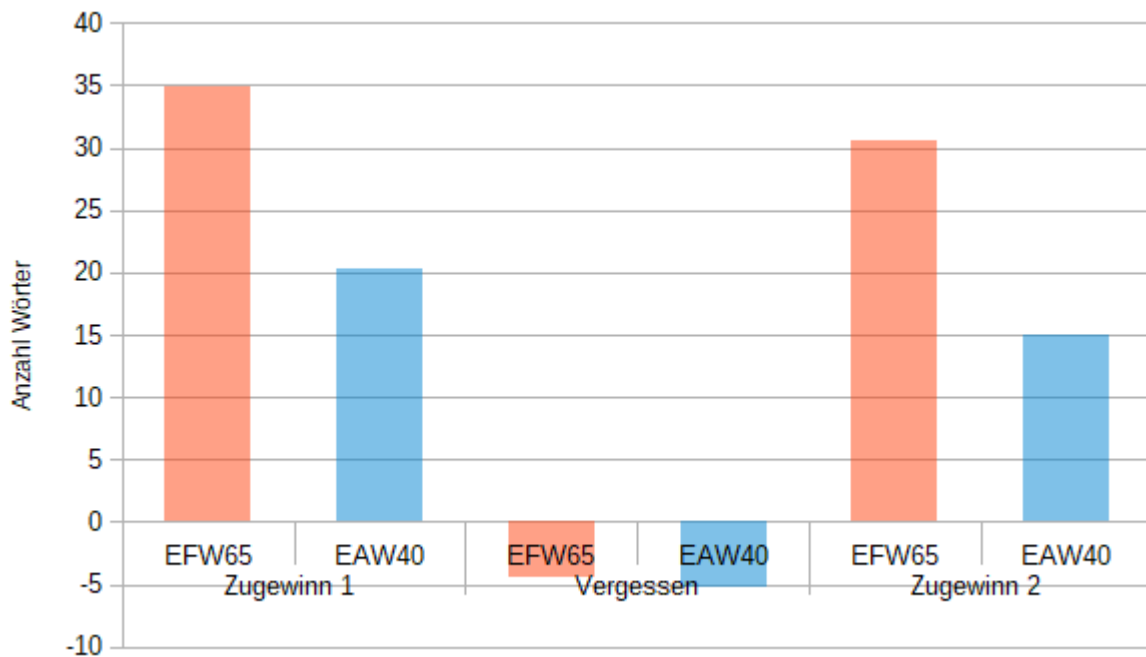


Abbildung 52: Zugewinn und Vergessen beim insgesamt zu erwerbenden Wortschatz

Es zeigt sich, dass der mittelfristige Zugewinn (Zugewinn 2) im englischen Fachwortschatz im Schnitt 30,57 Wörter und im Allgemeinwortschatz im Schnitt 15,03 Wörter beträgt. Berechnet man aufgrund der gegebenen Testergebnisse die maximal möglichen Zugewinne bzw. das maximal mögliche Vergessen ($M_{\max}$), so lassen sich prozentuale Werte für Zugewinn und Vergessen ermitteln ($M\%$). Es zeigt sich, dass der unmittelbare Zugewinn (Zugewinn 1) beim Fachwortschatz etwa 72 % und beim Allgemeinwortschatz etwa 65 % beträgt. Dabei werden Fachbegriffe offenbar nicht nur besser gelernt, sondern auch deutlich seltener vergessen als Begriffe des Allgemeinwortschatzes. Während nur rund 8,5 % der Fachbegriffe vergessen werden, sind es beim Allgemeinwortschatz etwa 18 %. Der mittelfristige Zugewinn (Zugewinn 2) liegt deshalb für den Fachwortschatz bei 63 % und beim Allgemeinwortschatz bei 48 %.

7.5 Motivation

In den folgenden Abschnitten stehen die Auswirkungen des bilingualen Unterrichts auf die Motivation im Mittelpunkt. Dazu werden zunächst die Ergebnisse der CLIL-Gruppe mit denen der nicht-CLIL Gruppe verglichen. Anschließend werden die Daten nach Interessensgruppen und Geschlecht differenziert analysiert. Es gilt, die beiden Hypothesen zur Motivation von fremdsprachlich orientierten und naturwissenschaftlich orientierten Schülern zu überprüfen. In diesem Zusammenhang werden auch andere Prädiktoren untersucht, die einen Einfluss auf die Motivation der Schüler im bilingualen Unterricht haben. Zuletzt werden die Einschätzungen der CLIL-Schüler vor und nach Durchführung des Moduls miteinander verglichen.

7.5.1 Vergleich von Untersuchungs- und Kontrollgruppe

Die Reliabilitätsanalyse zur Skala intrinsische Motivation lieferte für alle vier Subskalen Cronbachs Alpha-Werte um 0,7 (IMO: $\alpha = 0,643$; NMO : $\alpha = 0,727$; KMO : $\alpha = 0,735$; DMO : $\alpha = 0,750$). Damit liegen die Werte für Nutzen, Kompetenzerleben und Druck im akzeptablen Bereich, der Wert für das Interesse knapp darunter¹¹³. Auch die Trennschärfen aller verwendeter Items lagen oberhalb von 0,3 und damit im akzeptablen Bereich.

Gruppe	N	Interesse M (SD; SE)	Nutzen M (SD; SE)	Kompetenz- erleben M (SD; SE)	Druck M (SD; SE)
CLIL	70	3,15 (0,78;0,09)	3,03 (0,71;0,09)	3,20 (0,89;0,11)	2,94 (0,87;0,10)
nicht-CLIL	71	2,58 (0,70;0,08)	2,52 (0,72;0,09)	3,43 (0,58;0,07)	3,38 (0,54;0,06)
Gesamt	141	2,87 (0,79;0,07)	2,78 (0,76;0,06)	3,32 (0,75;0,06)	3,17 (0,76;0,06)

Tabelle 53: Intrinsische Motivation in der CLIL- und in der nicht-CLIL Gruppe

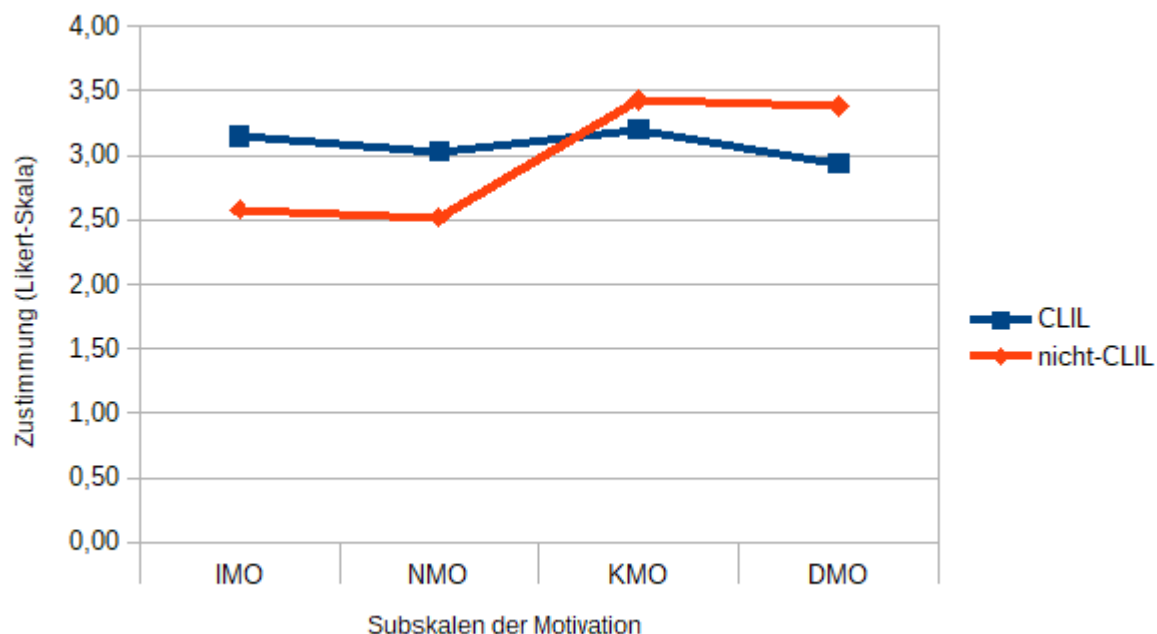


Abbildung 53: Intrinsische Motivation in der CLIL- und in der nicht-CLIL Gruppe

¹¹³ Die Cronbachs Alpha Werte der Skalen KMO und DMO lagen zunächst bei $\alpha = 0,634$ (KMO) und $\alpha = 0,693$ (DMO). Durch Streichung der Items KMO4 und DMO1 konnte der Wert jedoch verbessert werden. Für die Skala IMO war eine Verbesserung des Wertes durch Streichung eines Items nicht möglich.

Vergleicht man die Gruppenmittelwerte zu den vier Skalen der intrinsischen Motivation nach Deci und Ryan mittels t-Test¹¹⁴, so stellt man fest, dass das Interesse am Modul sowie die Einschätzung des Nutzens bei den bilingual unterrichteten Schülern signifikant größer sind als bei den monolingual unterrichteten Schüler (IMO: $t(139) = 4,622$; $p < 0,001$; NMO: $t(138) = 4,245$; $p < 0,001$)¹¹⁵. Die Berechnung der Effektgrößen mittels Cohens d zeigt zudem, dass in beiden Fällen ein starker Effekt vorliegt (IMO: $d = 0,73$; NMO: $d = 0,72$). Auf der anderen Seite zeigt sich, dass aber auch der subjektiv empfundene Druck in der CLIL-Gruppe signifikant größer ist als in der nicht-CLIL-Gruppe (DMO: $U=1783,500$; $z = -2,936$; $p = 0,003$)¹¹⁶. Auch im Kompetenzerleben schneiden die CLIL-Schüler signifikant schlechter ab als die Regelschüler, wobei auch hier die Mittelwertsunterschiede statistisch signifikant sind ($U = 2328,500$; $z = -0,661$; $p = 0,005$)¹¹⁷. Da keine Pretest-Daten zur intrinsischen Motivation im Fach Chemie vorliegen, soll zusätzlich überprüft werden, inwieweit die vier hier betrachteten Skalen mit weiteren chemiebezogenen Variablen korrelieren, und zwar insbesondere mit solchen, bei denen es zwischen CLIL- und nicht-CLIL-Gruppe signifikante Unterschiede gibt. In Abschnitt 7.1 wurde gezeigt, dass die CLIL-Schüler ein signifikant größeres Interesse am Fach Chemie sowie ein signifikant positiveres Selbstkonzept in diesem Fach haben. Es ist denkbar, dass auch die höhere Motivation im Modul auf eben diese Variablen und weniger auf die Verwendung der Fremdsprache zurückzuführen ist. Die Berechnung der Korrelationskoeffizienten Kendall-Tau für die Untersuchungs- und die Kontrollgruppe zeigt jedoch, dass nur das Fachinteresse signifikant mit dem Interesse am Modul korreliert. Interessanterweise ist die Korrelation jedoch negativ. Das ausgeprägtere Fachinteresse der CLIL-Schüler im Vergleich zu den nicht-CLIL-Schülern wirkt sich also negativ auf das Interesse am Modul aus. Dennoch ist, wie gesehen, das Interesse am Modul bei den CLIL-Schülern insgesamt signifikant höher.

114 Für die Sub-Skala Druck wurde dazu wegen der Ungleichheit der Varianzen ein Welch-Test durchgeführt.

115 Der Levene-Test ergab, dass in beiden Fällen Varianzhomogenität vorliegt (IMO: $F = 3,511$, $p = 0,063$; NMO: $F = 0,449$; $p = 0,504$). Der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend, wurde trotz Verletzung der Normalverteilungsannahme ein t-Test durchgeführt.

116 Die Codierung der Skala Druck wurde so angelegt, dass ein hoher Wert für DMO bedeutet, dass der Druck als gering empfunden wird. Da der Levene-Test ergab, dass keine Varianzhomogenität vorliegt (DMO: $F = 16,817$, $p < 0,001$), wurde der Mann-Whitney-U-Test durchgeführt.

117 Da der Levene-Test ergab, dass keine Varianzhomogenität vorliegt (KMO: $F = 20,487$, $p < 0,001$), wurde der Mann-Whitney-U-Test durchgeführt.

	FICH	SKCH
IMO	-0,129 (0,046)*	-0,47 (0,472)
NMO	-0,114 (0,076)	-0,061 (0,350)
KMO	-0,094 (0,157)	-0,109 (0,108)
DMO	-0,060 (0,362)	-0,023 (0,735)
* $p < 0,05$		

Tabelle 54: Korrelationen zur intrinsischen Motivation¹¹⁸

7.5.2 Vergleich der Interessensgruppen

Differenziert man innerhalb der CLIL- und der nicht-CLIL Gruppe nach den unterschiedlichen Interessensgruppen, ist erkennbar, ob sich der bilinguale Chemieunterricht für Schüler mit bestimmten Interessenslagen besser oder weniger gut eignet.

Die beiden Hypothesen hierzu lauten:

1. Die Werte für die Subskalen der intrinsischen Motivation (Interessen, Kompetenzerleben, Druck und Nutzen) sind bei den fremdsprachlich orientierten Schülern im bilingualen Unterricht höher als im Regelunterricht.
2. Die Werte für die Subskalen der intrinsischen Motivation (Interessen, Kompetenzerleben, Druck und Nutzen) sind bei den naturwissenschaftlich orientierten Schülern im bilingualen Unterricht geringer als im Regelunterricht.

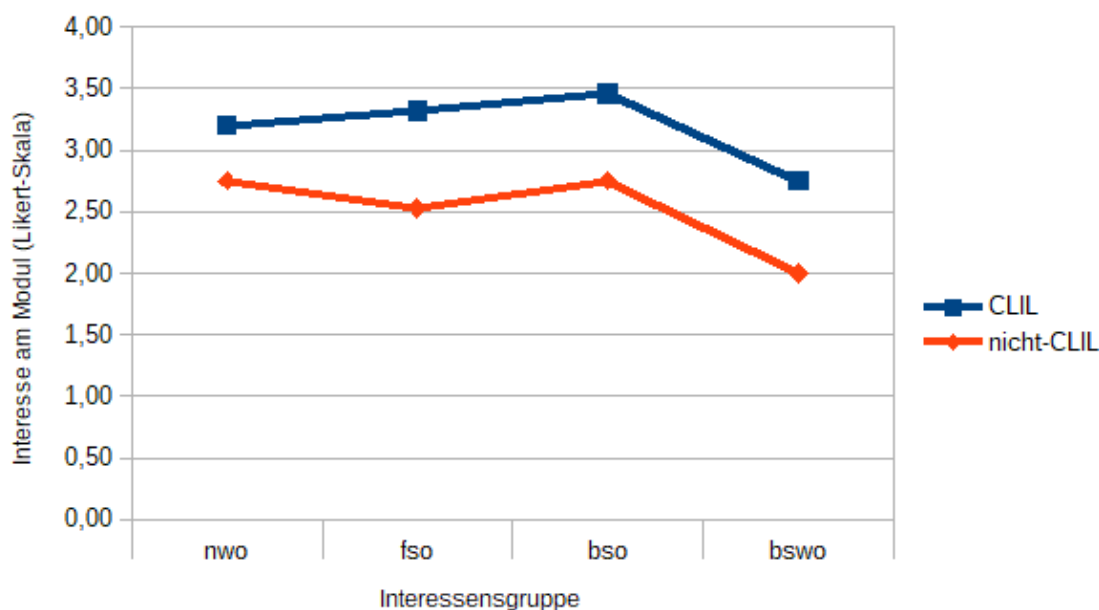
Es wird also davon ausgegangen, dass aufgrund der als besonders interessant empfundenen Fremdsprache sich fso-Schüler stärker für den bilingualen Unterricht motivieren lassen als für den Regelunterricht. Auf der anderen Seite steht die Befürchtung, dass nwo-Schüler aufgrund ihres mangelnden Interesses an der Fremdsprache und ihres gleichzeitig hohen Interesses am Fach Chemie durch die Verwendung der Fremdsprache demotiviert werden können. Der Vergleich der Mittelwerte zeigt, dass das Interesse am Modul in den CLIL-Gruppen stets höher ist als in den nicht-CLIL-Gruppen, und zwar unabhängig von der fachlichen Orientierung der Schüler.

¹¹⁸ Es wurde der Korrelationskoeffizient Kendall-Tau (b und c) berechnet, da die Inspektion der Streudiagramme in keinem der acht Fälle einen linearen Zusammenhang erkennen ließ (Bühner & Ziegler, 2009, S. 609ff.). Auf Grund der hohen Anzahl der Bindungen wurde Kendall-Tau dem Spearman Rangkorrelationskoeffizienten vorgezogen.

	Mittelwerte für das Interesse am Modul				
	nwo (SD;N)	fso (SD;N)	bso (SD;N)	bswo (SD;N)	kA (SD;N)
CLIL	3,20 (0,64;31)	3,32 (0,75;22)	3,46 (0,59;15)	2,75 (-;1)	3,75 (0,35;2)
nicht-CLIL	2,75 (0,75;14)	2,53 (0,65;28)	2,75 (0,73;16)	2,00 (0;2)	2,56 (1,09;4)

Tabelle 55: Interesse am Modul in den vier Interessensgruppen

Dabei ist der Unterschied bei den fso-Schülern am größten ($\Delta M = 0,78$) und bei den nwo-Schülern am kleinsten ($\Delta M = 0,45$), während die Differenz bei den bso-Schülern zwischen diesen beiden Werten liegt ($\Delta M = 0,71$)¹¹⁹.

**Abbildung 54: Interesse am Modul in den vier Interessensgruppen**

Es zeigt sich, dass die CLIL-Schüler im Mittel den Aussagen zum positiven Interesse am Modul eher zustimmen, während das bei den nicht-CLIL Schülern nicht der Fall ist¹²⁰. Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Orientierung (Faktorenstufen: fso, nwo, bso und bswo) ergibt für den Faktor Gruppe, wie bereits in Abschnitt 7.5.1 gesehen, einen hochsignifikanten Unterschied ($F(1,134) = 12,330$; $p < 0,001$). Sowohl für den Faktor Interessensgruppe ($F(4,134) = 0,923$; $p = 0,453$) als auch für die Interaktion der beiden

¹¹⁹ Die Werte für die bswo-Schüler werden im Folgenden nicht ausgewertet, da die entsprechenden Teilgruppen mit $N = 1$ bzw. $N = 2$ zu klein sind.

¹²⁰ Der Likert-Wert 3 bedeutet „stimmt eher“.

Faktoren können jedoch keine signifikanten Unterschiede festgestellt werden ($F(4,134) = 0,528$; $p = 0,716$)¹²¹. Auch bei der Einschätzung des Nutzens des Moduls erreichen die CLIL-Schüler über alle Interessensgruppen hinweg die höchsten Werte. Interessanterweise schätzen die nwo-Schüler den Nutzen eines in der Fremdsprache geführten Unterrichts etwas höher ein als die fso-Schüler. Die höchsten Werte erlangen hier jedoch die bso-Schüler. Sie sehen in der Kombination von Fremdsprache und Chemieunterricht den größten Nutzen ($\Delta M = 0,93$). Mit Werten von etwa 3 für die Skala NMO „stimmen“ die CLIL-Schüler damit den Aussagen zum persönlichen Nutzen des Moduls im Mittel „eher zu“, während in der nicht-CLIL Gruppe erkennbar mehr Schüler meinen, dass dies „eher nicht stimmt“.

	Mittelwerte für den Nutzen des Moduls				
	nwo (SD;N)	fso (SD;N)	bso (SD;N)	bswo (SD;N)	kA (SD;N)
CLIL	3,12 (0,63;31)	2,94 (0,72;22)	3,6 (0,60;15)	2,00 (-;1)	3,5 (0,71;2)
nicht-CLIL	2,42 (0,78;14)	2,57 (0,59;28)	2,67 (0,81;16)	1,83 (0,24;2)	2,75 (1,26;4)

Tabelle 56: Nutzen des Moduls in den vier Interessensgruppen

¹²¹ Während die Bedingung der Varianzhomogenität nach dem Levene-Test ($F = 0,978$; $p = 0,456$) erfüllt ist, zeigte der Kolmogorov-Smirnov-Test, dass die Daten nicht normalverteilt waren ($Z(135) = 0,095$; $p = 0,004$). Es wurde jedoch wieder, den Empfehlungen Bühners und Zieglers (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

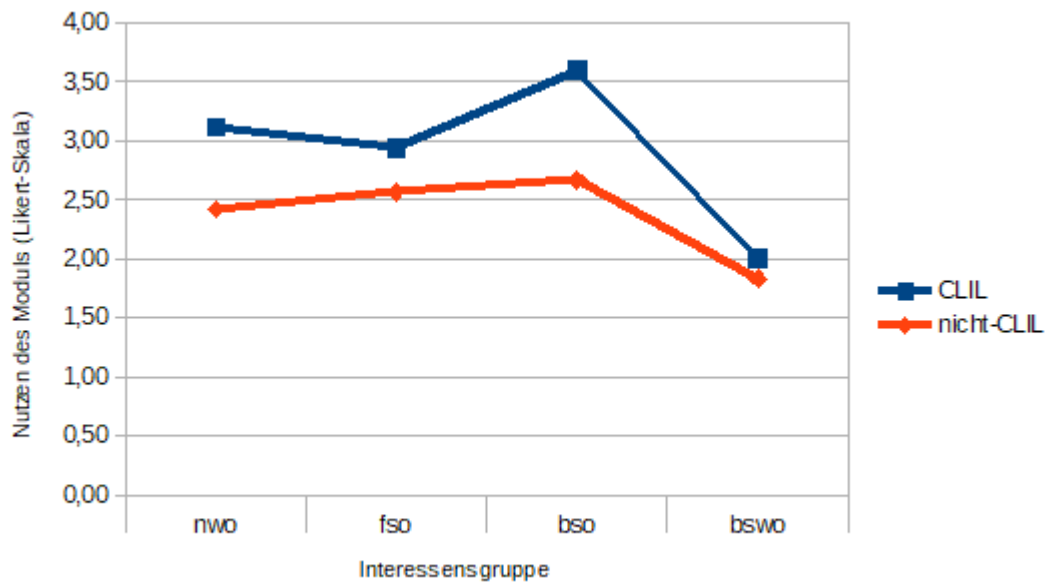


Abbildung 55: Nutzen des Moduls in den vier Interessensgruppen

Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Orientierung (Faktorenstufen: fso, nwo, bso und bswo) ergibt für den Faktor Gruppe, wie bereits in Abschnitt 7.5.1 gesehen, einen signifikanten Unterschied ($F(1,134) = 6,900$; $p = 0,01$). Auch für den Faktor Interessensgruppe ($F(4,134) = 3,032$; $p = 0,02$) - nicht jedoch für die Interaktion der beiden Faktoren - kann ein signifikanter Unterschied festgestellt werden ($F(4,134) = 0,915$; $p = 0,458$)¹²². Die wegen der Ungleichheit in den Gruppengrößen durchgeführten Post-hoc-Tests nach Gabriel ($p = 0,12$) und Hochberg ($p = 0,40$) zeigen, dass ein signifikanter Unterschied nur zwischen der sehr kleinen und deswegen von der Untersuchung ausgeschlossenen bswo-Gruppe und der bso-Gruppe zu finden ist, nicht aber zwischen den anderen Interessensgruppen¹²³. Beim Kompetenzerleben erkennt man, dass sich die nicht-CLIL Schüler durchweg kompetenter erleben als die CLIL-Schüler. Erwartungskonform ist der Unterschied bei den nwo-Schülern am größten und bei den fso- und den bso-Schülern deutlich kleiner.

¹²² Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 0,978$; $p = 0,456$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(135) = 0,085$; $p = 0,017$). Es wurde jedoch wieder, den Empfehlungen Bühners & Ziegler (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

¹²³ Die Post-hoc-Tests nach Tukey-HSD und nach Bonferroni liefern ähnliche Ergebnisse.

	Mittelwerte für das Kompetenzerleben während des Moduls				
	nwo (SD;N)	fso (SD;N)	bs0 (SD;N)	bswo (SD;N)	kA (SD;N)
CLIL	2,78 (0,65;31)	2,83 (0,73;21)	2,87 (0,53;15)	2,25 (-;1)	3,04 (0,41;2)
nicht-CLIL	3,39 (0,62;14)	3,21 (0,53;28)	3,30 (0,53;16)	3,13 (0,18;2)	3,31 (0,24;4)

Tabelle 57: Kompetenzerleben im Modul in den vier Interessensgruppen

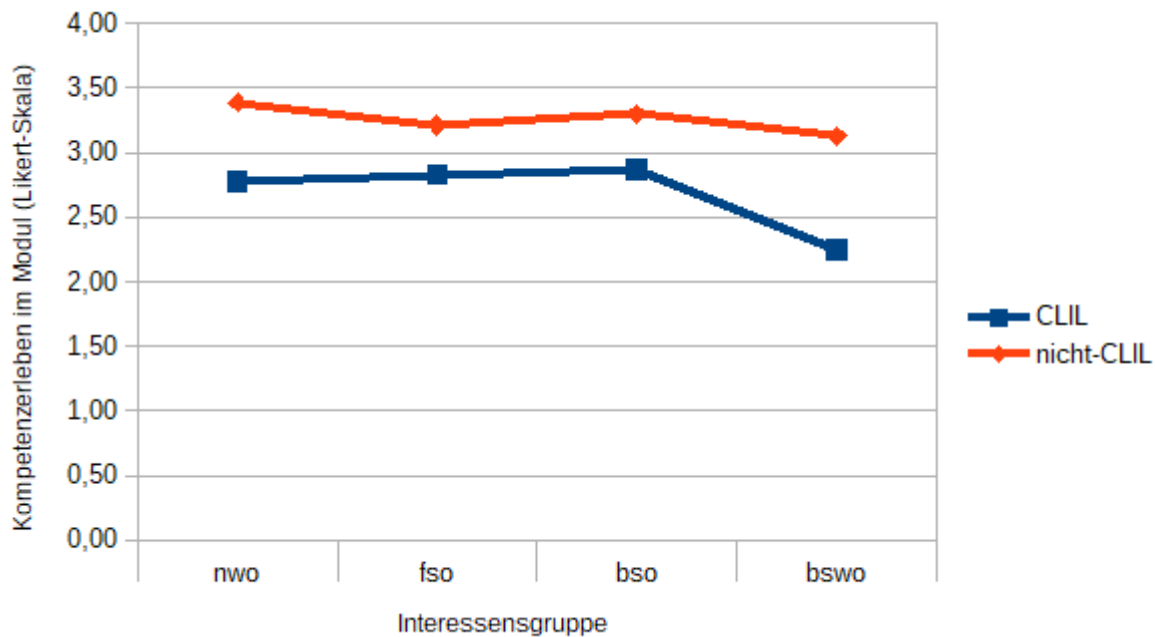


Abbildung 56: Kompetenzerleben im Modul in den vier Interessensgruppen

Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Orientierung (Faktorenstufen: fso, nwo, bso und bswo) ergibt für den Faktor Gruppe, wie bereits in Abschnitt 7.5.1 gesehen, einen signifikanten Unterschied ($F(1,133) = 7,104$; $p = 0,009$). Weder für den Faktor Interessensgruppe ($F(4,133) = 0,394$; $p = 0,812$) noch für die Interaktion der beiden Faktoren kann jedoch ein signifikanter Unterschied festgestellt werden ($F(4,133) = 0,342$; $p = 0,849$)¹²⁴. Die CLIL-Schüler empfinden über alle Interessensgruppen hinweg einen größeren Druck während des Moduls als die nicht-CLIL-Schüler, wobei der Unterschied bei den nwo-Schülern

¹²⁴ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 1,180$; $p = 0,316$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(134) = 0,101$; $p = 0,002$). Es wurde jedoch wieder, den Empfehlungen Bühners & Zieglers (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

wieder am größten und bei den fso- und den bso-Schülern deutlich kleiner ist¹²⁵. Wie bereits beim Kompetenzerleben scheint sich ein gewisses Interesse für die Fremdsprache, wie es fso- und bso-Schüler mitbringen, positiv auf Kompetenzerleben und Druck auszuwirken.

