

Effektivität des Gruppenpuzzles
im Chemieunterricht der Sekundarstufe I

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Pädagogik
(Dr. paed.)

Angenommen von der Fakultät Chemie
der Technischen Universität Dortmund

von
Markus Tepner
aus Rheinberg

Dortmund 2007

1. Gutachterin: Prof. Dr. Insa Melle

2. Gutachter: Prof. Dr. Bernd Ralle

Tag der Disputation: 11.12.2007

Homines dum docent discunt

(Während die Menschen lehren, lernen sie.)¹

¹ aus: SENECA: *Epistulae Morales ad Lucilium - Briefe an Lucilius über Ethik* (lat./dt., hg. von Franz Loretto) Buch 1, Brief 7, Abschnitt 8.

Inhaltsverzeichnis

1	EINLEITUNG UND HYPOTHESEN	5
2	KOOPERATIVES LERNEN	8
2.1	DEFINITION	8
2.2	MERKMALE	8
2.3	METHODEN	10
2.4	BEGRÜNDUNG	11
2.5	VORTEILE UND VORURTEILE	16
2.6	FORSCHUNGSSTAND	17
2.6.1	<i>Internationale Studien zum kooperativen Lernen.....</i>	<i>17</i>
2.6.2	<i>Internationale Studien zum Gruppenpuzzle.....</i>	<i>18</i>
2.6.3	<i>Nationale Studien zum kooperativen Lernen.....</i>	<i>19</i>
3	DAS GRUPPENPUZZLE	21
3.1	EINFÜHRUNG	21
3.2	ALLGEMEINER ABLAUF	21
3.3	DAS GRUPPENPUZZLE SEIFEN	22
3.3.1	<i>Die Voraussetzungen</i>	<i>23</i>
3.3.2	<i>Expertenmaterial</i>	<i>24</i>
3.3.3	<i>Arbeitsblätter</i>	<i>25</i>
3.3.4	<i>Durchführung</i>	<i>26</i>
4	STICHPROBE	28
4.1	STICHPROBENMERKMALE UND AUSWAHL DER STICHPROBE	28
4.2	BETEILIGTE SCHULEN.....	30
4.3	STICHPROBENGRÖÖE	32
4.4	VERTEILUNG VON ALTER UND GESCHLECHT DER SCHÜLER.....	32
4.5	LEHRKRÄFTE	33
4.6	EINORDNUNG.....	33
5	UNTERSUCHUNGSDESIGN.....	35
5.1	PROJEKTVERLAUF.....	35

5.2	KONZEPTION.....	35
5.3	ZENTRALE GÜTEKRITERIEN UND DEREN EINHALTUNG.....	37
5.3.1	<i>Objektivität</i>	37
5.3.2	<i>Reliabilität</i>	38
5.3.3	<i>Validität</i>	40
5.4	DETAILS ZUM UNTERSUCHUNGSDESIGN.....	43
5.4.1	<i>Allgemeines</i>	43
5.4.2	<i>Änderungen von Hauptuntersuchung I nach Hauptuntersuchung II</i>	45
5.4.3	<i>Ausrichtung</i>	46
6	VORSTELLUNG DER EINGESETZTEN ERHEBUNGSINSTRUMENTE.....	48
6.1	WISSENSTEST.....	48
6.2	MEINUNGSTESTS.....	49
6.2.1	<i>Einstellung zum Fach Chemie</i>	49
6.2.2	<i>Attraktivität der Unterrichtseinheit</i>	51
6.2.3	<i>Attraktivität des Gruppenpuzzles</i>	51
6.3	INDIVIDUELLE SCHÜLERBEURTEILUNGEN.....	52
6.4	AUDIOAUFZEICHNUNGEN.....	52
6.5	VIDEOAUFZEICHNUNGEN.....	53
6.6	STUNDENPROTOKOLLE.....	53
6.7	BEOBACHTUNGSBÖGEN.....	53
6.8	KLASSEN- UND LEHRERFEEDBACK.....	54
7	AUSWERTUNG.....	55
7.1	WISSENSTEST.....	55
7.1.1	<i>Einfluss der Unterrichtsmethode auf den Lernzuwachs</i>	55
7.1.2	<i>Abhängigkeit der Lernleistung von der Art der Wissensvermittlung</i>	61
7.1.3	<i>Wechselwirkungen zwischen Unterrichtsmethode und Chemiezensur</i>	63
7.1.4	<i>Abhängigkeit des Lernerfolgs vom Geschlecht</i>	65
7.1.5	<i>Abhängigkeit des Lernerfolgs von der Schulform</i>	65
7.1.6	<i>Einfluss der Übungsdauer auf den Lernerfolg</i>	65
7.1.7	<i>Nachhaltigkeit des erworbenen Wissens</i>	65
7.2	MEINUNGSTEST.....	66
7.2.1	<i>Einstellung zum Fach Chemie</i>	66