	Mittelwerte für den Druck während des Moduls				
	nwo (SD;N)	fso (SD;N)	bso (SD;N)	bswo (SD;N)	kA (SD;N)
CLIL	2,66 (0,59;31)	2,53 (0,73;22)	2,67 (0,73;15)	2,50 (-;1)	3,13 (0,18;2)
nicht-CLIL	3,43 (0,43;14)	3,18 (0,59;28)	3,09 (0,63;16)	3,00 (0,00;2)	3,50 (0,54;4)

Tabelle 58: Empfundener Druck im Modul in den vier Interessensgruppen

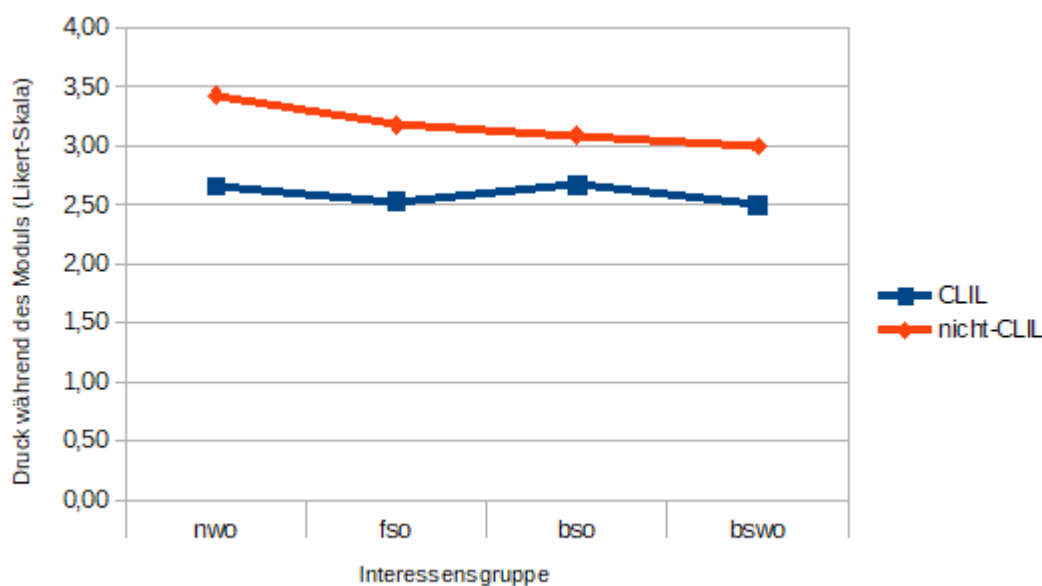


Abbildung 57: Empfundener Druck im Modul in den vier Interessensgruppen

Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Orientierung (Faktorenstufen: fso, nwo, bso und bswo) ergibt für den Faktor Gruppe, wie bereits in Abschnitt 7.5.1 gesehen, einen signifikanten Unterschied ($F(1,134) = 7,517$; $p = 0,007$). Weder für den Faktor Interessensgruppe ($F(4,134) = 1,154$; $p = 0,335$) noch für die Interaktion der beiden Faktoren kann jedoch ein signifikanter Unterschied festgestellt werden ($F(4,134) = 0,393$; $p = 0,813$)¹²⁶. Zusammenfassend ist festzuhalten, dass es in keiner der vier Subskalen der intrinsischen

¹²⁵ Die Codierung der Skala Druck wurde so angelegt, dass ein hoher Wert für DMO bedeutet, dass der Druck als gering empfunden wird.

¹²⁶ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 1,327$; $p = 0,236$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(135) = 0,082$; $p = 0,025$). Es wurde

Motivation statistisch signifikante Unterschiede zwischen den Interessensgruppen gibt, die gefundenen tendenziellen Unterschiede jedoch erwartungskonform sind. So ist der Interessensunterschied zwischen CLIL- und nicht-CLIL Schülern bei den fso-Schülern größer als bei den nwo-Schülern (siehe Seite 218). Gleichzeitig ist der Unterschied im Kompetenzerleben bei den CLIL-Schülern deutlich geringer als bei den nicht-CLIL-Schülern, die sich im Regelunterricht deutlich kompetenter fühlen (siehe Seite 221). Abschließend werden noch einmal die Gruppe der nwo- und der fso-Schüler in den Blick genommen, die Gegenstand der beiden Untersuchungshypothesen zur Motivation sind. Vergleicht man die Motivation der nwo-Schüler, die bilingual unterrichtet wurden, mit der Motivation der nwo-Schüler, die auf Deutsch unterrichtet wurden, so zeigt sich, dass sich die nwo-Schüler im Mittel etwas mehr für den CLIL-Unterricht interessieren und diesen auch nützlicher finden als den deutschsprachigen Unterricht. Auf der anderen Seite fühlen sie sich im deutschsprachigen Unterricht kompetenter und empfinden dort weniger Druck¹²⁷.

	Mittelwerte für die vier Subskalen der Motivation im Modul			
	IMO (SD;N)	NMO (SD;N)	KMO (SD;N)	DMO (SD;N)
CLIL	3,20 (0,64;31)	3,12 (0,63;31)	2,78 (0,65;31)	2,66 (0,59;31)
nicht-CLIL	2,75 (0,75;14)	2,42 (0,78;14)	3,39 (0,62;14)	3,43 (0,43;14)

Tabelle 59: Intrinsische Motivation der nwo-Schüler

Für die Untersuchungshypothese, wonach nwo-Schüler im deutschsprachigen Unterricht stärker motiviert sind als im bilingualen Unterricht, finden sich somit keine eindeutigen Belege. Nur zwei der vier Subskalen verhalten sich erwartungskonform (KMO, DMO), die Unterschiede sind aber - wie gesehen - nicht statistisch signifikant. Vergleicht man die Motivation der fso-Schüler, die bilingual unterrichtet wurden, mit der Motivation der auf Deutsch unterrichteten Schüler, so sind die Ergebnisse mit denen der nwo-Gruppe vergleichbar. Auch die fso-Schüler zeigen ein höheres Interesse für den bilingualen Unterricht und bewerten diesen als nützlicher. Wie die nwo-Schüler

jedoch wieder, den Empfehlungen Bühners & Zieglers (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

¹²⁷ Die Codierung der Skala Druck wurde so angelegt, dass ein hoher Wert für DMO bedeutet, dass der Druck als gering empfunden wird.

empfinden sie sich im bilingualen Unterricht weniger kompetent und spüren dort auch einen größeren Druck¹²⁸.

	Mittelwerte für die vier Subskalen der Motivation im Modul			
	IMO (SD;N)	NMO (SD;N)	KMO (SD;N)	DMO (SD;N)
CLIL	3,32 (0,75;22)	2,94 (0,72;22)	2,83 (0,73;21)	2,53 (0,73;22)
nicht-CLIL	2,53 (0,65;28)	2,57 (0,59;28)	3,21 (0,53;28)	3,18 (0,59;28)

Tabelle 60: Intrinsische Motivation der fso-Schüler

Auch für die zweite Untersuchungshypothese, wonach fso-Schüler im bilingualen Unterricht stärker motiviert sind als im Regelunterricht, finden sich somit keine eindeutigen Belege. Nur zwei der vier Subskalen verhalten sich erwartungskonform (IMO, NMO), die Unterschiede sind jedoch statistisch nicht signifikant.

7.5.3 Vergleich von Jungen und Mädchen

Wie in Abschnitt 4.3.2 dargestellt, zeigen Studien, dass Mädchen sich häufiger für Fremdsprachen und weniger für die Naturwissenschaften interessieren und sich in den mathematisch-naturwissenschaftlichen Fächer zudem als weniger kompetent erleben als Jungen. Inwieweit dies auch für den bilingualen Chemieunterricht zutrifft, soll im Folgenden untersucht werden. Dazu werden vier Subskalen der intrinsischen Motivation nach Deci und Ryan (Interesse, Nutzen, Kompetenzerleben und Druck) im Hinblick auf Geschlechterunterschiede analysiert.

	Mittelwerte für das Interesse am Modul	
	Junge (SD;N)	Mädchen (SD;N)
CLIL	3,10 (0,71;31)	3,45 (0,58;40)
nicht-CLIL	2,71 (0,67;27)	2,63 (0,70;35)

Tabelle 61: Interesse am Modul und Geschlecht

¹²⁸ Die Codierung der Skala Druck wurde so angelegt, dass ein hoher Wert für DMO bedeutet, dass der Druck als gering empfunden wird.

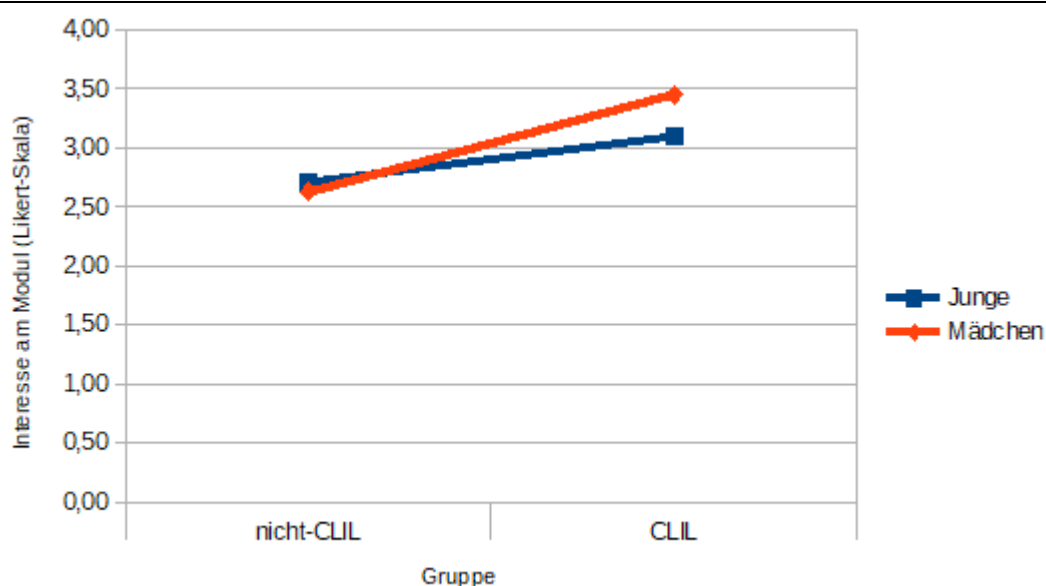


Abbildung 58: Interesse am Modul und Geschlecht

Für das Interesse zeigt sich, dass Mädchen deutlich stärker am bilingualen Unterricht interessiert sind als Jungen ($\Delta M = 0,35$), während Jungen den Regelunterricht etwas interessanter finden als Mädchen ($\Delta M = 0,08$). Insgesamt bewerten dabei sowohl Jungen als auch Mädchen den bilingualen Unterricht als deutlich interessanter als den Regelunterricht, wobei der Unterschied bei den Mädchen deutlich größer ist ($\Delta M = 0,82$). Die Daten belegen damit, dass das Interesse der Mädchen in der hier untersuchten Gruppe vom Wechsel in die Fremdsprache profitiert hat und sie sich stärker für den bilingualen Chemieunterricht interessieren als Jungen. Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Geschlecht (Faktorenstufen: Junge und Mädchen) ergibt für den Faktor Gruppe - wie bereits in Abschnitt 7.5.1 gesehen - einen signifikanten Unterschied ($F(1;132) = 27,272$; $p = 0,001$). Weder für den Faktor Geschlecht ($F(1;132) = 1,232$; $p = 0,269$) noch für die Interaktion der beiden Faktoren kann jedoch ein signifikanter Unterschied festgestellt werden ($F(1;132) = 3,443$; $p = 0,066$)¹²⁹.

¹²⁹ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 1,118$; $p = 0,344$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(133) = 0,100$; $p = 0,002$). Es wurde jedoch wieder, den Empfehlungen Bühners & Zieglers (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

	Mittelwerte für den Nutzen des Moduls	
	Junge (SD;N)	Mädchen (SD;N)
CLIL	3,17 (0,65;31)	3,16 (0,74;31)
nicht-CLIL	2,6 (0,84;27)	2,58 (0,59;35)

Tabelle 62: Nutzen des Moduls und Geschlecht

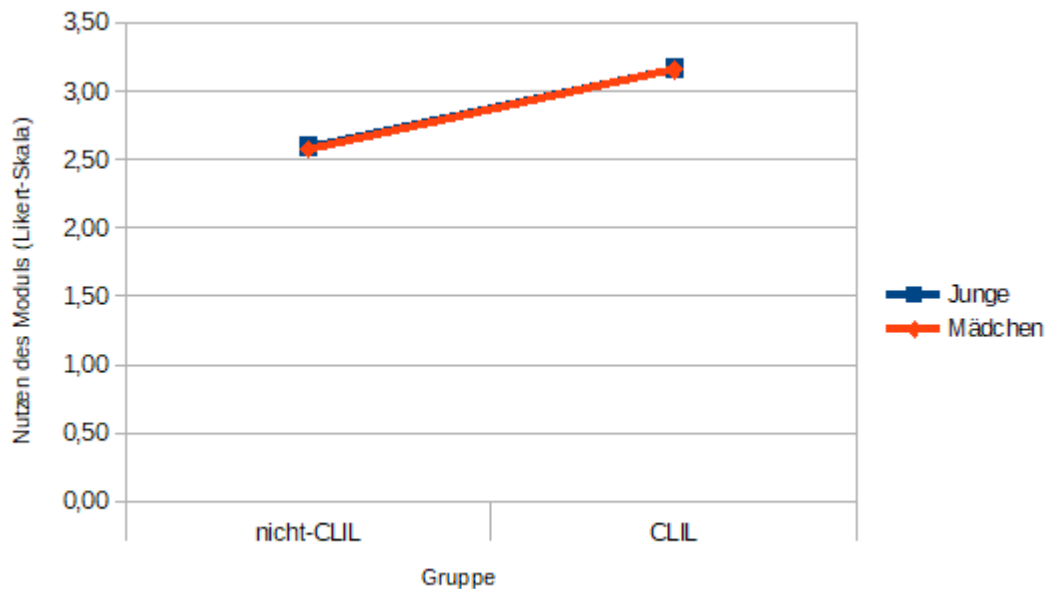


Abbildung 59: Nutzen des Moduls und Geschlecht

Der empfundene Nutzen des Moduls ist in den CLIL-Gruppen deutlich größer als in den nicht-CLIL-Gruppen. Die Geschlechtsunterschiede sind jedoch minimal. Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Geschlecht (Faktorenstufen: Junge und Mädchen) ergibt für den Faktor Gruppe - wie bereits in Abschnitt 7.5.1 gesehen - einen signifikanten Unterschied ($F(1;132) = 27,272$; $p = 0,001$). Weder für den Faktor Geschlecht ($F(1;132) = 1,232$; $p = 0,269$) noch für die Interaktion der beiden Faktoren kann ein signifikanter Unterschied festgestellt werden ($F(1;132) = 3,443$; $p = 0,066$)¹³⁰. Für das Kompetenzerleben finden sich sowohl bei Jungen als auch bei den Mädchen deutlich höhere Werte in der nicht-CLIL-Gruppe. Die Unterschiede zwischen den Geschlechtern sind minimal. Die Abnahme des Kompetenzerlebens

¹³⁰ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 1,897$; $p = 0,133$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(133) = 0,099$; $p = 0,003$). Es wurde jedoch wieder, den Empfehlungen Bühners & Zieglers (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

bei Verwendung der Fremdsprache ist bei den Mädchen etwas weniger ausgeprägt als bei den Jungen.

	Mittelwerte für das Kompetenzerleben im Modul	
	Junge (SD;N)	Mädchen (SD;N)
CLIL	2,81 (0,63;31)	2,82 (0,65;39)
nicht-CLIL	3,34 (0,54;27)	3,29 (0,49;35)

Tabelle 63 Kompetenzerleben im Modul und Geschlecht

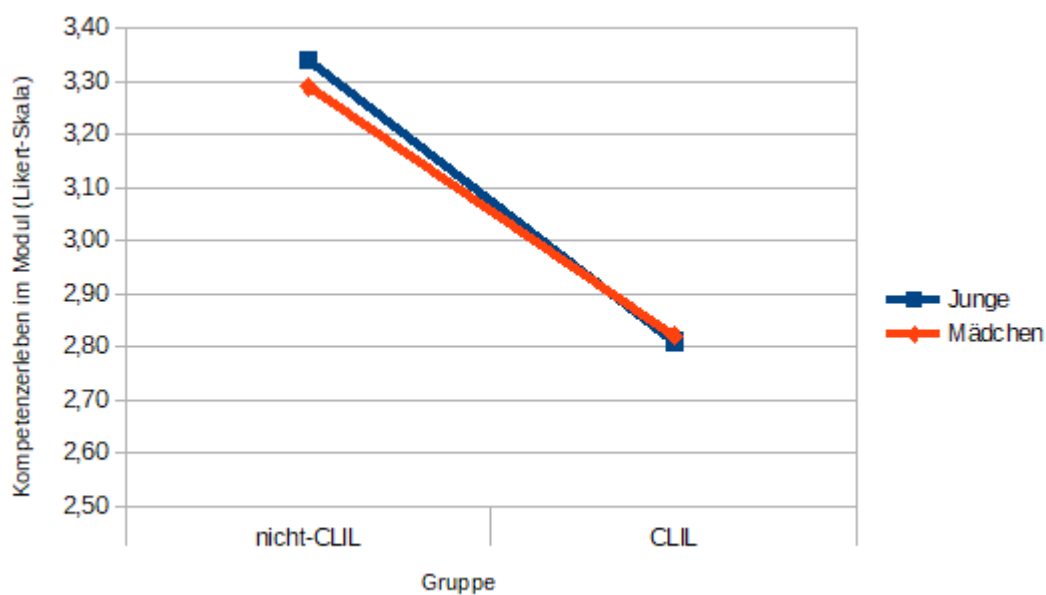


Abbildung 60: Kompetenzerleben im Modul und Geschlecht

Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorenstufen: CLIL und nicht-CLIL) und Geschlecht (Faktorenstufen: Junge und Mädchen) ergibt für den Faktor Gruppe - wie bereits in Abschnitt 7.5.1 gesehen - einen signifikanten Unterschied ($F(1;131) = 23,648$; $p = 0,001$). Weder für den Faktor Geschlecht ($F(1;131) = 0,060$; $p = 0,807$) noch für die Interaktion der beiden Faktoren kann ein signifikanter Unterschied festgestellt werden ($F(1;131) = 0,095$; $p = 0,758$)¹³¹.

¹³¹ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F=1,002$; $p = 0,394$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(132) = 0,167$; $p < 0,001$). Es wurde jedoch wieder, den Empfehlungen Bühners & Ziegler (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

Den Druck im Modul empfinden die CLIL-Schüler wieder als deutlich höher als die nicht-CLIL-Schüler¹³². Die Jungen verspüren jedoch einen etwas geringeren Druck im Regelunterricht als die Mädchen. Die Zunahme des Drucks bei Verwendung der Fremdsprache ist bei den Mädchen etwas geringer als bei den Jungen.

	Mittelwerte für den Druck im Modul	
	Junge (SD;N)	Mädchen (SD;N)
CLIL	2,61 (0,65;31)	2,65 (0,67;40)
nicht-CLIL	3,42 (0,53;27)	3,11 (0,55;35)

Tabelle 64: Empfundener Druck im Modul und Geschlecht

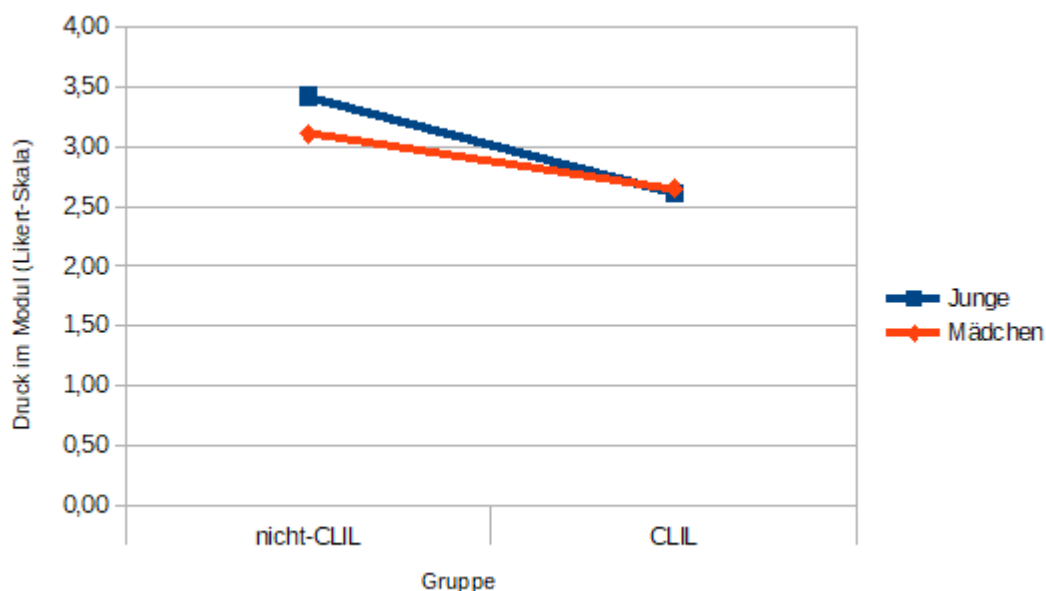


Abbildung 61: Empfundener Druck im Modul und Geschlecht

Die zweifaktorielle ANOVA mit den Faktoren Gruppe (Faktorstufen: CLIL und nicht-CLIL) und Geschlecht (Faktorstufen: Junge und Mädchen) ergibt für den Faktor Gruppe - wie bereits in Abschnitt 7.5.1 gesehen - einen signifikanten Unterschied ($F(1;132) = 20,518$; $p = 0,001$). Weder für den Faktor Geschlecht ($F(1;132) = 0,073$; $p = 0,787$) noch für die Interaktion der beiden Faktoren kann ein signifikanter Unterschied festgestellt werden ($F(1;132) = 0,020$; $p = 0,887$)¹³³.

¹³² Die Codierung der Skala Druck wurde so angelegt, dass ein hoher Wert für DMO bedeutet, dass der Druck als gering empfunden wird.

¹³³ Die Homogenität der Varianzen wurde mithilfe des Levene-Tests überprüft ($F = 1,118$; $p = 0,344$). Der Kolmogorov-Smirnov-Test zeigte, dass die Daten nicht normalverteilt waren ($Z(133) = 0,093$; $p = 0,007$). Es wurde

7.5.4 Prädiktoren für die Motivation im bilingualen Unterricht

Die vorangegangenen Abschnitte belegen, dass die CLIL-Schüler durch den bilingualen Unterricht nicht demotiviert werden, sondern in einzelnen Subskalen der intrinsischen Motivation im Vergleich zu den Regelschülern bessere oder zumindest vergleichbar gute Werte erzielen. Im Folgenden soll untersucht werden, welche personalen Faktoren sich dabei positiv auf die Motivation im bilingualen Unterricht auswirken. Die Daten können als Grundlage dazu dienen, interessierten Schülern Empfehlungen über die Teilnahme am bilingualen Chemieunterricht zu erteilen¹³⁴.

	IBU	NBU	KBU	DBU
FICH	0,120	-0,005	-0,100	-0,092
FIEN	0,226*	0,205*	0,169	0,193*
SKCH	-0,065	-0,010	-0,065	-0,193*
SKEN	0,085	0,088	0,018	0,024
KFT	-0,050	-0,008	-0,052	-0,077
CT	0,106	0,127	0,234**	0,134
ACLT	-0,057	-0,008	0,035	-0,018
** $p < 0,01$; * $p < 0,05$				

Tabelle 65: Korrelationen der Subskalen der intrinsischen Motivation¹³⁵

Die Analyse der Korrelationen zeigt, dass in Bezug auf das Fach Englisch das Fachinteresse signifikant mit dem Interesse am Modul, dem empfundenen Nutzen und dem Druck korrelieren. Dabei ist zu beachten, dass hohe Werte für DBU einen geringeren Druck bedeuten. Das heißt, dass solche Schüler, die ein größeres Interesse am Fach Englisch haben, das Modul nicht nur interessanter und nützlicher fanden als andere, sondern auch weniger Druck verspürten. Beim Kompetenzerleben ist keine statistisch signifikante Korrelation mit dem Fachinteresse festzustellen, wohl aber eine statistisch hochsignifikante Korrelation mit dem Ergebnis im C-Test. Es zeigt sich, dass Schüler, die im Fach Englisch leistungsstark sind, sich auch im bilingualen Unterricht

jedoch wieder, den Empfehlungen Bühners & Zieglers (2009, S. 372) folgend, eine zweifaktorielle ANOVA durchgeführt.

¹³⁴ Die Werte für FICH und SKCH weichen von denen in Abschnitt 7.5.1 ermittelten Werten ab, da letztere sich sowohl auf die CLIL- als auf auch die nicht-CLIL-Gruppe beziehen. Die hier ermittelten Daten beziehen sich hingegen nur auf die Schüler der CLIL-Gruppe.

¹³⁵ Da die Inspektion der Streudiagramme keinen linearen Zusammenhang erkennen ließ, wurde auf einen nonparametrischen Test ausgewichen und der Korrelationskoeffizient nach Kendall-Tau berechnet (Bühner & Ziegler, 2009, S. 611). Die Korrelationen nach Spearman lieferten ähnliche Ergebnisse.

kompetenter fühlen als weniger leistungsstarke Schüler. Für die chemiebezogenen Variablen (FICH, SKCH und ACLT) lassen sich - bis auf eine Ausnahme - zumindest *keine* statistisch signifikanten *negativen* Korrelationen finden. Ein großes Fachinteresse und eine größere Leistungsfähigkeit führen demnach nicht zu einer geringeren Motivation für den bilingualen Unterricht, wie für nwo-Schüler befürchtet (siehe Hypothese 7 in Abschnitt 5.3.3.1)¹³⁶. Interessant ist jedoch, dass Schüler mit einem ausgeprägten chemischen Selbstkonzept im bilingualen Unterricht statistisch signifikant mehr Druck verspüren als Schüler mit einem weniger ausgeprägten Selbstkonzept. Vermutlich merken sie in besonderem Maße, dass die Verwendung der Fremdsprache es ihnen nicht erlaubt, den Stoff gleichermaßen kompetent zu verarbeiten wie im Regelunterricht¹³⁷. Das kann daran liegen, dass Schüler, die sich im Fach Chemie für leistungsstark halten, höhere Ansprüche an ihre Kompetenzen haben als Schüler mit einem geringen Chemieselbstkonzept. Die Sprachbarriere kann sich für diese Schüler nachteilig auswirken, weil sie beispielsweise nicht im gewohnten Tempo und Umfang fachliche Inhalte erarbeiten können. Schüler mit geringem Chemieselbstkonzept hingegen können von der Verlangsamung des Unterrichts und den zusätzlichen schriftlichen Phasen profitieren, weil sie die Inhalte gründlicher erfassen und verstehen können als im Regelunterricht. Hinweise darauf, dass sich kognitive Fähigkeiten positiv auf die Motivation auswirken, etwa weil kognitiv leistungsstarke Kinder die zusätzliche Herausforderung durch die Verwendung der Fremdsprache anspricht (vgl. Abschnitt 3.2.4), finden sich in den Daten nicht. Abschließend sei darauf hingewiesen, dass sämtliche statistisch signifikanten Korrelationen als schwach eingestuft werden müssen, da die Werte für Kendall-Tau alle um 0,2 bzw. -0,2 liegen.

7.5.5 Entwicklung der Motivation bilingualer Schüler

Die Reliabilitätsanalyse zur Skala intrinsische Motivation vor Beginn des Modulunterrichts lieferte für alle vier Subskalen Cronbachs Alpha Werte um 0,7 (EIBU: $\alpha = 0,844$; ENBU : $\alpha = 0,762$; EKBU : $\alpha = 0,834$; EDBU : $\alpha = 0,649$). Damit liegen die Werte für Interesse, Nutzen und Kompetenzerleben im akzeptablen Bereich, der Wert für den Druck nur knapp darunter. Vergleicht man die Einschätzungen der Schüler zu Interesse, Nutzen, Kompetenzerleben und Druck vor Beginn des bilingualen Unterrichts mit den Angaben, die sie im Anschluss an das Modul machen, so fällt auf, dass die Werte für Interesse und Nutzen etwas niedriger ausfallen, während die Werte

¹³⁶ Die Beobachtung deckt sich mit den Ergebnissen zur Motivation der nwo-Schüler, die den bilingualen Unterricht ebenfalls tendenziell interessanter und nützlicher finden als den Regelunterricht (vgl. Abschnitt 8.4.2).

¹³⁷ Auch diese Beobachtung deckt sich mit den Ergebnissen zur Motivation der nwo-Schüler, die im bilingualen Unterricht mehr Druck verspüren als im Regelunterricht (vgl. Abschnitt 8.4.2).

für Kompetenzerleben und Druck deutlich höher sind. Das bedeutet, dass die Erwartungen der Schüler bezüglich der Frage, wie interessant und nützlich der bilinguale Unterricht ist, nicht ganz erfüllt wurden, während die Befürchtungen der Schüler im bilingualen Unterricht wenig kompetent zu sein und einen hohen Druck zu empfinden, nicht in dem Maße bestätigt wurden wie angenommen¹³⁸.

	M (SD;SE)		M (SD;SE)	ΔM (SD;SE)
EIBU	3,26 (0,66;0,08)	IBU	3,18 (0,74;0,09)	0,08 (0,72;0,09)
ENBU	3,16 (0,71;0,09)	NBU	3,06 (0,70;0,09)	0,10 (0,78;0,10)
EKBU	2,83 (0,64;0,08)	KBU	3,17 (0,75;0,09)	-0,34 (0,74;0,09)
EDBU	2,65 (0,66;0,08)	DBU	2,93 (0,79;0,10)	-0,29 (0,85;0,10)

Tabelle 66: Motivation für den bilingualen Unterricht vor und nach dem Modul

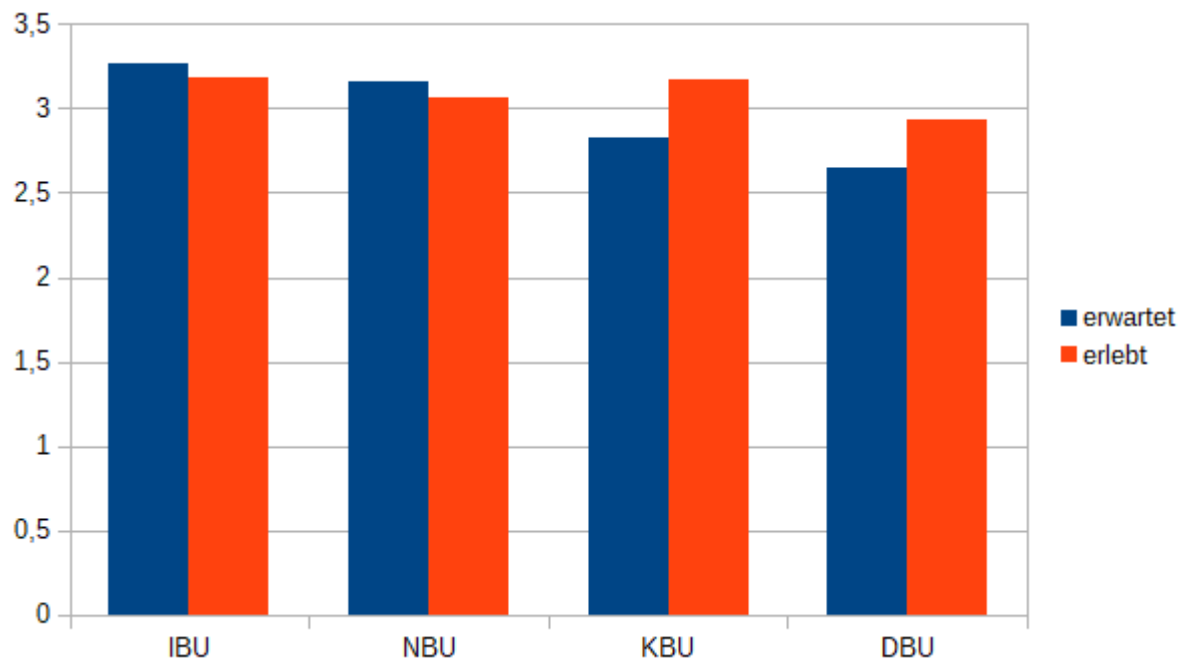


Abbildung 62: Motivation für den bilingualen Unterricht vor und nach dem Modul

Der Vergleich der Mittelwerte mittels t-Test für abhängige Stichproben zeigt, dass die beobachteten Mittelwertsunterschiede in Bezug auf das Interesse und den Nutzen nicht signifikant (IBU: $T(66) = 0,904$; $p = 0,369$; NBU: $T(65) = 1,056$; $p = 0,295$), jedoch die Unterschiede im Kompetenzerleben

¹³⁸ Die Codierung der Skala Druck wurde so vorgenommen, dass „4“ einen „geringen“ und „1“ einen „hohen Druck“ meint.

und im Druck statistisch hochsignifikant sind (KBU: $T(65) = -3,67$; $p = 0,001$; DBU: $T(65) = -2,770$; $p = 0,007$)¹³⁹.

8 Zusammenfassung der Ergebnisse und Fazit

Ziel der vorliegenden Untersuchung ist es, die Wirksamkeit des bilingualen Chemieunterrichts auf sachfachlicher, fremdsprachlicher und motivationaler Ebene zu untersuchen und empirisch abgesicherte Daten für Schüler der Jahrgangsstufe 8 des Gymnasiums ohne Vorerfahrung im bilingualen Unterricht bereitzustellen.

Sachfachlicher Lernerfolg

Hypothese 1

Die auf Deutsch unterrichteten Schüler sind den bilingual unterrichteten Schülern im Umgang mit dem vermittelten chemischen Fachwissen überlegen.

Anhand der Ergebnisse im Chemieleistungstest konnten keine signifikanten Leistungsunterschiede zwischen CLIL- und nicht-CLIL Schülern festgestellt werden. Die bilingual unterrichteten Schüler erzielen vergleichbar gute Ergebnisse wie die Regelschüler, und zwar sowohl unmittelbar im Anschluss an das Modul als auch sechs Wochen später. Sie sind diesen sogar geringfügig überlegen. Der Unterschied bestand jedoch schon vor Beginn des Moduls und verändert sich im Anschluss an das Modul nicht (siehe Abschnitt 7.3.1). Zu ähnlichen Ergebnissen kommen auch Kondring und Ewig (2006), Gregorczyk (2012), Rodenhauser (2016) und Haagen-Schützenhöfer et al. (2011) in ihren Untersuchungen zum bilingualen Biologie-, Physik- bzw. Chemieunterricht. Hartmannsgruber (2014) und Piesche (2016) hingegen sehen zwar ebenfalls signifikante Lernzuwächse bei den bilingual unterrichteten Schülern, in ihren Untersuchungen schneiden die auf Deutsch unterrichteten Schüler jedoch signifikant besser ab. Während die Studien von Hartmannsgruber (2014) und Piesche (2016) mit Realschülern der Klassen 6 und 7 durchgeführt wurden, betreffen die Untersuchungen von Kondring und Ewig (2006), Gregorczyk (2012), Rodenhauser (2016) sowie

¹³⁹ Die Werte waren für NBU, KBU und DBU normalverteilt, für IBU jedoch nicht, wie der Kolmogorov-Smirnov Test zeigt (NBU: $Z(65) = 0,081$; $p = 0,200$; KBU: $Z(65) = 0,096$; $p = 0,200$; DBU: $Z(65) = 0,105$; $p = 0,074$; IBU: $Z(67) = 0,141$; $p = 0,002$). Schiefe und Kurtosis deuteten jedoch auf keine wesentliche Abweichung hin. Deshalb wurde der Empfehlung Bühners & Zieglers (2009, S. 260f.) folgend, der t-Test der nonparametrischen Alternative vorgezogen.

Haagen-Schützenhöfer (2011) Gymnasialschüler der höheren Jahrgänge, d.h. Klasse 9 und älter. Für die Klasse 8 fehlten entsprechende Untersuchungen bislang. Die vorliegende Studie zeigt nun, dass auch diese Schüler den besonderen Herausforderungen des bilingualen Chemieunterrichts gewachsen sind, auch dann, wenn sie keinerlei Vorerfahrungen in dieser Unterrichtsform haben. Dabei wurde in dieser Studie, anders als in den oben zitierten Arbeiten, nicht nur das Fachwissen, sondern erstmals auch der Umgang damit unter Berücksichtigung höherer Denkopoperationen analysiert. Die eingesetzten Leistungstests orientieren sich damit sehr eng an den Anforderungen des schulischen Lernens wie sie durch die Bildungsstandards vorgegeben sind und erfassen Leistungen zu allen drei Anforderungsbereichen. Die Untersuchung liefert somit keine Belege für Hypothese 1, wonach die auf Deutsch unterrichteten Schüler den bilingual unterrichteten Schülern im Umgang mit dem vermittelten chemischen Fachwissen überlegen sind. Wenngleich die sachfachlichen Leistungen der CLIL-Schüler im Mittel denen der Regelschüler entsprechen, ergeben sich mit Blick auf unterschiedliche Schülergruppen deutliche Unterschiede. Dabei schneiden einzelne Schüler im bilingualen Unterricht besser ab, während andere im Regelunterricht bessere Ergebnisse erzielen. Es hat sich gezeigt, dass das Leistungsvermögen in Chemie hochsignifikant mit dem Ergebnis im Leistungstest zum Modul korreliert und, dass leistungsschwache Schüler, also solche, die weniger als ein Drittel der Punkte im ACL-Test erzielen, signifikant bessere Ergebnisse im Regelunterricht erreichen. Umgekehrt erzielen leistungsstarke Schüler im CLIL-Unterricht signifikant bessere Ergebnisse als im Regelunterricht (siehe Abschnitt 7.3.2)¹⁴⁰. Aus der vorliegenden Untersuchung geht somit zum einen hervor, dass sich eine hohe Leistungsfähigkeit in Chemie besonders positiv auf den Lernerfolg im bilingualen Unterricht auswirkt. Möglicherweise sind diese Schüler durch die zusätzlichen Anforderungen, die sich aus der Verwendung der Fremdsprache ergeben, zusätzlich motiviert und arbeiten besonders konzentriert mit. Das beobachtete geringere Kompetenzerleben im Vergleich zum Regelunterricht scheint sich somit für einzelne Schüler leistungsfördernd auszuwirken. Zum anderen zeigt sich, dass eine geringe Leistungsfähigkeit im Fach Chemie sich offenbar besonders negativ auf den Lernerfolg auswirkt, wenn der Unterricht bilingual erteilt wird. Offensichtlich erschwert der Einsatz der Fremdsprache das Verständnis dieser Schüler in besonderem Maße. Jedoch kann ein gutes Englischselbstkonzept dazu beitragen, diesen Leistungsabfall zu verringern, denn neben der Leistungsfähigkeit in Chemie korreliert auch das Englischselbstkonzept statistisch hochsignifikant mit dem Lernerfolg im Modul. Diese Ergebnisse

¹⁴⁰ Die Erkenntnis, dass Vorwissen ein guter Prädiktor für sachfachlichen Lernerfolg in den Naturwissenschaften ist, wird durch zahlreiche Studien bestätigt (Krapp & Seidel, 2014, S. 189).

decken sich mit den Ergebnissen anderer Studien. Auch Piesche (2016, S. 114) findet, dass, sich „im bilingualen Unterricht ein positives Englischselbstkonzept signifikant positiver auf die Testleistung auswirkte als im monolingualen Unterricht.“ Es kommt also, so folgert sie (Piesche 2016, S. 114), weniger auf „die objektiven Englischleistung, sondern eher auf das eigene Zutrauen der Lernenden in Englisch“ an. Wenngleich die durch den CL-Test festgestellten Leistungen der CLIL- und der nicht-CLIL-Schüler vergleichbar sind, so zeigen sich in der Rückbetrachtung des tatsächlich erteilten Unterrichts Differenzen beim Unterrichtsverlauf in den beiden Gruppen. Zwar wurden die Unterrichtsbedingungen in beiden Gruppen gleich gehalten, da alle Klassen vom gleichen Lehrer anhand gleicher Stundenverlaufspläne und Materialien unterrichtet wurden, dennoch zeigte sich, dass das Modul in der CLIL-Gruppe etwas länger, nämlich fast zwölf statt zehn Stunden dauerte. Außerdem ergaben sich im Regelunterricht immer wieder auch Phasen, in denen Sachverhalte ausführlicher diskutiert, durch Beispiele ergänzt oder durch weitere Fragen seitens der Schüler vertieft werden konnten. Auch Wiederholungen kamen im Regelunterricht häufiger vor, offensichtlich weil es Schülern in der Muttersprache leichter fällt, spontan entsprechende Fragen zu formulieren. Es ist zu bedenken, dass sich dadurch langfristig Lernvorteile für die Regelschüler ergeben können.

Hypothese 3

Hohe fremdsprachliche Kompetenzen führen zu besseren Leistungen im sachfachlichen Test. Bei Unterschreitung eines Mindestniveaus an fremdsprachlicher Kompetenz sind die sachfachlichen Leistungen im bilingualen Unterricht schlechter als im Regelunterricht. Überschreiten Schüler jedoch ein oberes Kompetenzniveau in der Fremdsprache, so sind die sachfachlichen Leistungen im bilingualen Unterricht besser als im Regelunterricht (*double threshold hypothesis*).

Aus den ermittelten Daten ergibt sich auch in dieser Studie lediglich ein tendenzieller Einfluss der Englischkompetenzen auf die Leistungen im Chemietest. Es zeigte sich, dass Schüler, die leistungsschwach in Englisch sind dann im Regelunterricht im Mittel bessere Ergebnisse erzielen als im bilingualen Unterricht, wenn sie auch leistungsschwach in Chemie sind. Auf der anderen Seite erzielten solche Schüler, die sowohl in Englisch als auch in Chemie leistungsstark sind, im Mittel im bilingualen Unterricht bessere Ergebnisse als im Regelunterricht (siehe Abschnitt 7.3.2). Die dritte Untersuchungshypothese, wonach hohe fremdsprachliche Kompetenzen zu besseren Leistungen im sachfachlichen Test führen, kann somit nicht bestätigt werden. Für die Annahme, dass bei Unterschreitung eines bestimmten Mindestniveaus an fremdsprachlicher Kompetenz die Leistungen im Regelunterricht besser sind als im bilingualen Unterricht, finden sich zwar Hinweise

in den vorliegenden Daten, die Leistungsunterschiede sind jedoch statistisch nicht signifikant. Das Gleiche gilt für die Annahme, dass Schüler, die ein gewisses oberes Kompetenzniveau in der Fremdsprache überschreiten, bessere Leistungen im bilingualen Unterricht als im Regelunterricht zeigen. Auch hier sind lediglich tendenzielle jedoch keine statistisch signifikanten Hinweise zu konstatieren. Die Beobachtungen von Lo und Lo (2014) für chinesische Immersionsschüler, die erhebliche Leistungsdefizite in den naturwissenschaftlichen Tests feststellten, wenn Schüler über geringe Fremdsprachenkompetenzen verfügten, konnten für die untersuchte Schülergruppe nicht bestätigt werden. Offenbar genügen die fremdsprachlichen Kompetenzen der Achtklässler, um angemessene Leistungen in den Chemietests zu erlangen, wenn der bilinguale Chemieunterricht nach dem vorgestellten CLIL-Konzept gestaltet ist (siehe Abschnitt 3.3). Zum gleichen Ergebnis kommt Hegerfeldt (2006) in ihrer Untersuchung zu einem bilingualen Chemiemodul in Klasse 9 eines Hamburger Gymnasiums, wobei die Schüler allerdings einem bilingualen Zweig entstammten, also bereits über Vorerfahrungen im bilingualen Unterricht verfügten. Einen signifikanten Einfluss der Englischkompetenzen auf das sachfachliche Lernen findet sich vor allem in solchen Studien, in denen - anders als in der vorliegenden Untersuchung - die Sachfachttests in der Fremdsprache sind (vgl. z.B. Zydati, 2007).

Hypothese 2

Die bilingual unterrichteten Schüler erzielen im kognitiv anspruchsvolleren Bereich (*higher order thinking skills*) bessere Leistungen als die auf Deutsch unterrichteten Schüler.

In der Forschung zum bilingualen Unterricht finden sich zahlreiche Hinweise auf die Überlegenheit bilingualer Schüler im Hinblick auf kognitiv anspruchsvollere Leistungen (z.B. Osterhage, 2007; Zydati, 2012; Bialystok & Martin, 2004; Bttger, 2020). Die Studie von Bttger (2020) zum bilingualen Unterricht in Modulform an Grundschulen nimmt dabei jedoch fr sich in Anspruch, dass Selektionseffekte, wie sie unter anderem von Rumlich (2016) nachgewiesen werden konnten, vernachlssigt werden knnen. Bttger (2020) begrndet die kognitiven Vorteile bilingual unterrichteter Grundschler mit einer greren kognitiven Flexibilitt als Folge des stndigen Wechsels zwischen zwei Sprachen. Diese wurde bereits zuvor fr den natrlichen Bilingualismus beobachtet (Diehr, 2016, S. 60). In der vorliegenden Untersuchung konnte jedoch keine berlegenheit der CLIL-Schler bei kognitiv anspruchsvolleren Aufgaben nachgewiesen werden. Statistisch signifikante Unterschiede in den Ergebnissen zu den Aufgaben des Anforderungsbereichs II und III des Chemietests wurden nicht gefunden. Vielmehr finden sich fr den Lernzuwachs durch das Modul in beiden Gruppen auch fr den Testteil zu den Anforderungsbereichen II und III nahezu

identische Werte (siehe Abschnitt 7.3.3). Für Hypothese 2 zur Überlegenheit bilingualer Schüler in den kognitiv anspruchsvolleren Aufgaben finden sich somit keine Belege. Die vermuteten Trainingseffekte für das Gehirn (Böttger 2019) konnten hier nicht nachgewiesen werden. Dabei liefert die vorliegende Studie jedoch keine Antwort auf die Frage, ob solche Effekte sich nicht doch einstellen, wenn bilingualer Unterricht über einen deutlich längeren Zeitraum, wie etwa im Rahmen eines bilingualen Zweigs, erteilt wird. Kognitive Vorteile werden auch für die Behaltensleistung bilingual unterrichteter Schüler vermutet. Heine (2010a) konnte im bilingualen Unterricht eine intensivere kognitive Auseinandersetzung der Schüler mit den sachfachlichen Inhalten beobachten. In diesem Zusammenhang wird häufig das Modell der Verarbeitungstiefe nach Craik und Lockhart (1972) zitiert, wonach eine intensivere semantische Verarbeitung neuer Wissensinhalte auch zu besseren Behaltensleistungen führt (siehe Abschnitt 4.1.3). Anhand der Ergebnisse der Chemieleistungstests der vorgelegten Studie können solche Effekte jedoch nicht bestätigt werden. Vielmehr finden sich für das Vergessen im Anschluss an das Modul in beiden Gruppen nahezu identische Werte (siehe Abschnitt 7.3.1). Piesche (2016, S. 89) kommt in ihrer Untersuchung zum bilingualen Physikunterricht zum gleichen Ergebnis.

(Fremd-) sprachlicher Lernerfolg

Hypothese 4

Der deutsche Fachwortschatz wird von den auf Deutsch unterrichteten Schülern im Anschluss an das Modul besser beherrscht als von den bilingual unterrichteten Schülern.

Da der bilinguale Unterricht den Lehrplänen des jeweiligen Sachfachs verpflichtet ist, muss im Sinne der Entwicklung einer doppelten Fachliteralität sichergestellt sein, dass Fachbegriffe von den Schülern sowohl in der L1 als auch in der L2 erworben werden. Laut Diehr (2016, S. 63f.) besteht hier Forschungsbedarf, auch weil nicht klar ist, inwiefern „in der Fremdsprache erworbene fachliche Kompetenzen auf die Schulsprache Deutsch“ übertragen werden können. Für die vorliegende Untersuchung wurde dabei von der Hypothese ausgegangen, dass der deutsche Fachwortschatz von den auf Deutsch unterrichteten Schülern im Anschluss an das Modul besser beherrscht wird als von den bilingual unterrichteten Schülern (Hypothese 4). Bei der Auswertung der Daten beider Gruppen konnte jedoch kein statistisch signifikanter Unterschied in den Ergebnissen zum deutschen Fachwortschatztest gefunden werden, und zwar weder beim rezeptiven noch beim produktiven Fachwortschatz. Vielmehr war der mittlere Zugewinn der CLIL-Schüler nur minimal kleiner als der der Regelschüler (siehe Abschnitt 7.4.1). Damit liefert die vorliegende Untersuchung keinerlei Hinweise auf die Überlegenheit der Regelschüler beim Erwerb des

deutschsprachigen Fachwortschatzes. Hypothese 4 kann deshalb nicht bestätigt werden. Es ist jedoch festzuhalten, dass die untersuchten Fachbegriffe ausschließlich Stoffbezeichnungen bzw. Bezeichnungen von Laborgeräten bzw. Laborverfahren betreffen. Viele englische Bezeichnungen der vorliegenden Reihe entsprechen dabei den deutschen (z.B. *Lithium/lithium*, *Rubidium/rubidium* oder *Alkalimetall/alkali metal*), so dass der Transfer in die Schulsprache Deutsch, der durch das Aushändigen einer zweisprachigen Vokabelliste unterstützt wurde, recht einfach gelingt, oft mit dem bloßen Verweis auf die Vokabelliste. Bohn & Doff (2010, S. 79f.) stellen für das Fach Biologie ebenfalls fest, dass sich viele Fachbegriffe „eins zu eins von einer Sprache in die andere übertragen“ lassen. Der englische Fachwortschatz wurde in dieser Studie durch zwei separate Tests erfasst. Der erste Test enthielt Fachbegriffe, die auch im Deutschen für die Schüler neu waren, der zweite Test solche, die im Deutschen bereits bekannt waren. Die Aufteilung wurde mit dem Ziel vorgenommen, den Zuwachs an deutschen und englischen Fachbegriffen als Ausdruck einer sich entwickelnden doppelten Fachliteralität vergleichen zu können. Die Testergebnisse belegen, dass die Fachbegriffe, die sowohl im Englischen wie im Deutschen neu waren, gleichgut von den CLIL-Schülern gelernt wurden. Sowohl direkt im Anschluss an das Modul als auch sechs Wochen später beherrschen die CLIL-Schüler etwa 70 % des englischen und des deutschen Fachwortschatzes. Statistisch signifikante Unterschiede zwischen den Ergebnissen des DFW-Tests und des EFW-Tests waren weder für den rezeptiven noch für den produktiven Fachwortschatz zu beobachten (siehe Abschnitt 7.4.2).

Hypothese 5

Der fremdsprachliche Zugewinn der bilingual unterrichteten Schüler betrifft vorwiegend den Bereich des Fachwortschatzes. Zugewinne im Bereich des Allgemeinwortschatzes sind gering.

Die Frage, wie hoch der Anteil der Allgemeinsprache im bilingualen Chemieunterricht ist, kann zumindest für den Wortschatz anhand der vorliegenden Daten beantwortet werden. Die Ergebnisse zu den Fachwortschatz- und Allgemeinwortschatztests zeigen, dass die Schüler im Modul durchschnittlich 45 neue englische Wörter gelernt haben, von denen zwei Drittel der Fachsprache (30 Vokabeln) und ein Drittel dem Allgemeinwortschatz (15 Vokabeln) zuzuordnen sind. Das Verhältnis von erworbenem Fach- zu Allgemeinwortschatz beträgt somit für das Modul 2:1 (siehe Abschnitt 7.4.3). Die Beobachtung deckt sich mit den Ergebnissen von Merikivi und Pietilä (2014), die den englischen Wortschatz von Neuntklässlern in Finnland untersuchten. Während die bilingual unterrichteten Schüler den Regelschülern sowohl im Allgemein- wie im Fachwortschatz überlegen

waren, fanden die Forscher ebenfalls die größten Vorteile der bilingual unterrichteten Schüler im Bereich des Fachwortschatzes. Hypothese 5, wonach der fremdsprachliche Zugewinn der bilingual unterrichteten Schüler vorwiegend den Bereich des Fachwortschatzes betrifft und Zugewinne im Bereich des Allgemeinwortschatzes eher gering sind, kann deshalb so nicht bestätigt werden. Wenngleich der Anteil des Fachwortschatzes überwiegt, so ist der Zuwachs im englischen Wortschatz insgesamt doch beachtlich, denn die Schüler lernten im Schnitt 3,75 neue Vokabeln pro Stunde¹⁴¹. Im Englischunterricht nach dem Cornelsen Lehrwerk Access lernen sie, laut Vorgaben der Autoren, in Klasse 8 bestenfalls 3,9 neue Vokabeln pro Stunde, nämlich dann, wenn sie am Ende der Unit alle neu zu erlernenden Vokabeln beherrschen¹⁴². Da dies vermutlich nicht für alle Schüler der Fall sein wird, dürfte der mittlere Vokabelzuwachs pro Stunde im Fremdsprachenunterricht geringer sein als im vorliegenden Modul. Die guten Ergebnisse der Tests zum englischen Wortschatz belegen somit auch die in der Literatur vielfach zitierte besonders starke Auswirkung bilingualen Unterrichts auf den fremdsprachlichen Wortschatz (vgl. Dalton-Puffer, 2008; Wode et al., 1996; Zydatiß, 2007; Kampmeier, 2013). Dabei ist zu bedenken, dass durch die hier durchgeführten Tests nur der Zuwachs in dem vom Lehrer im Vorfeld definierten Wortschatz erfasst wird. Individuelle Zugewinne, die sich in den für den bilingualen Unterricht typischen Unterrichtssituationen ergeben, wenn im Rahmen von Partner- oder Gruppenarbeit Bedeutungen ausgehandelt werden, werden hier nicht erfasst. Es kann somit davon ausgegangen werden, dass der tatsächliche Zuwachs im englischen Wortschatz der Schüler noch höher ist.

Motivation

Die Auswertung der Subskalen zur intrinsischen Motivation zeigt, dass die Mittelwerte für das Interesse am Modul sowie für die Einschätzung des Nutzens bei den CLIL-Schülern statistisch signifikant höher sind als bei den auf Deutsch unterrichteten Schülern. Der CLIL-Unterricht wird also von den Schülern als interessanter und nützlicher bewertet als der Regelunterricht. Dabei liegen die Mittelwerte der Likert-Skalen in der CLIL-Gruppe oberhalb von 3, was bedeutet, dass die Schüler im Mittel finden, dass die Aussagen „CAE hat mir Spaß gemacht“ bzw. „in CAE haben wir nützliche Dinge gelernt“ „eher stimmen“. Für den Regelunterricht beträgt der Mittelwert etwa 2,5, was bedeutet, dass mehr Schüler finden, dass dieselben Aussagen für den Chemieunterricht auf Deutsch „wenig“ bzw. „nicht stimmen“ (siehe Abschnitt 7.5.1). Damit belegen die Daten zum einen

141 Zum Erlernen der 45 Vokabeln wurden insgesamt 12 Unterrichtsstunden benötigt.

142 Das Lehrwerk Access schlägt 31 Unterrichtsstunden für die Durchführung der einzelnen Units vor, in welchen im Schnitt ca. 120 Vokabeln eingeführt werden.

die Beobachtung, dass Chemie zu den eher unbeliebten Fächern gehört, und zum anderen, dass das Interesse der Schüler durch den Einsatz der Fremdsprache gesteigert werden kann.