7.2.2	<i>Attraktivität der Unterrichtseinheit</i>	69
7.2.3	<i>Attraktivität des Gruppenpuzzles</i>	69
7.3	INDIVIDUELLE SCHÜLERBEURTEILUNGEN	69
7.4	AUDIOAUFZEICHNUNGEN	70
7.5	VIDEOAUFZEICHNUNGEN	72
7.6	STUNDENPROTOKOLLE	72
7.7	BEOBACHTUNGSBÖGEN	72
7.8	KLASSEN- UND LEHRERFEEDBACK	73
8	ERGEBNISSE UND DISKUSSION	74
8.1	WISSENSTEST	74
8.1.1	<i>Einfluss der Unterrichtsmethode auf den Lernzuwachs</i>	74
8.1.2	<i>Abhängigkeit der Lernleistung von der Art der Wissensvermittlung</i>	89
8.1.3	<i>Wechselwirkungen zwischen Unterrichtsmethode und Ausgangsleistung</i> ..	97
8.1.4	<i>Abhängigkeit des Lernerfolgs vom Geschlecht</i>	100
8.1.5	<i>Abhängigkeit des Lernerfolgs von der Schulform</i>	102
8.1.6	<i>Einfluss der Übungsdauer auf den Lernerfolg</i>	103
8.1.7	<i>Nachhaltigkeit des erworbenen Wissens</i>	106
8.2	MEINUNGSTEST	110
8.2.1	<i>Einstellung zum Fach Chemie</i>	110
8.2.2	<i>Attraktivität der Unterrichtseinheit</i>	113
8.2.3	<i>Attraktivität des Gruppenpuzzles</i>	116
8.3	INDIVIDUELLE SCHÜLERBEURTEILUNGEN	117
8.4	AUDIOAUFZEICHNUNGEN	119
8.5	VIDEOAUFZEICHNUNGEN	123
8.6	STUNDENPROTOKOLLE	124
8.7	BEOBACHTUNGSBÖGEN	124
8.8	KLASSEN- UND LEHRERFEEDBACK	124
9	FAZIT UND AUSBLICK	126
10	ZUSAMMENFASSUNG	129
11	LITERATUR	134
	INTERNETQUELLEN	146

12	ANHANG	150
12.1	VERZEICHNISSE	150
12.1.1	<i>Abbildungsverzeichnis</i>	<i>150</i>
12.1.2	<i>Tabellenverzeichnis</i>	<i>152</i>
12.1.3	<i>Abkürzungsverzeichnis</i>	<i>153</i>
12.1.4	<i>Inhaltsverzeichnis Anhang.....</i>	<i>154</i>

1 Einleitung und Hypothesen

Spätestens die PISA-Studie im Jahr 2000 hat deutlich gemacht, dass der traditionelle Unterricht des deutschen Schulsystems nicht uneingeschränkt zum gewünschten Erfolg führt. Diese wenig ermutigenden Ergebnisse konnten durch PISA 2003 noch untermauert werden: Über 50 % der untersuchten Schüler konnten ihre naturwissenschaftliche Kompetenz im Verlauf eines Schuljahres nicht weiter entwickeln, sondern bauten diese zum Teil sogar ab (PRENZEL 2006, S. 8). Diese Tatsache sowie sich wandelnde Berufsanforderungen, verändertes Schülerverhalten² und eine zunehmend pluralistische Gesellschaft machten und machen die Suche nach neuen Methoden und Konzepten erforderlich.

Eine viel versprechende Unterrichtsform, mit der Schüler adäquater auf die beruflichen Anforderungen und ihren Platz in einer Toleranz und Teamfähigkeit fordernden Gesellschaft vorbereitet werden können, scheint das Kooperative Lernen zu sein. Darauf deuten zumindest aktuelle Forschungsergebnisse hin.