Hartmannsgruber (2014), der auch auf die Kurzsкала zur intrinsischen Motivation zurückgreift, findet ebenfalls hohe Werte für das Interesse im bilingualen Biologieunterricht. Noch höhere Werte stellt er jedoch für den Regelunterricht fest. Das kann daran liegen, dass Biologie in der Beliebtheitsskala der Schüler vor Chemie anzusiedeln ist, und der Motivationsvorteil durch den Einsatz der Fremdsprache deshalb weniger zum Tragen kommt. Auch in Bezug auf den Druck und das Kompetenzerleben gibt es statistisch signifikante Unterschiede zwischen der CLIL- und der nicht-CLIL Gruppe. Dabei empfinden die CLIL-Schüler einen größeren Druck und fühlen sich weniger kompetent als die nicht-CLIL Schüler (siehe Abschnitt 7.5.1). Die von Piesche (2016, S. 153) gefundenen signifikant schlechteren Werte beim Kompetenzerleben von Sechstklässlern im bilingualen Physikunterricht können somit für das Fach Chemie in Klasse 8 bestätigt werden.

Ähnliche Beobachtungen können für das Fach Biologie von Hartmannsgruber (2014) jedoch nicht gemacht werden. Er konstatiert lediglich geringfügig erhöhte Werte für den Druck und etwas niedrigere Werte für das Kompetenzerleben in der Gruppe der bilingual unterrichteten Siebtklässler. Diese Unterschiede sind jedoch statistisch nicht signifikant (Hartmannsgruber, 2014, S.89f.)

Für die vorliegende Unterrichtsreihe zeigt sich jedoch, dass die Sorge der CLIL-Schüler bezüglich des höheren Drucks und des geringeren Kompetenzerlebens im bilingualen Unterricht zum Ende der Unterrichtsreihe deutlich abnimmt. Im Anschluss an das Modul bewerten die CLIL-Schüler nämlich sowohl ihr Kompetenzerleben als auch den empfundenen Druck signifikant positiver als vor Beginn des Moduls (siehe Abschnitt 7.5.5). Wenngleich sich die Werte für beide Skalen immer noch statistisch signifikant von den Werten für den Regelunterricht unterscheiden, zeigt sich zumindest, dass die anfänglichen Befürchtungen der Schüler, den Anforderungen des bilingualen Unterrichts nicht gewachsen zu sein, nicht ganz bestätigt wurden. Auch die insgesamt guten Ergebnisse der Chemietests der CLIL-Gruppe deuten auf eine hohe Kompetenz der CLIL-Schüler hin (siehe Abschnitt 7.3.1). Dies kann als Hinweis auf den Erfolg der eingesetzten *scaffolding*-Maßnahmen verstanden werden. Es erscheint deshalb denkbar, dass sich durch die erfolgreiche Teilnahme an weiteren Chemiemodulen die Motivation für den bilingualen Unterricht weiter verbessern lässt. Die Einbußen bei Interesse und Nutzen sind demgegenüber gering. Zwar bewerten die CLIL-Schüler im Anschluss an das Modul dieses als weniger interessant bzw. weniger nützlich als vor dem Modul, der Unterschied ist jedoch gering und statistisch nicht signifikant. Die Werte für Interesse und Nutzen sind, wie gesehen, signifikant größer als die entsprechenden Werte für den Regelunterricht (siehe Abschnitt 7.5.5). Dieses Ergebnis deckt sich mit den Beobachtungen von Lasagabaster & Doiz (2015, S. 20) zu spanischen CLIL-Schülern. Die Forscher finden, dass die

intrinsische Motivation und die motivationale Stärke, also die Bereitschaft der Schüler, sich im auf Englisch geführten bilingualen Unterricht anzustrengen, über einen Zeitraum von zwei bzw. drei Jahren unverändert hoch bleibt, auch wenn diese nicht für die Teilnahme am bilingualen Unterricht ausgewählt wurden. Es scheint laut den Autoren, dass die instrumentelle Motivation aufgrund der großen Bedeutung des Englischen als *lingua franca* besonders motivierend wirkt (Lasagabaster & Doiz, 2015, S. 20f.).

Hypothesen 6 und 7

Im bilingualen Chemieunterricht sind die Werte für die Motivation im Mittel bei fremdsprachlich orientierten Schülern höher als im regulären Chemieunterricht. Im bilingualen Chemieunterricht sind die Werte für die Motivation im Mittel bei den naturwissenschaftlich orientierten Lernern geringer als im regulären Chemieunterricht.

Betrachtet man die vier Subskalen der intrinsischen Motivation differenziert nach den vier Interessensgruppen, so sind zunächst keine statistisch signifikanten Unterschiede zu beobachten. Alle CLIL-Schüler empfinden - unabhängig von ihren Vorlieben für die Naturwissenschaften oder für Fremdsprachen – bilingualen Unterricht als interessanter und nützlicher, verspüren dort allerdings auch einen größeren Druck und fühlen sich weniger kompetent als im Regelunterricht (siehe Abschnitt 7.5.2). Es konnte kein statistisch signifikanter Effekt der Orientierung auf die intrinsische Motivation festgestellt werden. Somit können die beiden Untersuchungshypothesen (Hypothesen 6 und 7), wonach die fremdsprachlich orientierten Schüler im bilingualen Unterricht und die naturwissenschaftlich orientierten Schüler im Regelunterricht stärker motiviert sind, nicht bestätigt werden. Auch die naturwissenschaftlich orientierten Schüler finden bilingualen Unterricht interessanter und nützlicher als den Regelunterricht und auch die fremdsprachlich orientierten Schüler verspüren im bilingualen Unterricht einen größeren Druck und fühlen sich weniger kompetent als im Regelunterricht. Verriere (2014, S. 240) kommt für das Fach Mathematik zu ähnlichen Ergebnissen und folgert, dass die These, „dass mathematikzugewandte Schüler/innen eher unter dem bilingualen Modul in ihrer Lernbereitschaft leiden als widerlegt angesehen werden muss“. Sie beobachtet in ihren Daten vielmehr Hinweise darauf, dass das Kompetenz- und Autonomieerleben sowie die Zufriedenheit dieser Schüler im bilingualen Unterricht zunehmen. Auch für die englischzugewandten Schüler stellt sie neben positiven Reaktionen auf das Modul in verschiedenen Klassen auch Einbußen in den Kategorien Zufriedenheit, Kompetenzerleben und Gefühl der sozialen Eingebundenheit fest. Die These, nach der englischzugewandte Schüler vom bilingualen Modul profitieren, lehnt sie deshalb ebenfalls ab (Verriere, 2014, S. 240). Dabei finden

sich in der vorliegenden Studie die höchsten Werte über alle Subskalen der KIM hinweg betrachtet bei den bso-Schülern, also bei jenen Schülern, die sich sowohl für Naturwissenschaften als auch für Sprachen interessieren. Zum gleichen Ergebnis kommt auch Rodenhauser (2016, S. 269), die die größte intrinsische Motivation bei den „Allroundern“ findet. Dies entspricht ihrer Meinung nach insofern den Erwartungen, da durch den bilingualen Unterricht beiden Vorlieben dieses Schülertyps - sowohl den fremdsprachlichen als auch den biologischen - entsprochen wird. Betrachtet man die einzelnen Subskalen jedoch etwas genauer, so lassen sich dennoch Effekte einer fremdsprachlichen bzw. naturwissenschaftlichen Disposition erkennen. Es zeigt sich etwa, dass das Interesse der fso-Schüler am CLIL-Unterricht etwas größer ist als das der nwo-Schüler, während das Interesse der nwo-Schüler am Regelunterricht höher ist als das der fso-Schüler. Zwar finden sowohl die fso- als auch die nwo-Schüler den CLIL-Unterricht interessanter, der Unterschied ist jedoch bei den fso-Schülern größer. Ihr Interesse für den Chemieunterricht ist sogar größer als das der nwo-Schüler, wenn dieser in der Fremdsprache erteilt wird. Die Unterschiede sind statistisch nicht signifikant, deuten jedoch auf einen positiven Effekt des Fremdspracheninteresses auf das Interesse am bilingualen Unterricht hin. Dieser Effekt wird anhand der durchgeführten

Korrelationsuntersuchungen bestätigt. Man erkennt nämlich, dass das Interesse der CLIL-Schüler am Modul statistisch signifikant mit dem Englischinteresse korreliert (siehe Seite 229). Warum aber interessieren sich fso-Schüler dann nicht signifikant mehr für den bilingualen Unterricht als nwo-Schüler? Der scheinbare Widerspruch lässt sich auf den Unterschied zwischen Fremdsprachen- und Englischinteresse zurückführen. Da fso-Schüler auf dem Fragebogen ankreuzen, sich mehr für *Sprachen* als für Naturwissenschaften zu interessieren, bezieht sich ihr Urteil nicht nur auf *Englisch* sondern auch auf weitere Fremdsprachen, wie etwa Französisch oder Latein, die an der untersuchten Schule unterrichtet werden. Offensichtlich sind Englisch- und Fremdspracheninteresse bei den Schülern unterschiedlich ausgebildet. In der Gruppe der fremdsprachlich orientierten Schüler befinden sich vermutlich auch solche, die sich mehr für andere Sprachen interessieren und die deshalb nicht in gleichem Maße für den auf Englisch geführten bilingualen Unterricht zu begeistern sind. Damit bestätigen diese Ergebnisse die Beobachtung von Lipski-Buchholz (2019, S. 264) für fremdsprachlich interessierte Schüler im Fach Mathematik. Auch sie stellt eine positivere motivationale Disposition im bilingualen Unterricht als im regulären Mathematikunterricht fest. Die Frage, ob damit langfristig aber auch eine größere Auseinandersetzung mit Inhalten der Fächer Chemie oder Mathematik verbunden ist oder sich das größere Interesse nur auf die fremdsprachlichen Anteile des bilingualen Unterrichts bezieht, kann aufgrund der Daten allerdings nicht beantwortet werden (vgl. auch Lipski-Buchholz, 2019, S. 264). Die Vermutung, dass sich eine hohe Affinität für Englisch positiv auf die Motivation im bilingualen Unterricht auswirkt, wird auch

durch weitere Korrelationsuntersuchungen bestätigt. Diese zeigen nämlich, dass das Englischinteresse nicht nur mit dem Interesse für den bilingualen Unterricht, sondern auch mit dem empfundenen Druck statistisch signifikant korreliert. Schüler mit ausgeprägtem Englischinteresse empfinden nämlich weniger Druck im bilingualen Unterricht (siehe Seite 229). Die Zunahme des empfundenen Drucks ist in der fso-Gruppe deutlich geringer als in der nwo-Gruppe. Die Beobachtung deckt sich mit den Ergebnissen von Piesche (2016) für den bilingualen Physikunterricht. Sie findet, dass die fremdsprachliche Orientierung die nachteiligen Effekte auf die Angst im bilingualen Unterricht abmildert. Auch auf das Kompetenzerleben scheint das Englischinteresse einen Einfluss zu haben. Wenngleich die Korrelation statistisch nicht signifikant ist, zeigt sich, dass die Abnahme des Kompetenzerlebens bei Verwendung der Fremdsprache bei den fso-Schülern geringer ausfällt als bei den nwo-Schülern. Sie fühlen sich in einem auf Englisch geführten Chemieunterricht sogar geringfügig kompetenter als die nwo-Schüler. Wichtiger für das Kompetenzerleben im Modul scheint jedoch in der untersuchten Gruppe die Leistungsstärke in Englisch. Korrelationsuntersuchungen ergeben, dass die Ergebnisse im C-Test statistisch hochsignifikant mit dem Kompetenzerleben im Modul korrelieren (siehe Seite 229). Auch in anderen Studien konnte ein positiver Effekt der Englischleistungen auf die Motivationsentwicklung nachgewiesen werden (vgl. Cummings, 1979; Marsh et al., 2000 und Farrell, 2011). Im Gegensatz dazu hat das Englischselbstkonzept in der vorliegenden Untersuchung keinen signifikanten Einfluss auf das Kompetenzerleben oder den Druck im Modul. Piesche (2016) kommt hier für den bilingualen Physikunterricht zu einem anderen Ergebnis. Ein positives Englischselbstkonzept verringert in Piesches (2016) Untersuchung nämlich die nachteiligen Effekte des bilingualen Unterrichts und wirkt sich positiv auf das Kompetenzerleben aus. Die objektiv festgestellten Englischleistungen hingegen moderieren die Motivationsentwicklung in dem von Piesche (2016) untersuchten Modul nur geringfügig. Deshalb folgert sie, dass es auf die eigene Wahrnehmung der Kompetenz ankommt und nicht auf die tatsächlichen Leistungen in Englisch. Auf der anderen Seite zeigt sich, dass das Chemieselbstkonzept statistisch signifikant mit dem empfundenen Druck korreliert und Schüler einen höheren Druck im bilingualen Unterricht verspüren, wenn sie über ein ausgeprägtes Chemieselbstkonzept verfügen (siehe Seite 229). Das Ergebnis deckt sich wieder mit den Ergebnissen von Piesche (2016). Sie stellt nämlich fest, dass ein positives Selbstkonzept und ein großes Interesse an den Naturwissenschaften, die von ihr beobachteten nachteiligen Effekte bilingualen Unterrichts noch verstärken. Dies kann ein Hinweis auf das Vorliegen von Kontrasteffekten nach dem *Internal/External-Frame-of-Reference-Modell* von Marsh (1986) sein. Demnach gehen Schüler, die sich für gute Naturwissenschaftler halten, davon aus, dass sie in Fremdsprachen weniger begabt sind. In einem überwiegend in der Fremdsprache geführten

Unterricht empfinden sie dann entsprechend mehr Druck. Während sich ein großes naturwissenschaftliches Interesse bzw. ein positives naturwissenschaftliches Selbstkonzept negativ auf den Druck oder das Kompetenzerleben im bilingualen Chemie- bzw. Physikunterricht auswirken können, sind entsprechende Ergebnisse zum bilingualen Biologieunterricht widersprüchlich. Rodenhauser (2016), die die Schüler gezielt nach ihrem Interesse an biologischen bzw. fremdsprachlichen Kursinhalten befragt, ist der Meinung, dass „Fremdsprachler“ im Labor den größten Druck hinsichtlich der biologischen Kursanteile und „Naturwissenschaftler“ den größten Druck hinsichtlich der fremdsprachlichen Kursanteile empfinden. Auf der anderen Seite konstatiert Scheersoi (2008) keine vergleichbaren Effekte der Orientierung auf die Motivation im bilingualen Unterricht und sieht den Grund dafür darin, dass das Fach Biologie relativ nah an der verbalen Domäne anzusiedeln ist.

Gendereffekte

Hinweise auf Gendereffekte in Bezug auf die Motivation im bilingualen Chemieunterricht konnten in der vorliegenden Untersuchung nicht gefunden werden. Eine vermutete positivere Auswirkung des Gebrauchs der Fremdsprache auf die stärker fremdsprachlich interessierten Mädchen konnte nicht nachgewiesen werden. So befinden sich in der Gruppe der fremdsprachlich orientierten Schüler zwar viermal so viele Mädchen wie Jungen (siehe Seite 180), eine größere intrinsische Motivation für den bilingualen Unterricht resultiert daraus jedoch nicht. In Bezug auf Interesse, Nutzen, Kompetenzerleben und Druck unterscheiden sich die Werte der Mädchen im bilingualen Unterricht nicht signifikant von denen der Jungen. Genau wie die Jungen finden die Mädchen den bilingualen Unterricht signifikant interessanter und nützlicher, wenngleich sie sich dort weniger kompetent fühlen und einen höheren Druck verspüren (siehe Abschnitt 7.5.3). Gleiches gilt für den Regelunterricht im Fach Chemie. Auch hier ergeben sich keine statistisch signifikanten Unterschiede in Bezug auf die intrinsische Motivation von Mädchen und Jungen. Auf der anderen Seite zeigen die Daten auch, dass sich die Jungen der vorliegenden Stichprobe erwartungsgemäß stärker für Chemie interessieren als die Mädchen (siehe Seite 180), dass sie jedoch durch den Einsatz der Fremdsprache nicht stärker demotiviert werden (siehe Abschnitt 7.5.3). Anhand der hier ermittelten Daten lassen sich jedoch zumindest tendenzielle Unterschiede zwischen der Motivation von Jungen und Mädchen feststellen (vgl. Abschnitt 7.5.3). So zeigt sich, dass die Mädchen der vorliegenden Stichprobe im Regelunterricht etwas weniger motiviert sind als die Jungen. Außerdem fallen ihr Interesse als auch ihre Bewertung des Nutzens des Chemieunterrichts schlechter aus als bei den Jungen. Zudem empfinden sie sich im Regelunterricht weniger kompetent und verspüren einen größeren Druck als die Jungen. Durch den Einsatz der Fremdsprache

vergrößert sich das Interesse der Mädchen so weit, dass es größer ist als das der Jungen. Der zusätzliche Druck, der sich möglicherweise aus dem geringeren Kompetenzerleben im bilingualen Unterricht ergibt, ist bei den Mädchen ebenfalls etwas kleiner als bei den Jungen. Wenngleich diese Unterschiede statistisch nicht signifikant sind, decken sie sich mit den Beobachtungen Verrieres (2014). Sie stellt für die Mädchen ebenfalls tendenziell bessere Werte für die Zufriedenheit und die empfundene Schwierigkeit im bilingualen Mathematikunterricht fest (Verriere (2014, S. 255). Zu ähnlichen Ergebnissen kommt auch Rodenhauser (2016, S. 207f.) für das Fach Biologie. Für keine der von ihr untersuchten Skalen (Interesse, Druck, Kompetenzerleben und wahrgenommene Wahlfreiheit) ergeben sich statistisch signifikante Effekte des Geschlechts, die einen nennenswerten Effekt zur Folge hätten¹⁴³. Auch Piesche (2016, S. 142) sieht in ihrer Untersuchung zum bilingualen Physikunterricht keine statistisch signifikanten Genderunterschiede im Hinblick auf Kompetenzerleben, Angst, intrinsische Motivation und kognitive Mitarbeit im bilingualen Unterricht (Piesche, 2016, S. 132). Insbesondere für die von ihr beobachteten Nachteile in Kompetenzerleben und Angst der bilingualen Schüler ergibt sich keine moderierende Wirkung des Geschlechts. Die Nachteile betreffen Jungen wie Mädchen gleichermaßen.

Fazit

Summa summarum ist festzuhalten, dass die ermittelten Daten zur sachfachlichen, fremdsprachlichen und motivationalen Wirksamkeit des bilingualen Chemieunterrichts nicht gegen den Einsatz bilingualer Module im Chemieunterricht ab Klasse 8 sprechen. Das hier durchgeführte Modul erweist sich weder als fachlich zu schwer noch als fremdsprachlich zu wenig ergiebig und führt zu einem insgesamt größeren Interesse der Schüler. Auch auf die Frage, welche Schüler in besonderem Maße vom bilingualen Unterricht profitieren, und, in welchen Fällen pädagogische Maßnahmen ergriffen werden müssen, gibt die vorgelegte Studie Antworten. Die kognitiven Leistungen der Schüler haben, wie gesehen, keinen Einfluss auf den Lernerfolg. Mit dem Mythos, bilingualer Chemieunterricht sei nur etwas für besonders begabte Schüler, muss deshalb aufgeräumt werden. Während jedoch für leistungsschwache Schüler mit Defiziten zu rechnen ist, und hier weitere Maßnahmen, etwa im Rahmen von Binnendifferenzierung, nötig werden, ist dies bei leistungsstärkeren Schülern nicht zu erwarten. Im Gegenteil, die Studie zeigt, dass Schüler, die

143 Rodenhauser (2016, S. 207) unterscheidet bei den Skalen Interesse und Druck außerdem danach, ob diese sich auf die fremdsprachlichen bzw. biologischen Kursinhalte beziehen. Sie macht dabei nur für „Interesse Englisch“ und „Druck Biologie“ statistisch signifikante Unterschiede aus, die jedoch aufgrund der geringen Effektstärken zu vernachlässigen sind.

besonders leistungsstark in Chemie sind, im bilingualen Unterricht sogar bessere Leistungen erzielen als im Regelunterricht. Auf der anderen Seite haben die Leistungen im Fach Englisch keinen Einfluss auf die sachfachlichen Leistungen. Auch Schüler, die in Englisch leistungsschwach sind, müssen nicht mit Nachteilen im bilingualen Chemieunterricht rechnen, wenn dieser didaktisch geschickt erteilt wird. Das Selbstkonzept im Fach Englisch hat hingegen einen erkennbar positiven Einfluss auf die sachfachlichen Leistungen. Es scheint deshalb sinnvoll, Schülern Ängste, ihr Englischleistungen würden für den bilingualen Chemieunterricht nicht ausreichen, im Vorfeld zu nehmen, und etwa auf entsprechende Ergebnisse aus der Forschung hinzuweisen. Dies ist umso wichtiger, da sich gezeigt hat, dass schlechte Englischleistungen dazu führen, dass Schüler sich weniger kompetent fühlen, obwohl sie am Ende keine schlechteren Leistungen in Chemie zeigen als Regelschüler. Besonders wichtig scheint auch das Interesse für Englisch. Es wurde deutlich, dass Schüler mit einem hohen Englischinteresse den bilingualen Chemieunterricht interessanter finden und weniger Druck verspüren. Auf der anderen Seite muss berücksichtigt werden, dass ein hohes Chemieselbstkonzept sich negativ auswirken kann und dazu führt, dass Schüler einen höheren Druck im bilingualen Unterricht verspüren. Die kann daran liegen, dass die Schüler merken, dass sie chemische Inhalte nicht im gleichen Tempo aufnehmen und verarbeiten zu können, wie im Regelunterricht. Weil aber gerade die leistungsstarken Schüler unter ihnen sogar bessere Leistungen im bilingualen Unterricht erzielen als im Regelunterricht, gilt es im Vorfeld auch hier entsprechend aufzuklären. Die Daten deuten außerdem an, dass durch den wiederholten Einsatz bilingualer Module die motivationale Ausgangslage der Schüler weiter verbessert werden kann. Vieles spricht somit dafür, dem Fach Chemie, zumindest im Rahmen von Modulen, eine größere Beachtung unter den bilingualen Sachfächern zukommen zu lassen.

Literaturverzeichnis

- Abendroth-Timmer, D. (2007). *Akzeptanz und Motivation. Empirische Ansätze zur Erforschung des unterrichtlichen Einsatzes von bilingualen und mehrsprachigen Modulen*. Frankfurt am Main: Peter Lang.
- Adams, R., & Wu, M. (2002). *PISA 2000 Technical Report*. Von <https://www.oecd.org/pisa/data/33688233.pdf> abgerufen
- Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Student's language proficiency in English. *Educational Research and Evaluation* (12), S. 75-93.
- Altrichter, H., & Posch, P. (2007). *Lehrerinnen und Lehrer erforschen ihren Unterricht*. Bad Heilbrunn: Klinkhardt.
- Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Apelt, W. (1981). *Motivation und Fremdsprachenunterricht*. Leipzig: VEB.
- Apelt, W., & Koernig, H. (1994). Affektivität im Fremdsprachenunterricht. *Fremdsprachenunterricht* (47), S. 161-168.
- Apelthauer, E. (2002). Motivation und Mehrsprachigkeit: Veränderungsprozesse während der Aneignung. *Materialien Deutsch als Fremdsprache* (65), S. 143-164.
- Appel, J. (2003). Bilingual biology. Ausbildung für den bilingualen Sachfachunterricht in den naturwissenschaftlichen Fächern. *Praxis des neusprachlichen Unterrichts* 50(2), S. 124-130.
- Aristov, N. (2013). Was ist bekannt über den bilingualen Unterricht in den Naturwissenschaften (Chemie)? Ein Überblick. *Chemkon* (20), S. 169-174.
- Artelt, C., & Wirth, J. (2014). Kognition und Metakognition. In T. Seidel, & A. Krapp, *Pädagogische Psychologie* (S. 167-192). Weinheim: Beltz.
- Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, . . . M. Weiß, *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske + Budrich.
- Atkinson, J. (1975). *Einführung in die Motivationsforschung*. Stuttgart: Klett.

- Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control. In K. Spence, & J. Spence, *The psychology of learning and motivation: Advances in research and theory* (S. 89-195). New York: Academic Press.
- Austad, N. K. (2013). *Vocabulary Testing in CLIL: The Effect of Incidental Vocabulary Learning*. Von <https://www.duo.uio.no/bitstream/handle/10852/36658/VocabularyxTestinginxCLIL.pdf?sequence=1> abgerufen
- Ausubel, D. (1968). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart & Winston.
- Bachmann, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Baddeley, A. (2012). Working memory: Theories, models and controversies. *Annual Review of Psychology* (63), S. 1-29.
- Badertscher, H., & Bieri, T. (2009). *Wisenserwerb im Content and Language Integrated Learning. Empirische Befunde und Interpretationen*. Bern: Haupt.
- Baker, A. (2002). Der Erwerb von akademischen Sprachfähigkeiten im Englischen durch. In B. Krück, & K. Loeser, *Innovation im Fremdsprachenunterricht 2: Fremdsprachen als Arbeitssprachen*. (S. 133-152). Frankfurt am Main: Peter Lang.
- Barke, H.-D., & Harsch, G. (2001). *Chemiedidaktik heute. Lernprozesse in Theorie und Praxis*. Berlin: Springer.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., & Schneider, W. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Becker, H.-J., Hildebrandt, H., & Kühlmann, J. (2012). Chemiedidaktik 2011. *Nachrichten aus der Chemie* (60), S. 352-359.
- Beetz, P., Blell, G., & Klose, D. (2005). Den Anderen ein Stück näher: Fremdverstehen in bilingualen Lehr- und Lernkontexten Geschichte-Englisch. In G. Blell, & R. Kupetz, *Bilingualer Sachfachunterricht und Lehrerbildung für den Bilingualen Unterricht. Forschung und Praxisberichte* (S. 15-50). Frankfurt am Main: Lang.
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development* (70), S. 636-644.
- Bialystok, E., & Martin, M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science* (7), S. 325-339.
- Bloom, B. (1956). *The cognitive domain*. New York: David McKay.
- Bohn, M., & Doff, S. (2010). Biologie bilingual: Die Perspektive der Unterrichtspraxis. In S. Doff, *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (S. 72-88). Tübingen: Narr.

- Bohrmann-Linde, C. (2012). Auf dem Weg zu einer Fachdidaktik Bilingualer Chemieunterricht. In B. Diehr, & L. Schmelter, *Bilingualen Unterricht weiterdenken* (S. 183-200). Frankfurt am Main: Peter Lang.
- Bohrmann-Linde, C. (2013). Chemie. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 295-302). Seelze: Friedrich.
- Bonnet, A. (2002). 47% - Das Spracherwerbspotential englischsprachigen Chemieunterrichts. In S. Breidbach, G. Bach, & D. Wolff, *Bilingualer Sachfachunterricht* (S. 125-140). Frankfurt am Main: Peter Lang.
- Bonnet, A. (2004). *Chemie im bilingualen Unterricht*. Wiesbaden: Springer Fachmedien.
- Bonnet, A. (2007). Fach, Sprache, Interaktion - Eine Drei-Säulen-Methodik für CLIL. *Fremdsprachen Lehren und Lernen* (36), S. 126-141.
- Bonnet, A. (2010). Forschungsmethoden und Forschungsinstrumente. In C. Surkamp, *Metzler Lexikon Fremdsprachendidaktik. Ansätze – Methoden – Grundbegriffe* (S. 67-72). Stuttgart: Metzler.
- Bonnet, A. (2012). CLIL im Fach Chemie - Wachsende Orchidee und Motor der Integration. In B. Diehr, & L. Schmelter, *Bilingualen Unterricht weiterdenken. Programme, Positionen, Perspektiven* (S. 201-218). Frankfurt am Main: Peter Lang.
- Bonnet, A. (2016). Two for the price of one? Das Verhältnis von sachfachlicher und fremdsprachlicher Bildung beim Content and Language Integrated learning. In B. Diehr, A. Preisfeld, & L. Schmelter, *Bilingualen Unterricht weiterdenken und erforschen* (S. 37-56). Frankfurt am Main: Peter Lang.
- Bonnet, A., & Breidbach, S. (2013). Bilingualer Unterricht: Bildungstheoretische Grundlegung. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 26-32). Seelze: Friedrich.
- Boosch, A. (1983). Motivation und Einstellung. In G. Solmecke, *Motivation und Motivieren im Fremdsprachenunterricht* (S. 21-56). Paderborn: Schöningh.
- Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Bos, W., Bonsen, M., & Gröhlich, C. (2009). *Kess 7 - Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7. HANSE-Hamburger Schriften zur Qualität im Bildungswesen Band 5*. Münster: Waxmann.
- Bosenius, P. (2009). Content and language integrated learning: A model for multiliteracy ? In S.-A. Ditze, & A. Halbach, *Bilingualer Sachfachunterricht (CLIL) im Kontext von Sprache, Kultur und Multiliteralität* (S. 15-25). Frankfurt am Main: Peter Lang.

- Böttger, H. (2019). *Studie: Besser in Deutsch und Mathe dank bilinguaem Unterricht in der Grundschule*. Von <https://www.news4teachers.de/2019/04/studie-besser-in-deutsch-und-mathe-dank-bilinguaem-unterricht/> abgerufen
- Böttger, Heiner, & Müller, T. (2020). *Schulversuch Lernen in zwei Sprachen Bilinguale Grundschule Englisch. Abschlussbericht der wissenschaftlichen Evaluation*. Von https://www.researchgate.net/publication/344156958_Schulversuch_Lernen_in_zwei_Sprachen_-_Bilinguale_Grundschule_Englisch abgerufen
- Breidbach, S. (2013). Geschichte und Entstehung des Bilingualen Unterrichts in Deutschland: Bilingualer Unterricht und Gesellschaftspolitik. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 11-17). Seelze: Friedrich.
- Breidbach, S., & Viebrock, B. (2012). CLIL in Germany: Results from Recent Research in a Contested Field of Education. *International CLIL Research Journal 1* (4), S. 1-16.
- Bruegelmann, H. (2005). *Schule verstehen und gestalten. Perspektiven der Forschung auf Probleme*. Von https://www.pedocs.de/volltexte/2019/16758/pdf/Bruegelmann_2005_Schule_verstehen.pdf abgerufen
- Brunnert, R., Harsch, G., & Heimann, R. (2009). *Organic Chemistry goes bilingual - Bilingualer Chemieunterricht*. Münster: Schöningh.
- Brush, T., & Saye, J. (2002). *A Summary of Research Exploring Hard and Soft Scaffolding for Teachers and Learners Using a Multimedia Supported Learning Environment*. Von Journal of Interactive Online Learning 1 (2): <http://www.ncolr.org/jiol/issues/pdf/1.2.3.pdf> abgerufen
- Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Burmeister, P., & Daniel, A. (2002). How effective is late partial immersion? Some findings from a secondary school program in Germany. In P. Burmeister, P. Thorsten, & A. Rhode, *An Integrated View of Language Development. Papers in Honor of Henning Wode* (S. 499-5017). Trier: Wissenschaftlicher Verlag Trier.
- Butzkamm, W. (1993). Bilingualer Unterricht - Fragen an die Forschung. *Die neueren Sprachen* (92), S. 151-161.
- Butzkamm, W. (2000). Über die planvolle Mitbenutzung der Muttersprache im bilingualen Sachfachunterricht. In G. Bach, & S. Niemeier, *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (S. 97-114). Frankfurt am Main: Peter Lang Verlag.
- Bybee, R. (2002). Scientific Literacy - Mythos oder Realität? In W. Gräber, P. Nentwig, T. R. Koballa, & R. H. Evans, *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 21-43). Opladen: Leske und Budrich.

- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction* 8 (4), S. 293-332.
- Chiu, M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction* 18(4), S. 321-336.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge.
- Coyle, D. (2006). Content and language integrated learning: motivating learners and teachers. *Scottish Languages Review* (13), S. 1-18.
- Coyle, D., Hood, P., & Marsh, D. (2012). *CLIL Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Craik, F., & Lockhart, R. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behaviour*, S. 671-684.
- Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*. San Francisco: Jossey-Bass.
- Cummings, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Work. Pap. Biling.* (9), S. 1-43.
- Cummings, J. (1978). The Cognitive Development of Children in Immersion Programms. *Canadian Modern Language Review* (34), S. 855-883.
- Cummings, J. (1984). *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. Clevedon: Multilingual Matters.
- Dallinger, S. (2015). *Die Wirksamkeit bilingualen Sachfachunterrichts: Selektionseffekte, Leistungsentwicklung und die Rolle der Sprachen*. Von https://phbl-opus.phlb.de/frontdoor/deliver/index/docId/67/file/Dissertation_Sara+Dallinger.pdf abgerufen
- Dalton-Puffer, C. (2004). Academic language functions in content and language integrated classrooms (CLIL) : defining and hypothesizing. *Views* 13 (1), S. 23-48.
- Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated Learning (CLIL) Classrooms*. Amsterdam: Benjamins Publishing Company.
- Dalton-Puffer, C. (2008). Outcomes and processes of content and language integrated learning (CLIL): current research from Europe. In W. Delanoy, & L. Volkmann, *Future Perspectives for English Language Teaching* (S. 139- 157). Heidelberg: Universitätsverlag Winter.
- Dalton-Puffer, C. (2013). Diskursfunktionen und generische Ansätze. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 138-145). Seelze: Friedrich.
- Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning. A research agenda. *Language Teaching*, S. 545-559.

- Day, E., & Shapson, S. (1996). *Studies in immersion education*. Clevedon: Multilingual Matters.
- Deci, E. L., & Ryan, R. M. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist* 55 (1), S. 68-78.
- Deci, E., & Ryan, R. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik* (39), S. 223-238.
- Deci, E., & Ryan, R. (2003). *Intrinsic Motivation Inventory*. Von www.psych.rochester.edu/SDT/measures/intrins.html abgerufen
- Demuth, R., Gräsel, C., Parchmann, I., & Ralle, B. (2008). *Chemie im Kontext - Von der Innovation zur nachhaltigen Verbreitung eines Unterrichtskonzepts*. Münster: Waxmann.
- Diehr, B. (2012). What's in a name? Terminologische, typologische und programmatische Überlegungen zum Verhältnis der Sprachen im Bilingualen Unterricht. In B. Diehr, & L. Schmelter, *Bilingualen Unterricht weiterdenken. Programme, Positionen, Perspektiven* (S. 17-36). Frankfurt am Main: Peter Lang.
- Diehr, B. (2016). Doppelte Fachliteralität im bilingualen Unterricht. Theoretische Modelle für Forschung und Praxis. In B. Diehr, *Bilingualen Unterricht weiterentwickeln und erforschen* (S. 57-84). Frankfurt am Main: Peter Lang.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Longman.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in Second Language Research: Construction, Administration and Processing (2nd ed.)*. New York: Routledge.
- Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making Sense of Secondary Science - Research into Children's Ideas*. London, New York: Routledge.
- Ellis, R. (1991). *Instructed Second Language Acquisition*. Oxford: Blackwell.
- Eurydice. (2006). *Content and Language Integrated Learning (CLIL) at School in Europe*. Brüssel: Eurydice.
- Farrell, M. (2011). Bilingual competence and students' achievement in physics and mathematics. *International Journal of Bilingual Education and Bilingualism* (14), S. 335-345.
- Fehling, S. (2005). *Language Awareness und bilingualer Unterricht. Eine komparative Studie*. Frankfurt am Main: Peter Lang.
- Ferber, N. (2014). *Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I*. Berlin: Logos Verlag.
- Field, A. P. (2009). *Discovering statistics using SPSS*. Los Angeles: Sage.
- FMKS. (2022). *Bilinguale Kitas und Schulen*. Von <https://www.fmks.eu/mehrsprachigkeit/rezensionen-paper-fachaufsaetze.html> abgerufen

- Frisch, S. (2016). Sprachwechsel als integraler Bestandteil bilingualen Unterrichts. In B. Diehr, A. Preisfeld, & L. Schmelter, *Bilingualen Unterricht weiterentwickeln und erforschen* (S. 85-102). Frankfurt am Main: Peter Lang.
- Gablasova, D. (2015). Learning technical words through L1 and L2: Completeness and accuracy of word meanings. *English for Specific Purposes* (39), S. 62-74.
- George, D., & Mallery, P. (2002). *SPSS for Windows step by step: A simple guide and reference, 11.0 Update. 4. Auflage*. Boston: Allyn & Bacon.
- Glaserfeld von, E. (1997). *Radikaler Konstruktivismus - Ideen, Ergebnisse, Probleme*. Frankfurt am Main: Suhrkamp.
- Gnutzmann, C., & Jakisch, J. (2013). Language Awareness und Metakognition. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht* (S. 167-174). Seelze: Friedrich.
- Gregorczyk, B. (2012). An empirical study on the acquisition of content in a CLIL-based chemistry course: A preliminary report. *Latin American Journal of Content & Language Integrated Learning*, S. 9-32.
- Grotjahn, R. (1995). Der C-Test: State of the art. *Zeitschrift für Fremdsprachenforschung* 6 (2), S. 37-60.
- Grotjahn, R. (2002). *Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis*. Bochum: AKS.
- Grotjahn, R. (2003). Konzepte für die Erforschung des Lehrens und Lernens fremder Sprachen: Forschungsmethodologischer Überblick. In K.-R. Bausch, H. Christ, & H.-J. Krumm, *Handbuch Fremdsprachenunterricht* (S. 493-499). Tübingen: A. Francke.
- Grünewald, A. (2006). *Multimedia im Fremdsprachenunterricht*. Frankfurt am Main: Peter Lang.
- Haagen-Schützenhöfer, C., Mathelitsch, L., & Hopf, M. (2011). Fremdsprachiger Physikunterricht: Fremdsprachlicher Mehrwert auf Kosten fachlicher Leistung? *Zeitschrift für Didaktik der Naturwissenschaften* (17), S. 223-260.
- Hallet, W. (1998). The Bilingual Triangle - Überlegungen zu einer Didaktik des bilingualen Sachfachunterrichts. *Praxis des neusprachlichen Unterrichts* 45 (2), S. 115-125.
- Hallet, W. (2002). Auf dem Weg zu einer bilingualen Sachfachdidaktik. Bilinguales Lernen als fremdsprachige Konstruktion wissenschaftlicher Begriffe. *Praxis des neusprachlichen Unterrichts* 49, S. 115-126.
- Hallet, W. (2007). Zwischen Bildungsstandards und Mehrsprachigkeit: Kompetenzerwerb im Bilingualen Unterricht. In P. Bosenius, J. Donnerstag, & A. Rohde, *Der bilinguale Unterricht. Englisch aus der Sicht der Fachdidaktiken* (S. 17-36). Trier: Wissenschaftlicher Verlag.
- Hamann, M., & Jördens, J. (2014). Offene Aufgaben codieren. In D. Krüger, I. Parchmann, & H. Schecker, *Methoden der naturwissenschaftsdidaktischen Forschung* (S. 169-178). Berlin: Springer.

- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer, *Leistung und Leistungsdiagnostik* (S. 127-143). Berlin: Springer.
- Hartmannsgruber, M. (2014). *Bilinguale Biologie*. Baltmannsweiler: Schneider.
- Haß, F. (2006). *Fachdidaktik Englisch: Tradition. Innovation. Praxis*. Stuttgart: Klett.
- Hasselhorn, M., & Mähler, C. (1998). Wissen das auf Wissen baut: Entwicklungspsychologische Erkenntnisse zum Wissenserwerb und zum Erschließen von Wirklichkeit im Grundschulalter. In J. Kahlert, *Wissenserwerb in der Grundschule. Perspektiven erfahren, vergleichen, gestalten* (S. 73-89). Bad Heilbrunn: Klinkhardt.
- Hastings, A. (2002). Error analysis of an English C-Test: Evidence for integrated processing. In R. Grotjahn, *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (S. 53-66). Bochum: AKS.
- Heckhausen, H. (1963). *Erfolg und Misserfolg in der Leistungsmotivation*. Meisenheim: Hain.
- Heckhausen, H. (1977). Motivation: Kognitionspsychologische Aufspaltung eines summarischen Konstrukts. *Psychologische Rundschau* (28), S. 175-189.
- Hegerfeldt, A. (2006). Die Fremdsprache im Anfangsunterricht Chemie. *Praxis Fremdsprachenunterricht* (3), S. 36-41.
- Heine, L. (2010a). *Problem solving in a foreign language. A study in Content and Language Integrated Learning*. Berlin: Mouton de Gruyter.
- Heine, L. (2010b). Fremdsprache und konzeptuelle Repräsentation: bilingualer Unterricht aus kognitiver Perspektive. In S. Doff, *Bilingualer Sachfachunterricht in der Sekundarstufe* (S. 199-212). Tübingen: Narr.
- Heine, L. (2013). Empirische Erforschung des Bilingualen Unterrichts. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 216-220). Seelze: Friedrich Verlag.
- Held, G. (2017). *Foreign Language Anxiety and Motivation in the CLIL Classroom*. Von <http://dspace.unive.it/bitstream/handle/10579/11789/840652-1213611.pdf?sequence=2> abgerufen
- Heller, K., & Perleth, C. (2000). *KFT 4 - 12 + R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision ; Manual*. Göttingen: Beltz.
- Helmke, A., & Weinert, F. (1997). Bedingungsfaktoren schulischer Leistung. In F. Weinert, *Enzyklopädie der Psychologie. Themenbereich D: Praxisgebiete. Serie 1: Pädagogische Psychologie: Vol. 3. Psychologie des Unterrichtens und der Schule* (S. 71-176). Göttingen: Hogrefe.
- Helmke, A., Schrader, F.-W., Wagner, W., Nold, G., & Schröder, K. (2008). Selbstkonzept, Motivation und Englischleistung. In E. Klieme, *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 244-257). Weinheim: Beltz.

- Hoffmann, G. (2022). *Didaktische Grundsätze zum Aufbau bilingualer Unterrichtssequenzen*. Von https://arbeitsplattform.bildung.hessen.de/fach/bilingual/Magazin/bilinguale_Unterrichtssequenzen.pdf abgerufen
- Hoffmann, L., Häussler, P., & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik*. Kiel: IPN.
- Hollm, J. (2013). Bilinguales Lehren und Lernen als schulische Innovation aus Sicht der Beteiligten. In J. Hollm, A. Hüttermann, J.-U. Keßler, G. Schlemminger, & B. Ade-Thurw, *Bilinguales Lehren und Lernen in der Sekundarstufe I: Sprache, Sachfach und Schulorganisation* (S. 167-192). Landau: Verlag Empirische Pädagogik.
- Hülden, F. (2014). Sachfachlicher Lernerfolg und sprachlicher Zugewinn im Bilingualen Unterricht. Eine empirische Untersuchung im Rahmen eines bilingualen Moduls im Fach Chemie. *Fremdsprachen Lehren und Lernen (1)*, S. 109-125.
- Jansen, M., Schroeders, K., & Stanat, P. (2013). Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften. In H. A. Pant, P. Stanat, U. Schroeders, A. Ropplet, T. Siegle, & C. Pöhlmann, *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 347-365). Münster: Waxmann.
- Jäppinen, A. (2005). Thinking and content learning of mathematics and science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland. *Language and Education (19)*, S. 148-169.
- Jexenflicker, S., & Dalton-Puffer, C. (2010). The CLIL Differential. Comparing the Writing of CLIL and NON-CLIL Students in Higher Colleges of Technology. In C. Dalton-Puffer, T. Nikula, & U. Smit, *Language Use and Language Learning in CLIL Classrooms* (S. 169-189). Amsterdam.
- Kampmeier, V., & Rosendahl, J. (2013). *Evaluation Gy_ZP-2013 (MSA). Englisch am Gymnasium*. Soest.
- Kauertz, A., Fischer, H., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften (16)*, S. 135-153.
- Kemper, A.-K., & Becker, H.-J. (2016). Bilingualer Physik/Chemieunterricht in der Fremdsprache Deutsch - Erkenntnisse aus einer Fallstudie in Frankreich. *Chemkon 23 (1)*, S. 1-3.
- Kersten, S. (2012). Wortschatzarbeit im bilingualen Unterricht. In F. Lenz, *Bilinguales Lernen. Unterrichtskonzepte zur Förderung sachfachbezogener und interkultureller Kompetenz* (S. 195-206). Frankfurt am Main: Peter Lang.
- Kieweg, W. (2003). Die Rolle der Emotionen beim Fremdsprachenlernen. *Der fremdsprachliche Unterricht - Englisch (3)*, S. 4-11.

- Klingauf, M. (2002). Chemie auf Englisch: Bilingualer Unterricht in einem ungewöhnlichen Sachfach. In C. Finkbeiner, *Lehren und Lernen in zwei Sprachen. Bilingualer Unterricht* (S. 49-61). Hannover: Schroedel.
- KMK. (2005). *Bildungsstandards im Fach Chemie für den mittleren Bildungsabschluss*. Neuwied: Luchterhand Verlag.
- KMK. (2011). *Empfehlungen der Kultusministerkonferenz zur Stärkung der Fremdsprachenkompetenz (Beschluss der Kultusministerkonferenz vom 8.12.2011)*. Von https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2011/2011_12_08-Fremdsprachenkompetenz.pdf abgerufen
- KMK. (2013). *Konzepte für den bilingualen Unterricht - Erfahrungsbericht und Vorschläge zur Weiterentwicklung: Beschluss der Kultusministerkonferenz vom 17.10.2013*. Von https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2013/2013_10_17-Konzepte-bilingualer-Unterricht.pdf abgerufen
- Koch, A. (2005). *Bilingualer naturwissenschaftlicher Anfangsunterricht: Entwicklung - Erprobung - Evaluation*. Von https://macau.uni-kiel.de/receive/diss_mods_00001531?lang=en abgerufen
- Köller, O., & Baumert, J. (2008). Entwicklung schulischer Leistungen. In R. Oerter, & L. Montada, *Entwicklungspsychologie* (S. 735-768). Weinheim: Beltz.
- Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education* (32), S. 448-470.
- Köller, O., Fischer, H. E., Mayer, J., Sumfleth, E., Hartmann, S., Katzenbach, M., . . . Zilker, I. (2008a). *Evaluation der Standards in den Fächern Biologie, Chemie und Physik für die Sekundarstufe I (ESNaS). Band 1: Projektbeschreibung*.
- Köller, O., Fischer, H., Mayer, J., Sumfleth, E., Hartmann, S., Katzenbach, M., . . . Zilker, I. (2008b). *Evaluation der Standards in den Fächern Biologie, Chemie und Physik für die Sekundarstufe I (ESNaS). Band 2: Aufgabenkonstruktionsanleitung für die Aufgabenentwicklung für den Kompetenzbereich "Umgang mit Fachwissen"*.
- Kondring, B., & Ewig, M. (2006). Aspekte der Leistungsmessung im bilingualen Biologieunterricht. *Zeitschrift für Didaktik Der Biologie (ZDB) - Biologie Lehren und Lernen* 14(1), S. 49-62.
- Krapp, A., & Ryan, R. (2002). Selbstwirksamkeit und Lernmotivation: Eine kritische Betrachtung von Bandura aus der Sicht der Selbstbestimmungstheorie und der pädagogisch-psychologischen Interessenstheorie. *Zeitschrift für Pädagogik. Selbstwirksamkeit und Motivationsprozess in Bildungsinstitutionen* (44), S. 54-82.
- Krapp, A., & Seidel, T. (2014). *Pädagogische Psychologie*. Weinheim: Beltz.

- Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.
- Krechel, H.-L. (2003). Bilingual Modules. In M. Wildhage, & E. Otten, *Praxis des bilingualen Unterrichts* (S. 194-216). Berlin: Cornelsen Verlag Scriptor.
- Krechel, H.-L. (2013). Organisationsformen und Modelle in weiterführenden Schulen. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht - Content and Language Integrated Learning* (S. 74-80). Seelze: Friedrich.
- Krechel, H.-L. (kein Datum). Organisationsformen und Modelle in weiterführenden Schulen. In *Handbuch Bilingualer Unterricht* (S. 74-80). Seelze: Friedrich Verlag GmbH.
- Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* (33), S. 149-174.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., . . . Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests: A teacher's book*. Bristol: Longmans, Green and Company.
- Lamsfuß-Schenk, S. (2008). *Fremdverstehen im bilingualen Geschichtsunterricht. Eine Fallstudie*. Frankfurt am Main: Peter Lang.
- Landis, J., & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* (33) 1, S. 159-174.
- Lasagabaster, D., & Doiz, A. (2015). *A Longitudinal Study on the Impact of CLIL on Affective Factors*. Von https://www.researchgate.net/publication/284709599_A_Longitudinal_Study_on_the_Impact_of_CLIL_on_Affective_Factors abgerufen
- Laupenmühlen, J. (2012). Das Bilinguale im Biologieunterricht - Konzeptbildung in zwei Sprachen. In B. Diehr, & L. Schmelter, *Bilingualen Unterricht weiterdenken. Programme, Positionen, Perspektiven* (S. 163-182). Frankfurt am Main: Peter Lang.
- Lehmann, R., Peek, R., Gänsfuß, R., & Husfeldt, V. (2011). *LAU - Aspekte der Lernausgangslage und der Lernentwicklung*. Münster: Waxmann.
- Leisen, J. (1991). Über Sprachprobleme im deutschsprachigen Fachunterricht am Beispiel des Physikunterrichts. *Zielsprache Deutsch* (3), 143-151.
- Leisen, J. (2005). Wechsel der Darstellungsformen. Ein Unterrichtsprinzip für alle Fächer. *Der Fremdsprachliche Unterricht* (78), S. 9-11.

- Leisen, J. (2013). *Handbuch Sprachförderung im Fach - Sprachsensibler Fachunterricht in der Praxis*. Stuttgart: Klett.
- Lenhard, W., Lenhard, A., & Schneider, W. (2017). *Ein Leseverständnistest für Erst- bis Siebtklässler II (ELFE II)*. Göttingen: Hogrefe.
- Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, . . . J. Rost, *PISA 2003: Der Bildungsstandard der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 147-175). Münster: Waxmann.
- Lienert, G., & Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lipski-Buchholz, K. (2019). *Bilingualer Mathematikunterricht : Motivation der Schülerinnen und Schüler für Fremdsprache und Mathematik*. Von https://publikationsserver.tu-braunschweig.de/rsc/viewer/dbbs_derivate_00045562/Diss_Lipski-Buchholz_Kathrin.pdf?page=1 abgerufen
- Lo, Y., & Lo, E. (2014). A meta-analysis of the effectiveness of English-medium education in Hong-Kong. *Review of Educational Research* (84), S. 47-73.
- Long, M. (1980). *Input, interaction and second-language acquisition*. Los Angeles: University of California.
- Long, M. (1991). Focus on form: a design feature in language teaching and methodology. In K. de Bot, R. Ginsberg, & J. Kramsch, *Foreign language research in cross-cultural perspective* (S. 39-52). Amsterdam: Benjamins.
- Lose, J. (2007). The language of scientific discourse: Ergebnisse einer empirisch-deskriptiven Interaktionsanalyse zur Verwendung fachbezogener Diskursfunktionen im bilingualen Biologieunterricht. In D. Caspari, W. Hallet, A. Wegner, & W. Zydatiß, *Bilingualer Unterricht macht Schule* (S. 97-108). Frankfurt am Main: Peter Lang.
- Luprichova, J. (2013). *Modernization of Teaching English as a foreign Language by Means of CLIL Methodology*. Nitra: Constantine the Philosopher University.
- Madrid, D. (2011). Monolingual and bilingual student's competence in social sciences. In D. Madrid, & S. Hughes, *Studies in bilingual education* (S. 195-222). Bern: Lang.
- Marsh, D. (2000). *Using languages to learn and learning to use languages*. Jyväskylä, Finnland: University of Jyväskylä.
- Marsh, D. (2002). *Clil/Emile: The European Dimension : Actions, Trends and Foresight Potential*. Jyväskylä/Finnland: UniCom.
- Marsh, H. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal* (26), S. 191-225.

- Marsh, H., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology* 96 (1), S. 56-67.
- Marsh, H., Hau, K.-T., & Hong, C.-K. (2000). Late immersion and language of instruction in Hong Kong High School : Achievement growth in language and nonlanguage subjects. *Harvard Educational Review* (70), S. 302-347.
- Mäsch, N. (1993). The German model of bilingual education: an administrator's perspective. In H. Beatens Beardsmore, *European models of bilingual education* (S. 155-172). Clevedon: Multilingual Matters.
- Maxis-Gehrke, S., & Bonnet, A. (2001). Naturwissenschaften im Bilingualen Unterricht. In A. Bonnet, & P. Kahl, *Innovation und Tradition im Englischunterricht* (S. 153-180). Stuttgart: Klett.
- Mayring, P. (2000). *Qualitative Inhaltsanalyse. Forum Qualitative Sozialforschung*. Von <https://www.qualitative-research.net/index.php/fqs/article/view/1089/2383> abgerufen
- Mehisto, P., Marsh, D., & Frigols, M. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.
- Meindl, C. (2011). *Methodik für Linguisten: Eine Einführung in Statistik und Versuchsplanung*. Tübingen: Narr.
- Mentz, O. (2010). Alle Fächer eignen sich - oder doch nicht ? Überlegungen zu einem bilingualen Fächerkanon. In S. Doff, *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (S. 29-43). Tübingen: Narr.
- Mercer, N. (2008). Talk and the development of reasoning and understanding. *Human Development* (51), S. 91-100.
- Meyer, C. (2002). *Bedeutung, Wahrnehmung und Bewertung des bilingualen Geographieunterrichts. Studie zum zweisprachigen Erdkundeunterricht (Englisch) in Rheinland-Pfalz*. Von <https://ubt.opus.hbz-nrw.de/opus45-ubtr/frontdoor/deliver/index/docId/75/file/20021118.pdf> abgerufen
- Meyer, O. (2010). Towards quality-CLIL: successful planning and teaching strategies . *Puls* (33), S. 11-29.
- Mitchell, M. (1993). Situational Interest: Its Multifaceted Structure in the Secondary School Mathematics Classroom. *Journal of Educational Psychology* 85 (3), S. 424-436.
- Möller, J., & Trautwein, U. (2015). Selbstkonzept. In E. Wild, & J. Möller, *Pädagogische Psychologie* (S. 177-199). Berlin: Springer.

- MSW-NRW. (2012). *Bilingualer Unterricht. Biologie deutsch-englisch in der Sekundarstufe I*. Von https://www.schulentwicklung.nrw.de/cms/upload/bilingualer_Unterricht/documents/HR_BU_BioE_SekI_0912.pdf abgerufen
- MSW-NRW. (2015). *Chemie. Übersicht über die Operatoren*. Von <https://www.standardsicherung.schulministerium.nrw.de/cms/zentralabiturgost/faecher/getfile.php?file=3850> abgerufen
- MSW-NRW. (2021). *Statistik-Telegramm 2020/21*. Von <https://www.schulministerium.nrw/system/files/media/document/file/stattelegamm2020.pdf> abgerufen
- MSW-NRW. (2022). *Schule suchen*. Von https://www.schulministerium.nrw.de/BiPo/SchuleSuchen/pages/schulsuche/such_ergebnis.xhtml abgerufen
- Mummendey, H., & Grau, I. (2008). *Die Fragebogen-Methode*. Göttingen: Hogrefe.
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: CUP.
- Navarro-Pablo, M., & Garcia-Jimenez, E. (2018). *Are CLIL Students More Motivated? An Analysis of Affective Factors and their Relation to Language Attainment Porta Linguarum*. Von https://www.researchgate.net/publication/322506930_Are_CLIL_Students_More_Motivated_An_Analysis_of_Affective_Factors_and_their_Relation_to_Language_Attainment_Porta_Linguarum abgerufen
- Nerdel, C. (2017). *Grundlagen der Naturwissenschaftsdidaktik*. Berlin: Springer.
- Niemeier, S. (2005). Bilingualismus und "bilinguale" Bildungsgänge aus kognitiv-linguistischer Sicht. In G. Bach, & S. Niemeier, *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (S. 27-49). Frankfurt am Main: Lang.
- Nitz, S., Nerdel, C., & Prechtel, H. (2012). Entwicklung eines Erhebungsinstruments zur Erfassung der Verwendung von Fachsprache im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften* (18), S. 117-139.
- Nold, G. (2003). The impact of the learning culture on achievement in the English as a foreign language classroom - a view from Germany. In J. Rymarczyk, & H. Haudeck, *In search of the active learner. Untersuchungen zu Fremdsprachenunterricht, bilingualen und interdisziplinären Kontexten* (S. 163-184). Frankfurt am Main: Lang.
- Novak, J., & Canas, A. (2006). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Von <http://cmap.ihmc.us/docs/theory-of-concept-maps.php> abgerufen
- OECD. (2006). *Assessing scientific, reading and mathematical literacy. A framework for PISA 2006*. Paris: OECD Publishing Service.