Kooperative Lernformen wie z. B. Gruppenrallye, Kugellager oder Gruppenpuzzle wurden erst seit Mitte der 1990er Jahre in die Diskussion um das deutsche Bildungssystem aufgenommen, gehören aber mittlerweile zum festen Bestandteil aktueller Richtlinien und Lehrpläne. Mit ihrem Einsatz wuchs auch die Frage nach ihrem Erfolg bzw. ihrer Effektivität. Dabei seien Vorteile wie z. B. Stärkung des Selbstvertrauens, Training von sozialen Kompetenzen oder Förderung der Lernzufriedenheit unbestritten. Allerdings rückt insbesondere durch die PISA-Studie der Leistungsgedanke und damit auch die Effektivität in Bezug auf den Wissenszuwachs verstärkt in den Vordergrund.

Zur Effektivität kooperativer Unterrichtsformen gibt es vor allem außerhalb Europas bereits zahlreiche Studien bzw. Metaanalysen (vgl. JOHNSON et al. 2000, JOHNSON & JOHNSON 1989, SLAVIN 1995 oder LAZAROWITZ & HERTZ-LAZAROWITZ 1998), in denen dieses Unterrichtsprinzip überwiegend positiv abschneidet. Allerdings deuten einzelne Studien darauf hin, dass der kognitive Lernzuwachs bei schwächeren Schülern geringer ausfällt.

Kooperatives Lernen in Deutschland – speziell im Chemieunterricht – ist bisher nur wenig untersucht (z. B. EILKS 2003b, SUMFLETH et al. 2004), vorwiegend handelt es sich dabei

² Aus Gründen der Übersichtlichkeit wird im Folgenden „Schüler“ statt „Schülerinnen und Schüler“ verwendet. Unter dem Begriff „Lehrer“ sollen ebenfalls beide Geschlechter zusammengefasst sein.

um Erfahrungsberichte – z. B. im Rahmen Partizipativer Aktionsforschung (vgl. z. B. EILKS & RALLE 2002) – und selten um empirische Studien.

Untersuchungen zur tatsächlichen Effektivität im Sinne einer Vergleichsgruppenuntersuchung, die strengeren empirischen Anforderungen genügt, existieren für den deutschen Chemieunterricht bislang nicht.

Da Ergebnisse internationaler Studien nicht ohne weiteres auf das deutsche Schulsystem übertragen werden können, sollte in einer Studie anhand von Referenzgruppen evaluiert werden, ob das Gruppenpuzzle als Stellvertreter für eine kooperative Lernform die Lernleistung und die Einstellung der Schüler im direkten Vergleich mit „herkömmlichem“ Chemieunterricht erhöhen kann. Darunter soll in dieser Arbeit der im weitesten Sinne darlegend-informative bzw. fragend-entwickelnde, lehrerzentrierte Unterricht verstanden werden.

Ein weiterer Schwerpunkt der im Rahmen der vorliegenden Dissertation durchgeführten Untersuchungen war die Konzeption und Erprobung eines im realen Unterricht allgemein einsetzbaren Evaluationsinstrumentes, mit dem verschiedene Unterrichtsmethoden empirisch miteinander verglichen werden können.

Dem Gruppenpuzzle wird nachgesagt, dass es die Eigeninitiative der Schüler fördert, das Selbstvertrauen stärkt, die Aggressionsbereitschaft in der Klasse senkt und es zu einer höheren Wertschätzung der Schüler untereinander kommt (FREY-EILING & FREY 1999, S. 52). Da diese so genannten Soft Skills nur schwer quantitativ erfassbar sind, sollen sie von der Untersuchung ausgenommen werden.

Folgende **Hypothesen** liegen dieser Arbeit zugrunde:

1. Schüler, die nach der Gruppenpuzzlemethode unterrichtet werden, erzielen einen höheren Wissenszuwachs als inhaltsgleich frontal unterrichtete Schüler.
2. Die Methode des Gruppenpuzzles ist für Schüler attraktiver als vergleichbarer Frontalunterricht.

Folgende **Nebenhypothesen** wurden formuliert:

1. Das Gruppenpuzzle eignet sich für leistungsstarke bzw. -schwache Schüler gleichermaßen.
2. Bei Themen, die sich die Schüler selbst angeeignet haben, erzielen sie einen höheren Wissenszuwachs als bei Themen, die ihnen von ihren Lehrern bzw. Mitschülern vermittelt wurden.
3. Das Gruppenpuzzle beeinflusst die Einstellung der Schüler zum Fach Chemie.