- Ohlberger, S., & Wegner, C. (2018). *Bilingualer Sachfachunterricht in Deutschland und Europa - Darstellung des Forschungsstands*. Von <https://pub.uni-bielefeld.de/download/2932563/2956680/2390-Artikeltext-8634-3-10-20191022.pdf> abgerufen
- Ohlberger, S., & Wegner, C. (2019). CLIL modules and their affective impact on students with high English anxiety and low self-efficacy. *Apples - Journal of Applied Language Studies* 13 (3), S. 1-15.
- Osterhage, S. (2007). Sachfachkönnen (scientific literacy) bilingual und monolingual unterrichteter Biologieschüler: ein Kompetenzvergleich. In A. Hallet, D. Caspari, A. Wegner, & W. Zydatiß, *Bilingualer Unterricht macht Schule* (S. 41-50). Frankfurt am Main: Peter Lang.
- Otten, E., & Wildhage, M. (2003). Content and Language Integrated Learning - Eckpunkte einer "kleinen" Didaktik des bilingualen Sachfachunterrichts. In E. Otten, & M. Wildhage, *Praxis des bilingualen Unterrichts* (S. 12-45). Berlin: Cornelsen-Scriptor.
- Otwinowska, A. (2013). CLIL lessons in the upper primary: The interplay of affective factors and CALP. In D. Gabrys-Barker, & J. Bielska, *Affectivity in Second Language Acquisition* (S. 211-225). Multilingual Matters.
- Otwinowska, A., & Forys, M. (2015). They learn the CLIL way, but do they like it? Affectivity and cognition in upper-primary CLIL classes. *International Journal of Bilingual Education and Bilingualism* 20 (5), S. 457-480.
- Paviscic, C. I. (2011). *CLIL Teaching: an Opportunity to Teach Chemistry*. Von https://conference.pixel-online.net/conferences/ICT4LL2011/common/download/Abstract_pdf/pdf/CLIL11-323-ABS-Paviscic-ICT4LL2011.pdf abgerufen
- Pavlenko, A. (2009). Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning. In A. Pavlenko, *The Bilingual Mental Lexicon. Interdisciplinary Approaches* (S. 125-160). Bristol: Multilingual Matters.
- Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins.
- Pfeifer, P., Häuseler, K., & Lutz, B. (1992). *Konkrete Fachdidaktik Chemie*. München: Oldenbourg.
- Piesche, N. (2016). *CLIL im naturwissenschaftlichen Unterricht - Auswirkungen auf den Wissenszuwachs und die Motivation im Sachfach*. Von <https://phbl-opus.phlb.de/frontdoor/deliver/index/docId/459/file/Dissertation+Endfassung+17.02.16.pdf> abgerufen
- Pietilä, P., & Merikivi, R. (2014). The Impact of Free-time Reading on ForeignLanguage Vocabulary Development. *Journal of Language Teaching and Research* 5 (1), S. 28-36.
- Pokrivcakova, S. (2013). *CLIL Research in Slovakia*. Hradec Králové: University of Hradec Králové Press.

- Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education* 66 (2), S. 211-227.
- Preisfeld, A. (2016). Die Bedeutung bilingualen Experimentalunterrichts in Biologie für die fachliche und sprachliche Kompetenz. In B. Diehr, A. Preisfeld, & L. Schmelter, *Bilingualen Unterricht weiterentwickeln und erforschen* (S. 103-124). Frankfurt am Main: Peter Lang.
- Prüfer, K. (2012). Anwendungsbeitrag: Fragebogenentwicklung und -pilotierung im Rahmen des Dissertationsprojekts "Bilinguale Module im Mathematikunterricht". In S. Doff, *Fremdsprachenunterricht empirisch erforschen. Grundlagen - Methoden - Anwendung* (S. 136-149). Tübingen: Narr Francke Attempto.
- Prüfer, K. (2012). Bilinguale (englisch-deutsch) Module im Mathematikunterricht und ihre Auswirkung auf die Lernbereitschaft der Schüler/innen für das Sachfach: Theoretische Grundlagen und ausgewählte Aspekte eines Pilotmoduls. In F. Lenz, *Bilinguales Lernen* (S. 149-168). Frankfurt am Main: Peter Lang.
- Qualis-NRW. (2022). *Kernlehrplan Chemie für die gymnasiale Oberstufe - Übersicht*. Von <https://www.schulentwicklung.nrw.de/lehrplaene/lehrplannavigator-s-ii/gymnasiale-oberstufe/chemie/chemie-klp/aufgaben-ziele/aufgaben-und-ziele-des-faches.html> abgerufen
- Raynor, J., Brown, W., & Atkinson, J. (1974). Subjective aspects of achievement motivation immediately before an examination. In J. Atkinson, & J. Raynor, *Motivation and Achievement* (S. 155-171). New York: Winston.
- Read, J. (2011). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Reinmann, G., & Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten. In A. Krapp, & A. Weidenmann, *Pädagogische Psychologie* (S. 613-658). Weinheim: Beltz PVU.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., & Klieme, E. (2016). *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. Von https://www.researchgate.net/publication/311587945_PISA_2015_Eine_Studie_zwischen_Kontinuitat_und_Innovation abgerufen
- Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller, & O. Köller, *Emotionen, Kognitionen und Schulleistung* (S. 175-190). Weinheim: Beltz PVU.
- Renkl, A. (2008). Lernen und Lehren im Kontext der Schule. In A. Renkl, *Lehrbuch Pädagogische Psychologie* (S. 109-153). Bern: Huber.
- Richter, R., & Zimmermann, M. (2003). Und es geht doch: Naturwissenschaftlicher Unterricht auf Englisch. In M. Wildhage, & E. Otten, *Praxis des Bilingualen Unterrichts* (S. 116-146). Berlin: Cornelsen Verlag Scriptor.

- Riemeier, T. (2007). Moderater Konstruktivismus. In D. Krüger, & H. Vogt, *Theorien in der biomedizinischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 69-79). Berlin: Springer.
- Riese, J., & Reinhold, P. (2014). Entwicklung eines Leistungstests für fachdidaktisches Wissen. In D. Krüger, I. Parchmann, & H. Schecker, *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 257-267). Berlin: Springer.
- Rittersbacher, C. (2006). We're doing science. Bilingualer Sachfachunterricht an der Realschule. *Praxis Fremdsprachenunterricht* (3), S. 31-35.
- Rodenhauser, A. (2016). *Bilinguale biologische Schülerlaborkurse. Konzeption und Durchführung sowie Evaluation der kognitiven und affektiven Wirksamkeit*. Von <http://elpub.bib.uni-wuppertal.de/edocs/dokumente/fbc/biologie/diss2016/rodenhauser/dc1626.pdf> abgerufen
- Rodenhauser, A., & Preisfeld, A. (2016). A Glue from Snail Slime?! - Umsetzung und Evaluation eines bilingualen Moduls für den Biologieunterricht. In B. Diehr, A. Preisfeld, & L. Schmelter, *Bilingualen Unterricht weiterentwickeln und erforschen* (S. 221-244). Frankfurt am Main: Peter Lang.
- Ropohl, M. (2010). *Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion : Entwicklung und Analyse von Testaufgaben*. Berlin: Logos.
- Rumlich, D. (2012). (Sprach-)Tests in der Praxis: Die Studie „Development Of North Rhine-Westphalian CLIL Students“ (DENOCS). In S. Doff, *Fremdsprachenunterricht empirisch erforschen. Grundlagen - Methoden - Anwendung* (S. 169-178). Tübingen: Narr.
- Rumlich, D. (2016). *Evaluating Bilingual Education in Germany*. Frankfurt am Main: Peter Lang.
- Rymarczyk, J. (2003). *Kunst auf Englisch?* München: Langenscheidt-Longman.
- Scheersoi, A. (2008). Lernmotivation im bilingualen Biologieunterricht. In A. Scheersoi, & H. Klein, *Bilingualer Biologieunterricht* (S. 69-88). Aachen: Shaker.
- Schiefele, U. (2004). Förderung von Interessen. In G. Lauth, M. Grünke, & J. Brunstein, *Interventionen bei Lernstörungen* (S. 134-144). Göttingen: Hogrefe.
- Schiefele, U., & Schaffner, E. (2015). Motivation. In E. Wild, & J. Möller, *Pädagogische Psychologie* (S. 153-175). Berlin: Springer.
- Schiefele, U., & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen: ein Überblick zu Ergebnissen der Forschung. *Zeitschrift für Pädagogische Psychologie* (8), S. 1-13.
- Schiefele, U., Krapp, A., & Schreyer, I. (1993). Metanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* (25), S. 120-148.

- Schiepe-Tiska, A., Rönnebeck, S., & Neumann, K. (2019). Naturwissenschaftliche Kompetenz in PISA 2018 - aktueller Stand, Veränderungen und Implikationen für die naturwissenschaftliche Bildung in Deutschland. In K. Reiss, M. Weis, E. Klieme, & O. Köller, *PISA 2018. Grundbildung im internationalen Vergleich* (S. 211-240). Münster: Waxmann.
- Schiepe-Tiska, A., Simm, I., & Schmidtner, S. (2016). Motivationale Orientierungen, Selbstbilder und Berufserwartungen in den Naturwissenschaften in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller, *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann Verlag GmbH.
- Schlemminger, G., & Buchmann, M. (2013). Einstellungen von Schülerinnen und Schülern zum fremdsprachigen Sachfachunterricht. Ergebnisse einer Umfrage. In J. Hollm, A. Hüttermann, J. Keßler, G. Schlemminger, & B. Ade-Thurow, *Bilinguales Lehren und Lernen in der Sekundarstufe I: Sprache, Sachfach und Schulorganisation* (S. 209-228). Landau: Verlag Empirische Pädagogik.
- Schneider, W., Schlagmüller, M., & Ennemoser, M. (2017). *Lesegeschwindigkeits- und -verständnistest für die Klassen 5-12*. Göttingen: Hogrefe.
- Seikkula-Leino, J. (2007). CLIL Learning: Achievement Levels and Affective Factors. *Language and Education* 21 (4), S. 328-341.
- Selinker, L. (1972). Interlanguage. *IRAL* 10 (3), S. 31-54.
- Sharpe, T. (2001). Scaffolding in Action. Snapshots from the Classroom. In J. Hammond, *Scaffolding: Teaching and Learning in Language and Literacy Education*. (S. 31-48). Newtown: Peta.
- Siegler, R., DeLoache, J., & Eisenberg, N. (2008). *Entwicklungspsychologie im Kindes- und Jugendalter*. Heidelberg: Spektrum.
- Smit, U. (2008). The AILA research network - CLIL and immersion classrooms: applied linguistic perspectives. *Language Teaching* 41 (2), S. 295-298.
- Spörlein, E. (2003). *Das mit dem Chemischen find ich nicht so wichtig... - Chemielernen in der Sekundarstufe I aus der Perspektive der Bildungsdidaktik*. Opladen: Leske und Budrich.
- Stäudel, L. (2008). Mit Informationen umgehen - Übersetzung zwischen verschiedenen Darstellungsformen. *Unterricht Chemie* (106), S. 40-51.
- Steinkamp, M., & Maehr, M. (1983). Affect, Ability and Science Achievement: A Quantitative Synthesis of Correlation Research. *Review of Educational Research* (53), S. 369-396.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer, *Principles and practice in the study of language* (S. 125-144). Oxford: Oxford University Press.
- Sylvén, L. (2004). *Teaching in English or English teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition. PhD dissertation*. Göteborg University.

- Thürmann, E. (2010). Zur Konstruktion von Sprachgerüsten im bilingualen Sachfachunterricht. In S. Doff, *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (S. 137-153). Tübingen: Narr.
- Thürmann, E. (2013). Scaffolding. In Hallet, Wolfgang, & F. Königs, *Hanbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 236-242). Seelze: Friedrich.
- Ting, Y. (2013). *Why is Content and Language Integrated Learning (CLIL) Humanistic?* Von <http://old.hltmag.co.uk/feb13/mart02.htm> abgerufen
- Todtenhaupt, S. (1995). *Zur Entwicklung des Chemieverständnisses bei Schülern*. Frankfurt am Main: Peter Lang.
- van de Craen, P., Ceuleers, E., & Mondt, K. (2007a). Cognitive development and bilingualism in primary schools : teaching maths in a CLIL environment. In D. Marsh, & D. Wolff, *Diverse contexts - converging goals: CLIL in Europe* (S. 185-200). Frankfurt am Main: Peter Lang.
- van de Craen, P., Mondt, K., & Surmont, J. (2012). What content and language integrated learning has taught pedagogy and how to explain it. In O. Meyer, & D. Marsh, *Quality interfaces. Examining evidence & exploring solutions in CLIL* (S. 27-38). Eichstätt: Academic Press.
- van de Craen, P., Mondt, K., Allain, L., & Gao, Y. (2007b). Why and how CLIL works. An outline for a CLIL theory. *Vienna English Working Papers* 16(3), S. 70-78.
- Verriere, K. (2014). *Bilinguale Module im Mathematikunterricht und ihr Einfluss auf die Lernbereitschaft der Schüler/innen für das Sachfach*. Trier: Wissenschaftlicher Verlag Trier.
- Viebrock, B. (2007). *Bilingualer Erdkundeunterricht*. Frankfurt am Main: Peter Lang.
- Vögeding, J. (1995). *"Wenn in einen gesättigten Wasser Kochsalz gibt ... "*. Heidelberg: Groos.
- Vollmer, H. (2000). Bilingualer Sachfachunterricht als Inhalts- und Sprachlernen. In G. Bach, & S. Niemeier, *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (S. 51-73). Frankfurt am Main: Peter Lang.
- Vollmer, H. (2005). Förderung des Spracherwerbs im bilingualen Sachfachunterricht. In G. Bach, & S. Niemeier, *Bilingualer Unterricht Grundlagen, Methoden, Praxis, Perspektiven* (S. 131-150). Frankfurt am Main: Peter Lang.
- Vollmer, H. (2006). Fachlichkeit und Sprachlichkeit: Zwischenbilanz eines DFG-Projekts. *Zeitschrift für Fremdsprachenforschung* 17 (2), S. 201-244.
- Vollmer, H. (2009). Diskursfunktionen und fachliche Diskurkompetenz bei bilingualen und monolingualen Geografielernern. In S.-A. Ditze, & A. Halbach, *Bilingualer Sachfachunterricht (CLIL) im Kontext von Sprache, Kultur und Multiliteralität* (S. 165-185). Frankfurt am Main: Peter Lang.

- Vollmer, H. (2010). Bilingualer Unterricht als Inhalts- und Sprachenlernen. In G. Bach, & S. Niemeier, *Bilingualer Unterricht: Grundlagen, Methoden, Praxis, Perspektiven* (S. 47-70). Frankfurt am Main: Peter Lang.
- Vollmer, H. (2011). *Schulsprachliche Kompetenzen: Zentrale Diskursfunktionen*. Von <https://www.home.uni-osnabrueck.de/hvollmer/VollmerDF-Kurzdefinitionen.pdf> abgerufen
- Vollmer, H. (2013). Das Verhältnis von Sprach- und Inhaltslernen im Bilingualen Unterricht. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 124-131). Seelze: Friedrich.
- Vygotsky, L. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Walpuski, M., Kauertz, A., Kampa, M., Fischer, H., Mayer, J., & Sumfleth, E. (2010). ESNaS - Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrman, U. Hericks, & M. Lüders, *Bildungsstandards und Kompetenzmodelle - Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 171-184). Bad Heilbrunn: Julius Klinkhardt.
- Weiner, B. (1986). Attributions, emotion and action. In R. Sorrentino, & E. Higgins, *Handbook of motivation and cognition* (S. 281-312). New York: Guilford Press.
- Weller, F.-R. (1993). Bilingual oder zweisprachig? - Kritische Anmerkungen zu den Möglichkeiten und Grenzen fremdsprachigen Sachunterrichts. *DNS 92 (1)*, S. 8-22.
- Weskamp, R. (2001). *Fachdidaktik: Grundlagen & Konzepte*. Berlin: Cornelsen.
- Wilde, M., Bätz, K., Kovalevka, A., & Urhahne, D. (2009). Überprüfung einer Kurzskaala intrinsischer Motivation (KIM) . *Zeitschrift für Didaktik der Naturwissenschaften (15)*, S. 31-45.
- Wildhage, M. (2000). Bilingualer Unterricht in Niedersachsen. *Neusprachliche Mitteilungen (53)*, S. 212-219.
- Williams, M., & Burden, R. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge: Cambridge University Press.
- Witzigmann, S. (2011). *Bildende Kunst in der Zielsprache Französisch als Einstieg ins bilinguale Lehren und Lernen*. Frankfurt am Main: Peter Lang.
- Wode, Henning, Burmeister, P., Daniel, A., Kickler, K.-U., & Knust, M. (1996). Die Erprobung von deutsch-englisch bilinguaem Unterricht in Schleswig-Holstein: Ein erster Zwischenbericht. *Zeitschrift für Fremdsprachenforschung*, S. 15-42.
- Wolff, D. (1997). Zur Förderung von Sprachbewusstheit und Sprachlernbewusstheit im bilingualen Sachfachunterricht. *Fremdsprachen Lehren und Lernen*, S. 167-183.
- Wolff, D. (2011). Der bilinguale Sachfachunterricht (CLIL): Was dafür spricht, ihn als innovatives didaktisches Konzept zu bezeichnen. *Forum Sprache (6)*, S. 75-83.

- Wolff, D. (2017). *Stem - Content Learning in German - with CLIL*. Von Historical Development of CLIL: <https://www.goethe.de/en/spr/unt/kum/clg.html#> abgerufen
- Zaunbrecher, A., Retelsdorf, J., & Möller, J. (2009). Die Vorhersage von Englischleistungen am Anfang der Sekundarstufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 41 (3), S. 153-164.
- Zimbardo, P., & Gerrig, R. (2004). *Psychologie (16. Aufl.)*. München: Pearson.
- Zydati, W. (2002). Konzeptuelle Grundlagen einer eigenständigen Didaktik des bilingualen Sachfachunterrichts: Forschungsstand und Forschungsprogramm. In S. Breidbach, G. Bach, & D. Wolff, *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie* (S. 31-62). Frankfurt am Main: Peter Lang.
- Zydati, W. (2007). *Deutsch-Englische Züge in Berlin (DEZIBEL). Eine Evaluation des bilingualen Sachfachunterrichts an Gymnasien*. Frankfurt am Main: Peter Lang.
- Zydati, W. (2012). Linguistic Thresholds in the CLIL Classroom? The Threshold Hypothesis Revisited. *International CLIL Research Journal*, S. 17-28.
- Zydati, W. (2013). Kompetenzerwerb im Bilingualen Unterricht. In W. Hallet, & F. Königs, *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (S. 131-137). Seelze: Friedrich.

Anhang

Der Verlauf der Unterrichtsreihe

Stunde	Thema, Inhalte und Ergebnis	Material, Hausaufgabe
1	The reaction of sodium and water Elementares Natrium wird vorgestellt, geschnitten und zu Wasser gegeben, dem vorher Phenolphthalein zugesetzt wurde. Ergebnis: Eigenschaften des Natriums und sein Verhalten in Wasser	TB1: Sodium AB1: Sodium and water (HA)
2	Identifying the products of the reaction of sodium with water Natrium wird unter Wasser fixiert und das entweichende Gas pneumatisch aufgefangen, anschließend wird damit die Knallgasprobe durchgeführt. Ergebnis: Natrium reagiert mit Wasser zu Wasserstoff und gelöstem Natriumhydroxid (Natronlauge)	TB2: The reaction of sodium and water AB2: Sodium plus water - identifying one product of the reaction (HA)
3	Heating sodium Elementares Natrium wird auf einer Magnesiumrinne verbrannt, der entweichende Rauch wird in einem Becherglas aufgefangen. Anschließend werden dem Rauch Wasser und Phenolphthalein zugesetzt Ergebnis: Natrium verbrennt mit gelber Flamme zu festem Natriumoxid. In Wasser ist dieses gut löslich und reagiert weiter zu Natriumhydroxid.	AB3: Burning sodium (HA)
4	Sodium, sodium oxide and sodium hydroxide Der Versuch aus der Vorstunde wird mit Hilfe von AB3 ausgewertet. Anschließend tragen die Schüler ihr bisher erworbenes Wissen zu Natrium und seinen Verbindungen Natriumoxid und Natriumhydroxid in einer Tabelle zusammen Ergebnis: Natrium reagiert heftig mit Wasser zu Natriumhydroxid, an der Luft verbrennt es zu Natriumoxid, welches in Wasser zu Natriumhydroxid weiterreagiert	TB3: Burning sodium TB4: Sodium, sodium oxide and sodium hydroxide
5	Lithium, sodium and potassium Lithium, Natrium und Kalium werden vorgestellt, geschnitten und zu Wasser, dem Phenolphthalein zugesetzt wurde, gegeben. Ihre Eigenschaften und ihr Reaktionsverhalten in Wasser werden verglichen. Ergebnis: Alle drei Elemente sind weiche, reaktive Metalle, die mit Wasser zu einem Hydroxid und Wasserstoff reagieren. Lithium reagiert weniger heftig und Kalium noch heftiger als Natrium.	TB5: Lithium, sodium and potassium AB4: Comparing lithium, sodium and potassium (HA)
6	The alkali metals (1st main group) Anhand von AB5 werden die restlichen Alkalimetalle vorgestellt. Anschließend werden Stoffeigenschaften sowie die Reaktionen mit Wasser und Sauerstoff verglichen und Regeln bzgl. der Abstufung physikalischer und chemischer Eigenschaften erarbeitet Ergebnis: Die Dichte und die Reaktivität der AM steigen innerhalb der Gruppe, Härte und Schmelzpunkt sinken. Die Reaktionen mit Wasser liefern immer Hydroxide und Wasserstoff, bei der Reaktion mit Sauerstoff entstehen Oxide, die mit Wasser zu Hydroxiden weiterreagieren.	TB6: The alkali metals (1st main group) AB5: Properties of alkali metals (ggf. HA)

Anhang

7+8	The reaction of calcium with water Schülerversuch: Die Schüler geben Calciumspäne zu Wasser, führen mit dem entweichenden Gas die Knallgasprobe durch und dampfen die entstehende Suspension ein. Anschließend setzten sie dem weißen Rückstand Phenolphthalein zu. Die Ergebnisse werden anhand von AB6 gesammelt und im Plenum vorgestellt und ausgewertet. Ergebnis: Calcium reagiert ähnlich wie die AM mit Wasser zu Wasserstoff und Calciumhydroxid. Dieses ist weniger gut wasserlöslich als die Alkalimetallhydroxide und liefert in Wasser eine Lauge.	AB6: The reaction of calcium and water TB7: The reaction of calcium and water AB7: Results of the reaction of calcium and water (HA)
9	Calcium hydroxide and lime water Frisches Kalkwasser wird durch Zugabe von Calciumoxid zu Wasser und Filtern hergestellt. Anschließend wird die Kalkwasserprobe mit der Atemluft der Schüler durchgeführt. Ergebnis: Calciumoxid reagiert in Wasser zu Calciumhydroxid. Die wässrige Lösung ist Kalkwasser.	TB8: The production of fresh lime water
10	Alkaline earth metals (2nd main group) Anhand von AB7, Aufgabe 3 werden die Eigenschaften und das Reaktionsverhalten von Natrium und Calcium verglichen. Anschließend werden die Ergebnisse bzgl. der Erdalkalimetalle verallgemeinert und anhand von AB8 gefestigt. Ergebnis: Die Erdalkalimetalle sind härter und von höherer Dichte als die AM. Mit Wasser und Sauerstoff reagieren sie wie die AM, sind jedoch weniger reaktiv. Die Hydroxide sind schwer wasserlöslich.	TB9: Alkaline earth metals (2nd main group) AB8: Properties of the alkaline earth metals

Fragebogen vor Beginn des bilingualen Unterrichts im Fach Chemie

Hilfsmittel: Bleistift, Spitzer, Radiergummi

Die Auswertung des Fragebogens erfolgt anonym. Deshalb benötigst du einen Code-Namen.

Deinen Code-Namen erhältst du, wenn du

- die **zwei Anfangsbuchstaben** des **Vornamens** deines **Vaters**, falls unbekannt setze „DI“
- die **zwei Anfangsbuchstaben** des **Vornamens** deiner **Mutter**, falls unbekannt setze „TA“
- die ersten **vier Ziffern** deines **Geburtstags** (ohne Jahresangabe)
- deine **Klasse**

zusammenfügst.

So zum Beispiel:

Vorname Vater: KLAUS, Vorname MUTTER: GISELA, Geburtstag: 05.04.1998, Klasse 7b

CODE: **KL GI 0504 7b**

Mein Code-Name lautet:

Datum:

Anhang

	Stimmt nicht O	Stimmt wenig O	Stimmt eher O	Stimmt völlig O	Ich weiß es nicht O
Der CAD-Unterricht interessiert mich nicht.	O	O	O	O	O
Das Fach Englisch macht mir Spaß.	O	O	O	O	O
Ich finde CAE interessant.	O	O	O	O	O
Was ich in CAE lerne, brauche ich nicht wirklich.	O	O	O	O	O
Ich werde in CAE keine Probleme haben die Dinge zu verstehen.	O	O	O	O	O
CAD ist sehr schwer für mich.	O	O	O	O	O
Den Unterricht in Englisch kann ich gut verstehen	O	O	O	O	O
Ich mache mir Sorgen, dass CAE anstrengend für mich wird.	O	O	O	O	O

Teil 2

Persönliche Angaben

Bitte ergänze noch ein paar persönliche Angaben. Kreuze dazu das entsprechende Kästchen an.

Ich bin	☐ ein Junge	☐ ein Mädchen	☐ keine Angabe
Ich bin	☐ 12 Jahre	☐ 13 Jahre	☐ 14 Jahre
Chemienote (Klasse 7)	_____		
Englischnote (Klasse 7)	_____		

Welche der vier Aussagen trifft am ehesten auf dich zu? Kreuze an!

Ich mag Naturwissenschaften

☐ lieber als Sprachen
☐ nicht so gerne wie Sprachen
☐ genauso gerne wie Sprachen
☐ genauso wenig wie Sprachen

Danke für deine Mithilfe!

Fragebogen nach Ende des bilingualen Unterrichts im Fach Chemie

Hilfsmittel:

Bleistift, Spitzer, Radiergummi

Die Auswertung des Fragebogens erfolgt anonym. Deshalb benötigst du einen Code-Namen.

Deinen Code-Namen erhältst du, wenn du

- die **zwei Anfangsbuchstaben** des **Vornamens** deines **Vaters**, falls unbekannt setze „DI“
- die **zwei Anfangsbuchstaben** des **Vornamens** deiner **Mutter**, falls unbekannt setze „TA“
- die ersten **vier Ziffern** deines **Geburtstags** (ohne Jahresangabe)
- deine **Klasse**

zusammenfügst.

So zum Beispiel:

Vorname Vater: KLAUS, Vorname MUTTER: GISELA, Geburtstag: 0504.1998, Klasse 7b

CODE: **KL GI 0504 7b**

Mein Code-Name lautet:

Datum:

Code _____

Datum _____

Teil 1

Hier würden wir gerne etwas über deine Einschätzungen zu **Chemie auf Englisch (CAE)** erfahren. Versuche alle Aufgaben zu beantworten. Falls du dich absolut nicht entscheiden kannst, kreuz bitte „Ich weiß es nicht“ an.