2 Kooperatives Lernen

2.1 Definition

Nach JOHNSON & JOHNSON (1999, S. 1 ff) lassen sich Lernaktivitäten grundsätzlich in drei unterschiedliche Interaktionsmuster einteilen: wettbewerblich, individualistisch und kooperativ. Auf die beiden erstgenannten soll an dieser Stelle nicht eingegangen werden; sie sind lediglich der Vollständigkeit halber aufgeführt. JOHNSON & JOHNSON (1999, S. 5) definieren kooperatives Lernen (engl. „cooperative learning“) als *„the instructional use of small groups so that students work together to maximise their own and each other’s learning“*. Eine präzisere Definition für kooperatives Lernen gibt das Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien (THILLM, 2002, S. 20):

„Kooperatives Lernen bezeichnet Interaktionsformen, bei denen alle Mitglieder einer Gruppe gemeinsam und im wechselseitigen Austausch Kenntnisse und Fähigkeiten erwerben. Dabei sind alle Gruppenmitglieder aktiv am Lerngeschehen beteiligt und tragen gemeinsame Verantwortung für das Lernergebnis.“

Diese Definition umfasst viel mehr, als es bei einfacher Gruppenarbeit der Fall ist. Auf die Unterschiede wird in Kapitel 2.2 eingegangen.

2.2 Merkmale

Kooperatives Lernen basiert auf der Grundannahme, dass Lernen ein sozialer Prozess ist und Schüler gerne in Kontakt mit ihren Mitschülern sind. Dabei wird das Bedürfnis nach Interaktion positiv und konstruktiv genutzt (EILKS et al. 2005, S. 7). Es werden zum Beispiel gezielt Situationen erzeugt, in denen sich die Schüler gegenseitig etwas beibringen. Auf der Grundlage dieser wenigen aber wichtigen Annahmen lassen sich fünf Qualitätsmerkmale zusammenfassen, die kooperatives Lernen auszeichnen (EILKS et al. 2005, S. 8, nach JOHNSON und JOHNSON 1994):

- Positive gegenseitige Abhängigkeit
- Individuelle Verantwortungsübernahme / Verbindlichkeit
- Direkte, unterstützende Interaktion mit dem Gegenüber
- Häufiger Einsatz von Gruppen- und zwischenmenschlichen Kompetenzen
- Gemeinsame Reflexion des Prozesses

HUBER (2000, S. 57) ergänzt diese Merkmale um eine weitere „*Minimalbedingung*“, dem Spielraum für Entscheidungen.

Anhand der beschriebenen Merkmale wird deutlich, dass kooperatives Lernen weit mehr ist als eine Sozialform, wie es die traditionelle Gruppenarbeit ist. Es handelt sich vielmehr um eine Unterrichts-„Strategie“ bzw. um ein umfassendes Konzept, wie WEIDNER (2003) und EILKS et al. (2005) verdeutlichen.

Tabelle 1: Gegenüberstellung von Kooperativem Lernen und traditionellem Gruppenunterricht. (Verändert nach WEIDNER, 2003, S. 30 und EILKS et al. 2005, S. 7, basierend auf JOHNSON & JOHNSON 1999, S. 73.)

Kooperative Lerngruppen	Traditionelle Lerngruppen
Strukturierung einer positiven Abhängigkeit	Unstrukturierte und nicht immer positive Abhängigkeit
Die Einzelnen übernehmen Verantwortung für den eigenen Lernprozess sowie den der anderen Gruppenmitglieder	Einzelne meist nur für sich selbst verantwortlich
Heterogene Zusammensetzung der Gruppe erwünscht bzw. verstärkt	Oft homogene Zusammensetzung
Permanente, teamaufbauende Aktivitäten um Vertrauen, Verantwortung für das Gruppengeschehen und festen Zusammenhalt zu fördern	Kaum teamaufbauende Aktivitäten
Geteilte Führung	Ein Teammitglied meist selbsterklärter Leiter der Gruppe
Systematisches Lehren und Praktizieren von Sozialen Fertigkeiten (social skills)	Soziale Fertigkeiten werden oft vorausgesetzt (fehlen aber häufig)
Lehrer beobachtet ständig die Gruppenarbeit, dokumentiert seine Beobachtungen, gibt Rückmeldung über das Funktionieren im Team und interveniert, wenn nötig	Systematisches Feedback erfolgt weniger ausgeprägt
Evaluation/Prozess-Reflexion durch die Gruppe	Wenig Evaluation/Prozess-Reflexion durch die Gruppe

Aus der Tabelle sowie den Merkmalen wird ersichtlich, dass Nachteile traditioneller Gruppenarbeit – wie z. B. das Abwälzen der Verantwortung auf einzelne Teilnehmer oder das sich „Verstecken“ hinter den Gruppenmitgliedern – durch kooperatives Lernen vermieden werden soll.