	Stimmt nicht	Stimmt wenig	Stimmt eher	Stimmt völlig	Ich weiß es nicht
	☐	☐	☐	☐	☐
	1	2	3	4	

Gib an, inwiefern folgende Aussagen auf einer Skala von 1 bis 4 für dich zutreffen.

Beispiel

Falls ihr denkt, dass es wichtig ist, Chemie auf Englisch zu lernen, dann schreibt ihr:

CAE zu lernen ist wichtig.	☐	☐	☐	☐	☐
----------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Was denkst du über CAE?

1.					
CAE hat mir Spaß gemacht	☐	☐	☐	☐	☐
In CAE haben wir nützliche Dinge gelernt.	☐	☐	☐	☐	☐
CAE war sehr schwer für mich.	☐	☐	☐	☐	☐
CAE war nicht so stressig.	☐	☐	☐	☐	☐
Ich fand CAE interessant.	☐	☐	☐	☐	☐
In CAE habe ich sinnvolle Dinge für meine Zukunft gelernt.	☐	☐	☐	☐	☐
In CAE konnte ich dem Unterricht nicht folgen.	☐	☐	☐	☐	☐
CAE war relativ locker.	☐	☐	☐	☐	☐

2.					
CAE fand ich nicht interessant.	☐	☐	☐	☐	☐
Den Unterricht in CAE konnte ich gut verstehen	☐	☐	☐	☐	☐
CAE war anstrengend.	☐	☐	☐	☐	☐
CAE war langweilig.	☐	☐	☐	☐	☐
Was ich in CAE gelernt habe, brauche ich nicht wirklich.	☐	☐	☐	☐	☐
In CAE fand ich den Unterricht nicht schwer.	☐	☐	☐	☐	☐
Ich fand CAE anstrengend.	☐	☐	☐	☐	☐
CAE zu machen war Zeitverschwendung.	☐	☐	☐	☐	☐

Teil 2

Persönliche Angaben

Bitte ergänze noch ein paar persönliche Angaben. Kreuze dazu das entsprechende Kästchen an.

Ich bekomme Nachhilfe in Chemie	☐ ja	☐ nein
Mein Vater und/oder meine Mutter haben an einer Hochschule studiert.	☐ ja	☐ nein
Mein Vater und meine Mutter sind im Ausland aufgewachsen	☐ ja	☐ nein

Zuletzt ist noch deine Meinung gefragt. Bitte beantworte folgende Frage in einem Text:

*Sollte man Unterrichtsreihen auf Englisch im Chemieunterricht häufiger machen?
Bitte begründe deine Meinung möglichst ausführlich.*

Danke für deine Mitarbeit!

Fragebogen vor Beginn des Modulunterrichts im Fach Chemie

Hilfsmittel: **Bleistift, Spitzer, Radiergummi**

Die Auswertung des Fragebogens erfolgt anonym. Deshalb benötigst du einen Code-Namen.

Deinen Code-Namen erhältst du, wenn du

- die **zwei Anfangsbuchstaben** des **Vornamens** deines **Vaters**, falls unbekannt setze „DI“
- die **zwei Anfangsbuchstaben** des **Vornamens** deiner **Mutter**, falls unbekannt setze „TA“
- die ersten **vier Ziffern** deines **Geburtstags** (ohne Jahresangabe)
- deine **Klasse**

zusammenfügst.

So zum Beispiel:

Vorname Vater: KLAUS, Vorname MUTTER: GISELA, Geburtstag: 05.04.1998, Klasse 7b

CODE: KL GI 0504 7b

Mein Code-Name lautet:

Datum:

Code _____

Datum _____

Teil 1

Hier würden wir gerne etwas über die Fächer **Chemie** und **Englisch** erfahren. Versuche alle Aufgaben zu beantworten. Falls du dich absolut nicht entscheiden kannst, kreuz bitte „Ich weiß es nicht“ an.

	Stimmt nicht	Stimmt wenig	Stimmt eher	Stimmt völlig	Ich weiß es nicht
	☐	☐	☐	☐	☐
	1	2	3	4	

Gib an, inwiefern folgende Aussagen auf einer Skala von 1 bis 4 für dich zutreffen.

Beispiel

Falls ihr denkt, dass es wichtig ist, Englisch sprechen zu können, dann schreibt ihr:

Englischsprechen zu können ist wichtig. ☐ ☐ ☐ ☐ ☐

Was denkst du über die Fächer Chemie und Englisch?

1.					
Das Fach Chemie interessiert mich auch nach der Schule.	☐	☐	☐	☐	☐
Das Fach Englisch macht mir Spaß.	☐	☐	☐	☐	☐
In Chemie lernen wir nützliche Dinge.	☐	☐	☐	☐	☐
Chemie ist sehr schwer für mich.	☐	☐	☐	☐	☐
In Englisch kann ich dem Unterricht nicht folgen.	☐	☐	☐	☐	☐
Chemie ist nicht so stressig für mich.	☐	☐	☐	☐	☐
Der Chemieunterricht interessiert mich nicht.	☐	☐	☐	☐	☐
In Chemie habe ich keine Probleme die Dinge zu verstehen.	☐	☐	☐	☐	☐
2.					
Ich finde das Fach Chemie interessant.	☐	☐	☐	☐	☐
Mit dem Fach Englisch beschäftige ich mich gerne.	☐	☐	☐	☐	☐
Chemie zu lernen ist sinnvoll für meine Zukunft.	☐	☐	☐	☐	☐
In Chemie kann ich dem Unterricht nicht folgen.	☐	☐	☐	☐	☐
Englisch ist sehr schwer für mich.	☐	☐	☐	☐	☐
Chemie ist anstrengend.	☐	☐	☐	☐	☐
Chemie zu machen ist Zeitverschwendung.	☐	☐	☐	☐	☐
Den Unterricht in Englisch kann ich gut verstehen.	☐	☐	☐	☐	☐
3.					
Sich mit dem Fach Chemie befassen ist langweilig.	☐	☐	☐	☐	☐
Chemie ist relativ locker.	☐	☐	☐	☐	☐
Ich mag das Fach Englisch nicht.	☐	☐	☐	☐	☐
Was ich in Chemie lerne, brauche ich nicht wirklich.	☐	☐	☐	☐	☐
Den Unterricht in Chemie kann ich gut verstehen.	☐	☐	☐	☐	☐
In Englisch habe ich keine Probleme die Dinge zu verstehen.	☐	☐	☐	☐	☐
Im Fach Englisch zu lernen ist langweilig.	☐	☐	☐	☐	☐
Für mich ist Chemie anstrengend.	☐	☐	☐	☐	☐

Teil 2

Persönliche Angaben

Bitte ergänze noch ein paar persönliche Angaben. Kreuze dazu das entsprechende Kästchen an.

Ich bin	☐ ein Junge	☐ ein Mädchen	☐ keine Angabe
Ich bin	☐ 12 Jahre	☐ 13 Jahre	☐ 14 Jahre
Chemienote (Klasse 7)	☐		
Englischnote (Klasse 7)	☐		

Welche der vier Aussagen trifft am ehesten auf dich zu? Kreuze an!

Ich mag Naturwissenschaften	☐ lieber als Sprachen
	☐ nicht so gerne wie Sprachen
	☐ genauso gerne wie Sprachen
	☐ genauso wenig wie Sprachen

Danke für deine Mithilfe!

Fragebogen nach Ende des Moduls im Fach Chemie

Hilfsmittel: Bleistift, Spitzer, Radiergummi

Die Auswertung des Fragebogens erfolgt anonym. Deshalb benötigst du einen Code-Namen.

Deinen Code-Namen erhältst du, wenn du

- die **zwei Anfangsbuchstaben** des **Vornamens** deines **Vaters**, falls unbekannt setze „DI“
- die **zwei Anfangsbuchstaben** des **Vornamens** deiner **Mutter**, falls unbekannt setze „TA“
- die ersten **vier Ziffern** deines **Geburtstags** (ohne Jahresangabe)
- deine **Klasse**

zusammenfügst.

So zum Beispiel:

Vorname Vater: KLAUS, Vorname MUTTER: GISELA, Geburtstag: 05.04.1998, Klasse 7b

CODE: KL GI 0504 7b

Mein Code-Name lautet:

Datum:

Code _____

Datum _____

Teil 1

Hier würden wir gerne etwas über deine Einschätzungen zum **Modul** erfahren. Versuche alle Aufgaben zu beantworten. Falls du dich absolut nicht entscheiden kannst, kreuz bitte „Ich weiß es nicht“ an.

	Stimmt nicht	Stimmt wenig	Stimmt eher	Stimmt völlig	Ich weiß es nicht
	☐	☐	☐	☐	☐
	1	2	3	4	

Gib an, inwiefern folgende Aussagen auf einer Skala von 1 bis 4 für dich zutreffen.

Beispiel

Falls ihr denkt, dass es wichtig ist, Chemie zu lernen, dann schreibt ihr:

Chemie zu lernen ist wichtig.

☐ ☐ ☐ ☒ ☐

Was denkst du über das Modul?

1.					
Im Modul haben wir nützliche Dinge gelernt.	☐	☐	☐	☐	☐
Das Modul war sehr schwer für mich.	☐	☐	☐	☐	☐
Ich fand das Modul interessant.	☐	☐	☐	☐	☐
Das Modul war nicht so stressig für mich.	☐	☐	☐	☐	☐
Das Modul war sinnvoll für meine Zukunft.	☐	☐	☐	☐	☐
In dem Modul konnte ich dem Unterricht kaum folgen.	☐	☐	☐	☐	☐
Das Modul war relativ locker.	☐	☐	☐	☐	☐
Das Modul hat mich auch nach der Schule interessiert.	☐	☐	☐	☐	☐
2.					
Was ich in dem Modul gelernt habe, brauche ich nicht wirklich.	☐	☐	☐	☐	☐
Sich mit dem Modul zu befassen war langweilig.	☐	☐	☐	☐	☐
Den Unterricht im Modul habe ich gut verstanden.	☐	☐	☐	☐	☐
Das Modul war anstrengend.	☐	☐	☐	☐	☐
Das Modul war Zeitverschwendung.	☐	☐	☐	☐	☐
Im Modul hatte ich keine Probleme die Dinge zu verstehen.	☐	☐	☐	☐	☐
Für mich war das Modul anstrengend.	☐	☐	☐	☐	☐
Das Modul hat mich nicht interessiert.	☐	☐	☐	☐	☐

Teil 2

Persönliche Angaben

Bitte ergänze noch ein paar persönliche Angaben. Kreuze dazu das entsprechende Kästchen an.

Ich bekomme Nachhilfe in Chemie	☐ ja	☐ nein
Mein Vater und/oder meine Mutter haben an einer Hochschule studiert.	☐ ja	☐ nein
Mein Vater und meine Mutter sind im Ausland aufgewachsen.	☐ ja	☐ nein

Zuletzt ist noch deine Meinung gefragt. Bitte beantworte folgende Frage in einem Text:

*War der Unterricht im Modul anders als der Chemieunterricht, den du sonst gewohnt bist?
Bitte begründe deine Meinung möglichst ausführlich.*

Danke für deine Mitarbeit!

Test zum Umgang mit chemischem Fachwissen

Teil 1-3

Hilfsmittel:

Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Teil 1

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

Wie wird Natrium (Na, sodium) aufbewahrt?

A unter Paraffinöl B in einer Ampulle C unter Spiritus D unter Schutzgas

Hier ist die **Antwort A** richtig, also würde das Kreuz auf dem Antwortbogen bei **A** gemacht:

Nr.	A	B	C	D
Beispiel	X			

Aufgaben

1. Welche Beschaffenheit hat Natrium (Na, sodium)?

A hart wie Eisen B zähflüssig C weich wie Wachs D spröde

2. Was gilt für die Dichte (density) von Natrium im Vergleich zur Dichte von Wasser?

A etwas größer B kleiner C gleich groß D sehr viel größer

3. Wie verhält sich ein Natriumstück (Na, piece of sodium), das vorsichtig auf eine Wasseroberfläche gelegt wird?

A beginnt zu zittern
B bewegt sich auf der Oberfläche hin und her
C sinkt zu Boden und steigt dann wieder auf
D wird schwarz

4. Wie verändert sich die Form eines Natriumstücks (Na, piece of sodium), das auf eine Wasseroberfläche gelegt wird?

A dehnt sich aus B wird zur Kugel C zerfällt in Stücke D zieht Fäden

5. Erdalkalimetalle (alkaline earth metals) sind im Vergleich zu Alkalimetallen (alkaline metals) gleicher Periode

A härter B viel weicher C etwas weicher D gleich hart

6. Erdalkalimetalle (alkaline earth metals) sind im Vergleich zu Alkalimetallen (alkaline metals) gleicher Periode

A viel reaktiver B weniger reaktiv C gleich reaktiv D etwas reaktiver

7. Die Dichte (density) von Alkalimetallen (alkaline metals) ist im Vergleich zur Dichte (density) der Erdalkalimetalle (alkaline earth metals) gleicher Periode

A viel größer B etwas größer C gleich D geringer

.....

Teil 2

Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

Unter welcher Flüssigkeit wird Natrium aufbewahrt?

Paraffinöl

Hier ist das Lösungswort **Paraffinöl**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>Paraffinöl</i>

Aufgaben

8. Beschreibe eine charakteristische Stoffeigenschaft der Alkalimetalle (alkali metals).

9. Beschreibe eine zweite charakteristische Stoffeigenschaft der Alkalimetalle (alkali metals).

10. Notiere einen Reaktionspartner, mit dem Erdalkalimetalle (alkaline earth metals) besonders gut reagieren.

11. Notiere einen weiteren Reaktionspartner, mit dem Erdalkalimetalle (alkaline earth metals) besonders gut reagieren.

12. Notiere ein Reaktionsprodukt (Name oder Formel), welches leicht aus Alkalimetallen (alkali metals) gebildet werden kann.

13. Notiere ein weiteres Reaktionsprodukt (Name oder Formel), welches leicht aus Alkalimetallen (alkali metals) gebildet werden kann.

.....

Teil 3

Aufgabenstellung

Beantworte die Frage in **einem oder mehreren Sätzen** und schreibe diese(n) in das **entsprechende Kästchen auf dem Antwortbogen**.

14. Kalkwasser (lime water) ist an der Luft schlecht haltbar. Woran liegt das?

Test zum Umgang mit dem chemischen Fachwissen

Teil 4-6

Hilfsmittel:

Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Aufgabenstellung

Bitte hör genau zu, wenn dein Lehrer den folgenden Text vorliest. Was hier steht, kannst du sicher gut gebrauchen, wenn du die nächsten Aufgaben löst. Du kannst natürlich während des Tests immer mal wieder zurückblättern und nochmal nachlesen, was hier steht.

Wichtige Hintergrundinformationen zum nächsten Teil des Tests!

Die **Alkalimetalle** (alkali metals) (**AM**, 1. Hauptgruppe) sind weiche, äußerst reaktive Metalle. Sie reagieren besonders heftig mit Sauerstoff (oxygen) oder Wasser. Dabei entstehen Oxide (AM_2O) oder Hydroxide (AMOH). An der Luft laufen sie schnell an, da sie mit dem Sauerstoff (oxygen) und dem Wasser in der Luft reagieren. Oft werden sie deshalb unter einer Schutzflüssigkeit (Paraffinöl) aufbewahrt.

Mit Wasser reagieren sie besonders heftig, wobei Wasserstoff (hydrogen) und eine alkalische (alkaline) Lösung des Hydroxids (AMOH) entstehen. Durch die Knallgasprobe (pop test) lässt sich der Wasserstoff (hydrogen) und durch die Pinkfärbung von zugesetztem Phenolphthalein die alkalische Lösung nachweisen. Die Reaktionen sind stark exotherm. Während Natrium (sodium) dabei schmilzt, entzündet sich Kalium (potassium) sogar. Der entweichende Wasserstoff (hydrogen) brennt dann.

Bei der **Verbrennung von Alkalimetallen** entsteht das Oxid als weißer, gut wasserlöslicher Feststoff, der als Rauch entweicht. In Wasser reagieren das Oxid (AM_2O) weiter zum entsprechenden Hydroxid (AMOH).

Es gelten folgende Regeln:

1. Wässrige Lösungen von Hydroxiden sind alkalisch.
2. Wässrige Lösungen von Metalloxiden sind alkalisch, weil diese in Wasser zu Metallhydroxiden reagieren.

Die **Erdalkalimetalle** (alkaline earth metals) (**EAM**, 2. Hauptgruppe) sind weniger reaktiv, härter und von höherer Dichte (density) als die Alkalimetalle (alkali metals) gleicher Periode. In Wasser liefern die Erdalkalimetalle schwerlösliche Hydroxide ($\text{EAM}(\text{OH})_2$) sowie Wasserstoff (hydrogen). Bei ihrer Verbrennung entstehen Metalloxide (EAMO).

Gibt man **Calciumstücke zu Wasser** so entsteht schwerlösliches Calciumhydroxid, welches eine milchig-weiße Trübung hervorruft. Die Suspension ist alkalisch (Kalkmilch). Außerdem entweicht Wasserstoff (hydrogen). Durch Filtrieren der Kalkmilch erhält man Kalkwasser (lime water), eine wässrige Lösung von Calciumhydroxid. Sie dient zum Nachweis von Kohlenstoffdioxid.

An der Luft ist Kalkwasser schlecht haltbar, da es mit dem Kohlenstoffdioxid reagiert und sich die Lösung dadurch trübt. Frisches Kalkwasser kann deshalb z.B. durch Zugabe von Calciumoxid zu Wasser hergestellt werden, da dieses in Wasser zu Calciumhydroxid reagiert.

Die **Reaktivitäten** der Alkalimetalle (alkali metals) und der Erdalkalimetalle (alkaline earth metals) nehmen in der Gruppe von oben nach unten zu. Auch andere **Eigenschaften** ändern sich in der Gruppe systematisch. So steigt die **Dichte** (density) von oben nach unten, während die **Härte** abnimmt.

	1. Hauptgruppe		2. Hauptgruppe	
	<u>Alkalimetalle</u>		<u>Erdalkalimetalle</u>	
1.	Li	Lithium	Be	Beryllium
2.	Na	Natrium	Mg	Magnesium
3.	K	Kalium	Ca	Calcium
4.	Rb	Rubidium	Sr	Strontium
5.	Cs	Caesium	Ba	Barium

Teil 4Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

Wie wird Natrium (Na, sodium) aufbewahrt?

A unter Paraffinöl

B in einer Ampulle

C unter Spiritus

D unter Schutzgas

Hier ist die **Antwort A** richtig, also würde das Kreuz auf dem Antwortbogen bei **A** gemacht:

Nr.	A	B	C	D
Beispiel	X			

Aufgaben

15. Welche Formel hat Berylliumhydroxid?

A $\text{Be}_2(\text{OH})_2$

B $\text{Be}(\text{OH})_2$

C BeOH

D Be_2OH

16. Welche Formel hat Magnesiumoxid?

A MgO

B Mg_2O

C MgO_2

D Mg_2O_2

17. Gibt man elementares Strontium zu Wasser so wird die Lösung

A sauer.

B alkalisch.

C neutral.

D schwach sauer.

18. Rubidiumhydroxid ist in Lösung

A stark sauer.

B neutral.

C alkalisch.

D schwach sauer.

19. Gibt man Caesium zu Wasser so verläuft die Reaktion

A schwach endotherm.

B stark exotherm.

C stark endotherm.

D schwach exotherm.

20. Berylliumhydroxid ist

A gut wasserlöslich.

B schwer wasserlöslich.

C wasserlöslich.

D wasseranziehend.

21. Verbrennt man Beryllium, so entsteht ein weißer Rauch. Dieser besteht aus

A Beryllium.

B Berylliumhydrid.

C Berylliumhydroxid.

D Berylliumoxid.

22. Rubidiumoxid ist

A wasserunlöslich.

B schlecht wasserlöslich.

C wasserabweisend.

D wasserlöslich.

23. Rubidium ist im Vergleich zu Lithium

A etwas härter.

B weicher.

C gleich hart.

D wesentlich härter.

24. Rubidium ist im Vergleich zu Lithium

A reaktiver.

B gleich reaktiv.

C deutlich weniger reaktiv.

D etwas weniger reaktiv.

25. Rubidium läuft an der Luft im Vergleich zu Lithium

A schneller an. B gleich schnell an. C etwas langsamer an. D deutlich langsamer an.

26. Magnesium ist im Vergleich zu Barium

A etwas weicher. B härter. C gleich hart. D wesentlich weicher.

27. Beryllium ist im Vergleich zu Barium

A reaktiver. B gleich reaktiv. C deutlich reaktiver. D weniger reaktiv.

28. Die Dichte (density) von Beryllium ist im Vergleich zur Dichte (density) von Strontium

A deutlich höher. B gleich. C geringer. D etwas höher.

29. Beryllium ist im Vergleich zu Caesium

A härter. B etwas weicher. C gleich hart. D wesentlich weicher.

30. Beryllium ist im Vergleich zu Caesium

A weniger reaktiv. B gleich reaktiv. C deutlich reaktiver. D etwas reaktiver.

31. Die Dichte (density) von Barium ist im Vergleich zur Dichte (density) von Rubidium

A etwas geringer. B gleich. C höher. D deutlich geringer.

32. Welche Formel hat Strontiumhydroxid?

A SrOH_2 B $\text{Sr}(\text{OH})_2$ C SrOH D Sr_2OH

33. Die Formel von Rubidiumoxid lautet

A RbO B Rb_2O C RbO_2 D Rb_2O_2

.....

Teil 5

Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

Unter welcher Flüssigkeit wird Natrium aufbewahrt?

Paraffinöl

Hier ist das Lösungswort **Paraffinöl**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>Paraffinöl</i>

Aufgaben

34. Kaliumoxid lässt sich in einem Becher voll Wasser auffangen, weil es _____ ist.

35. Rubidiumoxid liefert in Wasser eine Lauge (alkaline solution), weil es zu _____ reagiert.

36. Formuliere die vollständig eingerichtete Reaktionsgleichung (Formeln und Koeffizienten) für die Verbrennung von Strontium.

37. Formuliere die vollständig eingerichtete Reaktionsgleichung (Formeln und Koeffizienten) für die Reaktion von Rubidiumoxid mit Wasser.

38. Notiere die vollständig eingerichtete Reaktionsgleichung (Formeln und Koeffizienten) für die Verbrennung von Rubidium.

.....

Teil 6

Aufgabenstellung

Beantworte die Frage in **einem oder mehreren Sätzen** und schreibe diese(n) in das **entsprechende Kästchen auf dem Antwortbogen**.

Aufgaben

39. Neben Kalkwasser (lime water) kann auch Barytwasser (wässrige Bariumhydroxidlösung) zum Nachweis für CO_2 verwendet werden. Erläutere, wie Barytwasser, ausgehend von Bariumoxid, hergestellt werden kann. Beschreibe dazu das experimentelle Vorgehen und erkläre den chemischen Hintergrund (mögliche Reaktionen bitte erläutern).

40. Barium wird in Vakuumröhren für Sonnenkollektoren zur Aufrechterhaltung des Vakuums (Unterdruck, luftleerer Raum) verwendet. Das Metall wird als Beschichtung in der Röhre aufgetragen. Ohne diese Beschichtung steigt der Druck im Laufe der Jahre wieder langsam an. Erkläre die Wirkung des Bariums (mögliche Reaktionen bitte erläutern).

41. Strontium zählt zu den gefährlichen Metallen. Es lässt sich schon durch Reiben an einem harten Gegenstand entzünden. Erkläre diese Eigenschaft (mögliche Reaktionen bitte erläutern)!

42. In der Nähe des Freiburger Hauptbahnhofs geriet vor einigen Jahren ein Güterwaggon in Brand, der etwa eine halbe Tonne Natrium (Na, sodium) transportierte. Die Feuerwehr löschte unter Missachtung der Gefahrgutkennzeichnung mit Wasser, wobei eine schwere Explosion auftrat, die einen Feuerwehrmann tötete und eine Brücke beschädigte. Erkläre, wie es zu dem Unfall bei den Löscharbeiten kommen konnte (mögliche Reaktionen bitte erläutern).

**Test zum Umgang mit dem
chemischen Fachwissen
der Klasse 7**

Teil 1-3

Hilfsmittel:

Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Teil 1

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

Aufgrund welcher unterschiedlichen Stoffeigenschaft lassen sich Stoffe durch Sieben trennen?

A Dichte B Siedetemperatur C Löslichkeit **D Partikelgröße**

Hier ist die **Antwort D** richtig, also würde das Kreuz auf dem Antwortbogen bei **D** gemacht:

Nr.	A	B	C	D
Beispiel				X

Aufgaben

- Den Übergang vom festen in den flüssigen Zustand nennt man
A sublimieren. B kondensieren. C erstarren. D schmelzen.
- Ein Stoff wird gasförmig,
A unterhalb der Schmelztemperatur.
B unterhalb der Siedetemperatur.
C oberhalb der Siedetemperatur.
D oberhalb der Schmelztemperatur.
- Welcher der folgenden Begriffe bezeichnet einen Aggregatzustand?
A sauer B wasserlöslich C fest D spröde
- Wenn die Teilchen sich ungeordnet bewegen und große Abstände zueinander haben, dann ist der Stoff:
A flüssig B fest C gasförmig D elastisch
- Der pH-Wert einer alkalischen Lösung ist
A größer als 7. B kleiner als 7. C gleich 7. D gleich 14.
- Stoffe können durch Eindampfen getrennt werden, wenn sie sich in folgender Stoffeigenschaft unterscheiden:
A Dichte B Löslichkeit C Siedetemperatur D Schmelztemperatur
- Kleine Wassertröpfchen in der Luft bilden
A einen Rauch. B einen Nebel. C eine Lösung. D eine Emulsion.

Teil 2

Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

Aufgrund ihrer unterschiedlichen _____ lassen sich Stoffe durch Sieben trennen.

Hier ist das Lösungswort **Partikelgröße**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>Partikelgröße</i>

Aufgaben

8. Einen Vorgang, bei dem aus Ausgangsstoffen (Edukten) neue Stoffe (Produkte) entstehen, nennt man _____.
9. Zur Identifizierung eines Stoffes untersucht man seine _____.
10. Wird Energie aus der Umgebung aufgenommen, so nennt man den Vorgang _____.
11. Stoffe, die aus einer Atomsorte bestehen und deshalb nicht weiter zerlegt werden können, nennt man _____.
12. Kohlenstoffdioxid ist kein Element, da es in Kohlenstoff und Sauerstoff zerlegt werden kann. Man sagt Kohlenstoffdioxid ist eine _____ aus Kohlenstoff und Sauerstoff.
13. Stoffe wie Kohlenstoffdioxid nennt man auch _____, da sie aus einer Reaktion mit Sauerstoff hervorgehen.

Teil 3

Aufgabenstellung

Beantworte die Frage in **einem oder mehreren Sätzen** und schreibe diese(n) in das **entsprechende Kästchen auf dem Antwortbogen**.

Aufgaben

14. Ein Sand-Wasser-Gemisch kann leicht mit Hilfe eines Filters getrennt werden. Beschreibe wie man vorgeht und auf welcher Stoffeigenschaft die Trennung beruht.

**Test zum Umgang mit dem
chemischen Fachwissen
der Klasse 7**

Teil 4-6

Hilfsmittel:

Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Aufgabenstellung

Bitte hör genau zu, wenn dein Lehrer den folgenden Text vorliest. Was hier steht, kannst du sicher gut gebrauchen, wenn du die nächsten Aufgaben löst. Du kannst natürlich während des Tests immer mal wieder zurückblättern und nochmal nachlesen.

Aber, bitte nichts unterstreichen und nichts auf dieses Blatt schreiben.

.....

Wichtige Hintergrundinformationen zum nächsten Teil des Tests!