Die obige Gegenüberstellung gilt es doch insofern kritisch zu betrachten, als dass die traditionelle Gruppenarbeit je nach Lehrer stark variieren kann. Dargestellt ist bewusst der „worst case“ um Unterschiede deutlicher zu machen. Aber nicht nur die Gruppenarbeit, sondern auch der traditionelle Frontalunterricht weist potenzielle Probleme auf. Hierbei gibt es drei Hauptprobleme (vgl. THILLM 2002, S. 17):

- Die Angleichung des Schüler-Denkens an das des Lehrers kann zum Verzicht auf die eigenen (Denk-)Leistungen führen.
- Eine negative Bewertung scheinbar unpassender Antworten kann Schüler entweder unsicher machen oder dazu führen, dass sie sich vom Unterricht distanzieren.
- Der Lehrer übernimmt die Hauptverantwortung für Unterrichtsinhalt und -prozess und somit für den Lernerfolg. Daraus können bei einigen Lehrern ein permanentes Überforderungserleben und Stress resultieren, weil sie letztlich die hauptsächlich Engagierten im Unterricht sind. Hinzu kommt eine enttäuschte Grundhaltung gegenüber den scheinbar unbefriedigenden Schülerleistungen.

Im nächsten Unterkapitel soll dargelegt werden, warum gerade kooperatives Lernen eine Möglichkeit sein kann, den Problemen traditionellen Unterrichts zu begegnen.

2.3 Methoden

In den letzten Jahrzehnten hat sich vor allem in den USA und Kanada eine Vielzahl von kooperativen Unterrichtsmethoden etabliert. Da diese jedoch kaum im deutschen Chemieunterricht eingesetzt werden, seien hier lediglich die in Deutschland populärsten Methoden genannt: GRUPPENPUZZLE (Expertenmethode, Jig-Saw-Puzzle), GRUPPENRECHERCHER (Group Investigation), KUGELLAGER (Karussellgespräch, „Innenkreis-Außenkreis“), LERNEN DURCH LEHREN, PROJEKTORIENTIERTER UNTERRICHT (Projektunterricht), GRUPPENRALLYE (Student Teams and Achievement Divisions, STAD), LERNEN AN STATIONEN (Lernzirkel, Stationenlernen), 1-2-4-ALLE (Pairs-to-Share), LERNFIRMA, KOOOPERATIVES ÜBEN MIT LERNKARTEIEN (vgl. EILKS & STÄUDEL 2005).

Dabei dürfte das Stationenlernen und an berufsbildenden Schulen besonders der Projektunterricht am populärsten sein. Da Lernzirkel oft individualistisch gestaltet werden, muss ein kooperatives Moment nicht zwangsläufig enthalten sein, weshalb diese Methode unter Vorbehalt einzuordnen ist.

2.4 Begründung

UNRUH & PETERSEN (2006, S. 55) formulieren plakativ: *„Kinder und Jugendliche von heute sind anders als früher“*. Und auf der einen Seite klagen viele Lehrer, dass Schüler heutzutage weniger wüssten und unkonzentrierter, regelloser sowie weniger motiviert seien (vgl. ebd.). Hinzu kommt eine *„Zunahme auffälligen Schülerverhaltens“* (JÜRGENS 1996, S. 26). Dabei fällt auf, dass die *„heutigen Kinder eher ichbezogen und im Vergleich zu früher weniger rücksichtsvoll seien“* (FÖLLING-ALBERS 1992, S. 34 ff). Damit folgen sie einem Trend, der ihnen von der erwachsenen Gesellschaft vorgelebt wird und die elterlichen Erziehungsziele stark beeinflusst: Eltern legen heute vielfach mehr Wert auf Selbstständigkeit und Durchsetzungsvermögen ihrer Kinder (GREEN & GREEN 2005, S. 16). Andererseits zeichnen sich die Kinder durch größeren Individualismus, mehr Selbstvertrauen und zahlreiche Zusatzqualifikationen aus. Dazu zählen z. B. das Präsentieren sowie der Umgang mit neuen Medien.