Stoffe haben typische Eigenschaften. Zu den Eigenschaften gehören **Aggregatzustand, Härte, Dichte, Siedetemperatur, Schmelztemperatur und Löslichkeit**. Ein Stoff kann fest, flüssig oder gasförmig sein. Wird ein Feststoff flüssig, so nennt man das **schmelzen**, der umgekehrte Vorgang heißt **erstarren**. Wird eine Flüssigkeit gasförmig so spricht man von **sieden**, der umgekehrte Vorgang heißt **kondensieren**.

Alle Stoffe bestehen aus **kleinsten Teilchen**, die sich gegenseitig anziehen. Im festen Zustand liegen sie eng beieinander. In der Flüssigkeit bewegen sie sich aneinander vorbei und im gasförmigen Zustand bewegen sie sich ungeordnet und mit noch höherer Geschwindigkeit durch den ganzen Raum. Hier sind die Teilchen weit auseinander. Je höher die Temperatur, desto schneller bewegen sich die Teilchen. Ist die Temperatur hoch genug, können die Anziehungskräfte überwunden werden und der Stoff wird flüssig. Die Temperatur heißt **Schmelztemperatur**. Ist die Temperatur noch höher lösen sich die Teilchen vollständig voneinander. Die Temperatur heißt **Siedetemperatur**.

Feststoffe können außerdem eine unterschiedliche **Härte** haben, d.h. es gibt sehr weiche und sehr harte Stoffe. Unter der **Dichte** versteht man die Masse eines Kubikzentimeters bzw. eines Milliliters eines Stoffes. Ist die Dichte eines Stoffes größer als die Dichte des Stoffes in der Umgebung, so sinkt dieser zu Boden. Im umgekehrten Fall steigt er nach oben. Ein Stoff löst sich in einer Flüssigkeit, wenn er sich in dieser sehr fein verteilt. Stoffe, die dies nicht tun, nennt man **unlöslich**.

Stoffe können als Gemisch oder als Reinstoff vorliegen. Ein **Gemisch** lässt sich mit verschiedenen **Trennverfahren** wieder trennen. Dazu gehören **Sieben, Filtrieren, Eindampfen und Destillieren**. Beim Sieben und Filtrieren gelingt die Trennung, falls die Partikel unterschiedlich groß sind. Beim Eindampfen und Destillieren kommt es auf unterschiedliche Siedetemperaturen der Stoffe an.

Gemische, bei denen man die Bestandteile mit bloßem Auge nicht mehr erkennen kann, heißen **homogen**.

Gemische, deren Bestandteile mit dem Auge zu erkennen sind, nennt man **heterogen**. Ein Gemisch aus verschiedenen Gasen, nennt man **Gasgemisch**. Ist ein Stoff in einer Flüssigkeit fein verteilt, so spricht man von einer Lösung. Ein unlöslicher Feststoff liefert in einer Flüssigkeit eine **Suspension**, während man bei einer unlöslichen Flüssigkeit in einer anderen Flüssigkeit von **Emulsion** spricht. Ein **Rauch** liegt vor, wenn ein Feststoff in Form kleiner Stückchen in einem Gas verteilt ist. Bei einer Flüssigkeit, die in Form kleiner Tröpfchen in einem Gas verteilt ist, spricht man von **Nebel**. Rauch und Nebel sind heterogene Gemische. Lösungen und Gasgemische sind homogen. Ein heterogenes Feststoffgemisch heißt **Gemenge**.

Eine chemische Reaktion mit Sauerstoff nennt man Oxidation, dabei entstehen **Oxide**. Bei einer Reaktion mit Schwefel entstehen **Sulfide**. Verläuft eine Oxidation unter Flammenbildung so spricht man von einer **Verbrennung**. Eine solche Reaktion, bei der Energie an die Umgebung abgegeben wird, nennt man **exotherm**. Die Verbrennung ist umso heftiger, je feiner zerteilt der Stoff ist. Den Stoffumsatz gibt man mit einem **Reaktionsschema** an. Man notiert die Namen der Ausgangsstoffe und fügt dazwischen ein Pluszeichen ein. Anschließend folgt ein Pfeil und auf der rechten Seite des Pfeils die Namen der Reaktionsprodukte, ebenfalls mit einem Pluszeichen getrennt.

Stoffe, die sehr reaktiv sind, nennt man **unedel**, wenig reaktive Stoffe heißen **edel**. Bei den Metallen sind Aluminium, Magnesium, Zink und Eisen besonders unedel. Kupfer, Silber und Gold sind hingegen edel. Eine Reaktion, bei der Sauerstoff von einem Stoff auf den anderen übertragen wird, heißt **Redoxreaktion**. Der Stoff, der Sauerstoff aufnimmt wird dabei oxidiert, der Stoff der Sauerstoff abgibt wird reduziert. Ein Stoff, der dem Partner den Sauerstoff entreißen kann, muss unedler sein als dieser. Ansonsten ist die Reaktion nicht möglich.

.....

Teil 4

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

Aufgrund welcher unterschiedlichen Stoffeigenschaft lassen sich Stoffe durch Sieben trennen?

A Dichte B Siedetemperatur C Löslichkeit **D Partikelgröße**

Hier ist die **Antwort D** richtig, also würde das Kreuz auf dem Antwortbogen bei **D** gemacht:

Nr.	A	B	C	D
Beispiel				X

Aufgaben

15. Bei welcher Abfolge steigt die Entzündbarkeit des Metalls von links nach rechts?

- A Draht - Blech - Pulver
- B Pulver - Blech - Draht
- C Blech - Draht - Pulver
- D Pulver - Draht - Blech

16. Was beobachtet man beim Erhitzen von Eisenoxid und Zink?

- A Eisen und Zink entstehen
- B Eisen und Zinkoxid entstehen
- C Eisen, Zink und Sauerstoff entstehen
- D keine Reaktion

17. Welche Stoffe müssen eingesetzt werden, um durch eine Redoxreaktion Blei und Magnesiumoxid zu gewinnen?

- A Blei und Magnesium
- B Bleioxid und Magnesiumoxid
- C Bleioxid und Magnesium
- D Blei und Magnesiumoxid

18. Welche Reaktion ist ein Beispiel für eine Reduktion?

- A Silber reagiert mit Sauerstoff zu Silberoxid
- B Silber reagiert mit Schwefel zu Silbersulfid
- C Silberoxid reagiert zu Silber und Sauerstoff
- D Silberoxid reagiert mit Magnesium zu Magnesiumoxid und Silber

19. Bei der Verbrennung von Aluminium entsteht ein weißer Rauch. Dieser besteht aus

- A Aluminium.
- B Sauerstoff.
- C Aluminiumoxid.
- D Aluminiumsulfid.

20. Warum lässt sich Nagellack nicht mit Wasser abwaschen?

- A Wasser und Nagellack reagieren miteinander.
- B Wasser und Nagellack vermischen sich.
- C Nagellack löst sich nicht in Wasser.
- D Nagellack nimmt Wasser auf.

21. Warum ist Regenwasser sauberer als Flusswasser?

- A die Schadstoffe werden beim Regnen von der Luft aufgenommen
- B die Schadstoffe bleiben beim Verdunsten im Flusswasser zurück
- C die Schadstoffe bleiben beim Regnen in der Wolke zurück
- D die Schadstoffe reagieren beim Verdunsten mit dem Sauerstoff

22. Cola ist

- A eine Emulsion.
- B eine Suspension.
- C eine Lösung.
- D ein Gemenge.

23. Trockene Gartenerde ist

- A eine Lösung.
- B eine Suspension.
- C ein Gemenge.
- D ein Reinstoff.

24. Luft ist ein

- A homogenes Gemisch.
- B heterogenes Gemisch.
- C Reinstoff.
- D Gemenge.

25. Fett schwimmt auf Wasser, weil es

- A eine höhere Dichte hat.
- B gut wasserlöslich ist.
- C eine geringere Dichte hat.
- D eine geringere Siedetemperatur hat.

26. Zucker lässt sich aus Zuckerwasser zurückgewinnen, durch

- A Sieben.
- B Filtrieren.
- C Abkühlen.
- D Erhitzen.

27. Welcher Stoff hat die größte Dichte?

- A Stoff A, hier wiegen 10 Milliliter 4 Gramm.
- B Stoff B, hier wiegen 20 Milliliter 4 Gramm.
- C Stoff C, hier wiegen 20 Milliliter 6 Gramm.
- D Stoff D, hier wiegen 20 Milliliter 10 Gramm.

28. Taucht man einen Luftballon in flüssigen Stickstoff, so kühlt sich die Luft darin stark ab. Als Folge davon wird der Ballon kleiner. Dies liegt daran, dass die Teilchen in der Luft

- A kleiner werden.
- B schneller werden.
- C sich auflösen.
- D näher zusammenrücken.

29. Größere Anziehungskräfte zwischen den Teilchen, haben zur Folge, dass die Siedetemperatur

- A sinkt.
- B steigt.
- C gleich bleibt.
- D Null wird.

30. Ein Stoff, dessen Siedetemperatur 7°C beträgt, ist bei Raumtemperatur (20°C)

- A fest.
- B flüssig.
- C gasförmig.
- D zähflüssig.

31. Ein Stoff, dessen Schmelztemperatur 100°C beträgt, ist bei Raumtemperatur (20°C)

- A fest.
- B flüssig.
- C gasförmig.
- D zähflüssig.

32. Stinkbomben enthalten Flüssigkeiten, die

- A leicht verdunsten.
- B gut wasserlöslich sind.
- C einen hohen Siedepunkt haben.
- D eine geringe Dichte haben.

33. Während sich auf Magnesium an der Luft allmählich eine Schicht aus Magnesiumoxid bilden kann, ist dies bei Kupfer nicht der Fall, da Kupfer
 A schwerer ist. B edler ist. C unedler ist. D weicher ist.

.....

Teil 5

Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende Wort bzw. das Reaktionsschema **auf dem Antwortbogen**.

Beispiel

Aufgrund ihrer unterschiedlichen _____ lassen sich Stoffe durch Sieben trennen.

Hier ist das Lösungswort **Partikelgröße**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>Partikelgröße</i>

Aufgaben

34. Sand sinkt in Wasser auf den Boden, weil es eine höhere _____ hat als das Wasser.

35. Eine Gasflasche, aus der über längere Zeit Gas entnommen wird, kühlt sich stark ab. Nach einigen Minuten bildet sich außen auf der Gasflasche Eis, da das Wasser aus _____ auf der Flasche erstarrt.

36. Notiere das Wortreaktionsschema (Stoffnamen unter Angabe des Aggregatzustands in Klammern, also keine Formeln!) für die Redoxreaktion von Magnesium mit Silberoxid.

37. Notiere das Wortreaktionsschema (Stoffnamen unter Angabe des Aggregatzustands in Klammern, also keine Formeln!) für die Verbrennung von Phosphor.

38. Notiere das Wortreaktionsschema (Stoffnamen unter Angabe des Aggregatzustands in Klammern, also keine Formeln!) für die Zerlegung von Silbersulfid in die Elemente.

.....

Teil 6

Aufgabenstellung

Beantworte die Fragen in **einem oder mehreren Sätzen** und schreibe diese(n) in das entsprechende Kästchen auf dem Antwortbogen.

.....

Aufgaben

39. Mineralwasser enthält neben Wasser noch feste Mineralstoffe, die im Wasser gelöst sind. Beschreibe, wie man die Mineralstoffe vom Wasser abtrennen kann. Erkläre auch, worauf das Trennverfahren beruht.

40. Eine traditionelle Glühbirne leuchtet, weil ein Metalldraht im Inneren sich so stark erhitzt, dass er hell aufglüht. Die Birnen enthalten außerdem ein Schutzgas wie z.B. Argon, das verhindert, dass Luft in das Innere strömt. Ist die Birne undicht, so reißt der Faden und die Glühbirne wird unbrauchbar. Erkläre, warum der Faden reißt! Erläutere auch chemische Reaktionen, die ablaufen.

41. Weißer Phosphor wird unter Wasser aufbewahrt, da er sich an der Luft spontan entzündet und unter großer Hitzeentwicklung verbrennt. Holzkohle hingegen entzündet sich nicht spontan an der Luft. Erkläre das unterschiedliche Verhalten der beiden Stoffe und gib an, welche Reaktionen ablaufen.

42. Wunderkerzen enthalten, neben anderen Bestandteilen, Eisenpulver und eine Verbindung, die - ähnlich wie Kupferoxid - gebundenen Sauerstoff enthält. Erkläre, wieso es beim Anzünden zur Funkenentwicklung kommt. Erläutere auch chemische Reaktionen, die ablaufen.

.....

Wortschatztest zur deutschen Fachsprache

Hilfsmittel: Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Teil 1

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

Es bildete sich eine **Suspension**.

Eine **Suspension** ist ein/e

- A löslicher Feststoff in einer Flüssigkeit
- B unlöslicher Feststoff in einer Flüssigkeit
- C unlösliche Flüssigkeit in einer Flüssigkeit
- D lösliche Flüssigkeit in einer Flüssigkeit

Hier ist die **Antwort B** richtig, also würde das Kreuz auf dem Antwortbogen bei **B** gemacht:

Nr.	A	B	C	D
Beispiel		X		

Aufgaben

1. In der Flasche befindet sich eine **Schutzflüssigkeit**.

Eine **Schutzflüssigkeit** schützt den Inhalt vor:

- A Verdunstung
- B chemischer Zersetzung
- C Reaktion an der Luft
- D Lichteinstrahlung

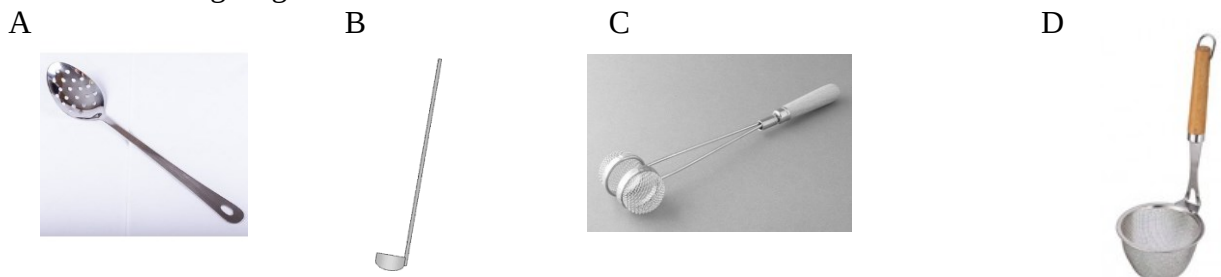
2. Er nahm etwas **Rubidium**.

Rubidium ist ein/e

- A Edelmetall.
- B Nichtmetall.
- C reaktives Metall.
- D Indikator.

3. Für das Experiment war ein **Sieblöffel** nötig.

Diese Abbildung zeigt einen **Sieblöffel** aus dem Labor:



4. Er benötigte mehr **Strontium**.

Strontium ist ein/e

- A Feststoff.
- B Verbindung.
- C Flüssigkeit.
- D Gas.

5. Er untersuchte die **Beryllium**-Probe genau.

Beryllium ist ein/e

- A Kristall.
- B Metall.
- C Mineral.
- D Legierung.

6. Bei dem Versuch blieb **Kalkmilch** übrig.

Kalkmilch ist

- A geschmolzenes Calciumcarbonat.
- B eine Lösung von Calciumnitrat in Wasser.
- C ein Gemisch von Calciumhydroxid und Wasser.
- D eine Emulsion von Calciumsulfat.

7. Bei der Reaktion wurde **Barium** eingesetzt.

Barium ist ein/e

- A Leichtmetall.
- B unedles Metall.
- C Edelmetall.
- D Nichtmetall.

8. Das **Caesium** war aufgebraucht.

Caesium ist ein/e

- A Flüssigkeit.
- B Gas.
- C Feststoff.
- D Gemisch.

9. Für das Experiment wurde eine **Magnesiumrinne** benötigt.

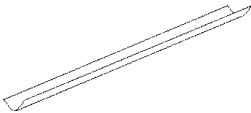
Welches Bild zeigt eine **Magnesiumrinne**?

A

B

C

D



10. Er fügte einige **Granalien** hinzu.

Granalien bedeutet, dass der Stoff vorliegt in Form von:

- A glatten Folien
- B dünnen Drähten
- C kleinen Körnern
- D dicken Plättchen

11. Anschließend kam das **pneumatische Prinzip** zur Anwendung.

Das **pneumatische Prinzip** bezieht sich auf:

- A das Mischen von Stoffen
- B das Lagern von Stoffen
- C das Zerkleinern von Stoffen
- D das Auffangen von Stoffen

Teil 2Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

Bei der Reaktion mit Schwefel können Sulfide entstehen.

(Verbindungen mit Schwefel)

Hier ist das Lösungswort **Sulfide**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>Sulfide</i>

Aufgaben

12. _____ (K) ist ein sehr reaktiver Stoff.

13. _____ (Li) reagiert mit Wasser.

14. _____ (Ca) ist ein auf der Erde häufig anzutreffendes Element.

15. In Wasser liefern _____ (Bezeichnung für Verbindungen die OH enthalten) meist Laugen.

16. Der Rauch enthält festes _____ (Verbindung aus Natrium und Sauerstoff).

17. Die _____ zählen zu den reaktivsten Metallen (Oberbegriff für die Metalle der 1. Hauptgruppe des PSE).

18. Die _____ reagieren leicht an der Luft (Oberbegriff für die Metalle der 2. Hauptgruppe des PSE).

19. Im Labor wird oft _____ (andere Bezeichnung für eine Natriumhydroxidlösung) benutzt.

20. _____ (andere Bezeichnung für eine Calciumhydroxidlösung) zählt zu den wichtigen Laborchemikalien.

Wortschatztest zur englischen Fachsprache

Hilfsmittel: Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Teil 1

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

Iron sulfide was produced in the reaction.

Das Wort **iron sulfide** bezeichnet eine Verbindung aus

A Blei und Schwefel

B Aluminium und Schwefel

C Zink und Schwefel

D Eisen und Schwefel

Hier ist die **Antwort D** richtig, also würde das Kreuz auf dem Antwortbogen bei **D** gemacht:

Nr.	A	B	C	D
Beispiel				X

Aufgaben

1. He needed a **protective liquid**.

Der Begriff **protective liquid** bezeichnet eine Substanz

A zum Schutz von Stoffen.

B zum Schutz von Glasgefäßen.

C zum Schutz vor Überhitzung.

D zum Schutz vor Verbrennungen.

2. He took some **rubidium**.

Das Wort **rubidium** bezeichnet ein

A Edelmetall.

B Nichtmetall.

C reaktives Metall.

D Indikator.

3. They needed a **sodium spoon**.

Der Begriff **sodium spoon** bezeichnet einen Löffel zum

A Abfüllen von Natrium.

B Fixieren von Natrium.

C Verbrennen von Natrium.

D Zerkleinern von Natrium.

4. They used **strotium** for the experiment.

Das Wort **strontium** bezeichnet eine/n

A Feststoff.

B Verbindung.

C Flüssigkeit.

D Gas.

5. He looked for the box with **beryllium**.

Das Wort **beryllium** bezeichnet ein/e/en

A Kristall.

B Metall.

C Mineral.

D Legierung.

6. At the end **lime milk** was left over.

Lime milk ist

- A Kaliumcarbonat in Wasser.
- B Natriumoxid in Wasser.
- C Calciumhydroxid in Wasser.
- D Calciumsulfat in Wasser.

7. They used some **barium**.

Das Wort **barium** bezeichnet ein

- | | | | |
|-----------------|-------------------|---------------|----------------|
| A Leichtmetall. | B unedles Metall. | C Edelmetall. | D Nichtmetall. |
|-----------------|-------------------|---------------|----------------|

8. He took more **caesium** than usual.

Das Wort **caesium** bezeichnet ein/e/en

- | | | | |
|----------------|--------|--------------|------------|
| A Flüssigkeit. | B Gas. | C Feststoff. | D Gemisch. |
|----------------|--------|--------------|------------|

9. A **magnesia tray** was needed for the experiment.

Der Begriff **magnesia tray** bezeichnet ein Laborgerät das

- A magnetisch ist.
- B scharf ist.
- C feuerfest ist.
- D lichtempfindlich ist.

10. He added a calcium **granule**.

Der Begriff **granule** bezeichnet

- | | | | |
|------------------|-------------------|------------------|----------------|
| A einen Überzug. | B einen Streifen. | C ein Stückchen. | D ein Gemisch. |
|------------------|-------------------|------------------|----------------|

11. The **pneumatic principle** plays a role in some experiments.

Das **pneumatic principle** kommt zur Anwendung, wenn Gase

- | | | | |
|--------------|------------------|------------------|----------------------------|
| A entstehen. | B sich abkühlen. | C sich erwärmen. | D zusammengedrückt werden. |
|--------------|------------------|------------------|----------------------------|

.....

Teil 2Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

In a reaction with sulphur sulfides (Verbindungen mit Schwefel) can be formed.

Hier ist das Lösungswort sulfides. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<u>sulfides</u>

Aufgaben

12. _____ (K) is a very reactive metal.

13. _____ (Li) reacts with water.

14. _____ (Ca) is important for your bones.

15. _____ (Bezeichnung für Verbindungen die OH enthalten) can be produced when putting a metal into water.

16. The smoke contained solid _____.
(Verbindung aus Na und O)

17. _____ are metals.
(Oberbegriff für die Metalle der 1. Hauptgruppe des PSE)

18. _____ are metals, too.
(Oberbegriff für die Metalle der 2. Hauptgruppe des PSE)

19. _____ (andere Bezeichnung für Natriumhydroxidlösung) is an important substance in chemical laboratories.

20. _____ (andere Bezeichnung für Calciumhydroxidlösung) is often used in chemical experiments.

Wortschatztest zur englischen Fach- und Allgemeinsprache

Hilfsmittel: Bleistift, Spitzer, Radiergummi

Gruppe A

Aufgabenheft

Bitte nicht in dieses Heft schreiben!

Alle Antworten werden auf einem Antwortbogen notiert.

Englische Fachsprache

Teil 1

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

The reaction needs **activation energy**.

Das Wort **activation energy** bezeichnet:

A Aktivenergie

B Aktivierungsenergie

C Aktionsenergie

D Aktivitätsenergie

Hier ist die **Antwort B** richtig, also würde das Kreuz auf dem Antwortbogen bei **B** gemacht:

Nr.	A	B	C	D
Beispiel		X		

Aufgaben

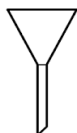
1. He needed a **clay triangle** for the experiment.

Kreuze die richtige Abbildung für **clay triangle** an.

A



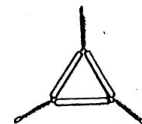
B



C



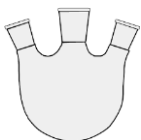
D



2. He used a **tripod**.

Kreuze die richtige Abbildung für **tripod** an.

A



B



C



D



3. She told them the **set-up** of the experiments.

Das Wort **set-up** bezeichnet:

A die Reihenfolge

B die Vorbereitung

C den Aufbau

D die Versuchsanleitung

4. The substance was **soluble**.

Das Wort **soluble** bedeutet:

- A löslich B spröde C elastisch D biegsam

5. They described the **properties** of the mixtures.

Das Wort **property** bedeutet:

- A Aggregatzustand B Eigenschaft C Bestandteil D Zusammensetzung

6. Different **compounds** were analysed.

Das Wort **compound** bedeutet:

- A Gemisch B Verbindung C Stoff D Probe

7. The **procedure** of the experiments was planned.

Das Wort **procedure** bedeutet:

- A Anordnung B Reihenfolge C Durchführung D Vorbereitung

8. He needed an **evaporating dish**.

Der Begriff **evaporating dish** bezeichnet ein Laborgerät zum

- A Trennen von Stoffen.
B Mischen von Stoffen.
C Kühlen von Stoffen.
D Eindampfen von Stoffen.

9. They used **forceps**.

Das Wort **forceps** bedeutet

- A Tiegelzange. B Stativklammer. C Pinzette. D Siedesteine.

10. He filled the bottle with **lime water**.

Lime water bedeutet

- A Schwefelwasser. B Kalkwasser. C Seifenwasser. D destilliertes Wasser.

.....

Teil 2Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

The water started to boil (sieden).

Hier ist das Lösungswort **boil**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>boil</i>

Aufgaben

11. He needed a _____ (Spatel) for the experiment.
12. They filled the _____ (Reagenzglas) with water.
13. During the experiment different _____ (Beobachtungen) were discussed.
14. Next he had to _____ (auflösen) the substance in water.
15. He prepared a _____ (Lösung) of salt in water.
16. _____ (Wasserstoff) is an explosive gas.
17. Some metals must be protected against _____ (Oxidation).
18. Wood has a lower _____ (Dichte) than water.
19. The _____ (Knallgasprobe) was done to identify the gas.
20. A _____ (Redoxreaktion) took place.
21. The substance is not acid but _____ (alkalisch).
22. _____ (Sauerstoff) is a gas in our air.
23. The reaction was _____ (exotherm).
24. The used a _____ (Brenner) to boil the substances.
25. They had to note down the chemical _____ (Gleichung) for the reaction.

Englische Allgemeinsprache

Teil 3

Aufgabenstellung

Finde die richtige Antwort und **kreuze das entsprechende Kästchen** auf dem **Antwortbogen an**.
Es ist immer **nur eine** Antwort richtig.

Beispiel

They **burnt** the paper.

Das Wort **burn** bezeichnet:

A falten

B verbrennen

C wegwerfen

D zerschneiden

Hier ist die **Antwort B** richtig, also würde das Kreuz auf dem Antwortbogen bei **B** gemacht:

Nr.	A	B	C	D
Beispiel		X		

1. The material was **exposed to** heat.

Der Ausdruck **to be exposed to** bedeutet:

A behandelt sein gegen

B beständig sein gegen

C ausgesetzt sein

D geschützt sein vor

2. They saw that the plastic was **floating**.

Das Wort **to float** bedeutet:

A fließen

B schmelzen

C schwimmen

D schäumen

3. The fire burnt **vigorously**.

Das Wort **vigorously** bedeutet:

A gleichmäßig

B langanhaltend

C heftig

D schwach

4. The substance started **to tarnish**.

Das Wort **to tarnish** bedeutet:

A kondensieren

B sich zersetzen

C anlaufen

D verdunsten

5. They removed the **residue** from the bottle.

Das Wort **residue** bedeutet:

A Belag

B Rückstand

C Verschluss

D Inhalt

6. They couldn't find the **sieves**.

Das Wort **sieve** bedeutet:

A Papiertuch

B Trichter

C Sieb

D Klammer

7. After a few minutes the substance **ignites**.

Das Wort **ignite** bedeutet:

A verdampfen

B entzünden

C aufschäumen

D trocknen

8. A gas was **released**.

Das Wort **released** bedeutet:

- A verflüssigt B aufgefangen C freigesetzt D eingeleitet

9. The **sphere** rolls down the table.

Das Wort **sphere** bedeutet:

- A Scheibe B Rad C Kugel D Rohr

10. The material **gleams**.

Das Wort **to gleam** bedeutet:

- A glänzen B kleben C erweichen D sich verfärben

Teil 4

Aufgabenstellung

Finde die passende **Ergänzung für die Lücken** und notiere das entsprechende **Wort auf dem Antwortbogen**.

Beispiel

They solved (lösen) the problem

Hier ist das Lösungswort **solved**. Also notierst du dieses Wort im entsprechenden **Kästchen auf dem Antwortbogen**.

Nr.	Lücke
Beispiel	<i>solved</i>

11. After some time the water became _____ (trübe).

12. A gas will _____ (entweichen) from the balloon.

13. The sugar was _____ (verbraucht) when they prepared the cake.

14. It's getting warmer so the temperature will _____ (steigen).

15. They had to _____ (erhitzen) the water first.

16. I love pizza with a thick _____ (Kruste).

17. The book about his life will _____ (enthalten) funny stories.

18. A _____ (Verbrennung) of diesel took place.

19. The substance became _____ (fest) after some time.
20. They put _____ (flüssig) chocolate on the cake.
21. The pupils _____(beobachten) wild animals in the forest.
22. The detective _____(schlussfolgert) that Peter was the thief.
23. It's getting colder. The temperature _____ (sinkt).

.....