UNRUH & PETERSEN sehen vor allem im selbstständigen Lernen *„eine sinnvolle Möglichkeit, den veränderten Kindern und Jugendlichen besser gerecht zu werden“* (ebd., S. 56). Von den kooperativen Unterrichtsformen besitzt insbesondere das Gruppenpuzzle das Potenzial auch für selbstständiges Arbeiten (siehe Kapitel 3).

Neben einer stärkeren Individualisierung auf der einen Seite fordert eine sich wandelnde Berufswelt immer mehr Teamfähigkeit von ihren Arbeitnehmern – selbstverständlich um im Sinne der heutigen Leistungsorientierung noch produktiver und damit gewinnbringender zu arbeiten. Diese Teamfähigkeit setzt ein hohes Maß an sozialer Kompetenz voraus, zu dem FRIEDE (1994) folgende Komponenten zählt: Kommunikations- und Kooperationsfähigkeit, Durchsetzungs- und Konfliktfähigkeit, Empathie und Flexibilität. GREEN & GREEN (2005) sehen in kooperativen Lernformen ein Trainingsmittel für soziale Kompetenzen, was durch JOHNSON, JOHNSON und HOLUBEC (1984) bestätigt wird.

Weiterhin nehmen das Erkennen von Unterschieden und deren Akzeptanz eine große Rolle in unserer zunehmend pluralistischen Gesellschaft ein. GREEN & GREEN (2005) bezeichnen dies *„als eine Grundvoraussetzung, um bei den Lernenden ein Gefühl der Zusammengehörigkeit, der Toleranz und des gegenseitigen Respekts zu entwickeln“* (S. 32). Kooperatives Lernen hilft, diese Kompetenzen zu stärken, indem es die notwendigen positiven Abhängigkeiten betont. Gerade ein frühzeitiges Lernen in und durch einen multi-kulturellen Kontext, wie er in einer multi-kulturellen Gesellschaft früher oder später stattfinden wird,

macht die Schüler vertraut mit alltäglichen Situationen des späteren Berufslebens. Auf der anderen Seite kann nur derjenige später beruflich erfolgreich sein, der fähig ist, im Team auch mit „unsympathischen“ Kollegen/Mitschülern erfolgreich zusammenzuarbeiten, sich also professionell zu verhalten, um Synergieeffekte der Gruppe zu nutzen.

Aus lerntheoretischer Sicht zeichnet sich guter Unterricht durch häufigen Methodenwechsel aus, da *„unterschiedliche Lernziele unterschiedliche Lehr- und Lernmethoden erfordern“* (HELMKE, zitiert nach EILKS & STÄUDEL 2005a). WEINERT (1998) begründet den Einsatz von unterschiedlichen Methoden mit der Art des Wissens, das vermittelt werden soll: Für „intelligentes Wissen“ empfiehlt er eine lehrergesteuerte, aber schülerzentrierte, direkte Instruktion. Um die lebenspraktische Nutzung von Kenntnissen und Fertigkeiten zu üben, eignen sich nach WEINERT (1998) vor allem kooperative Lernformen, wie z. B. die Projektmethode. Gleiches gilt für den Erwerb von variabel nutzbaren Schlüsselqualifikationen, die durch selbstständiges Arbeiten besonders geschult werden können.

Aber auch unterschiedliche Lerntypen verlangen nach verschiedenen Methoden, die durch kooperatives Lernen leichter bereitgestellt werden können als z. B. durch Frontalunterricht. Eine Auswahl an kooperativen Methoden mit ihren Variationen findet sich in GREEN & GREEN (2005, S. 125 – 136). Allen gemeinsam ist die hohe Schüleraktivität, die den Lehrer als Informationsgeber stark in den Hintergrund treten lässt. Um zu verstehen, warum diese Form des Lernens zurzeit so populär ist, soll zunächst der Frage nachgegangen werden, wie Lernen funktioniert. Der Vollständigkeit halber wird vorab eine kurze Definition von „Lernen“ gegeben:

ZIMBARDO und GERRIG (1999, S. 791) verstehen unter Lernen einen *„Prozess, der zu relativ stabilen Veränderungen im Verhalten oder Verhaltenspotential führt und auf Erfahrung beruht. L. [Lernen] ist nicht direkt beobachtbar, sondern wird aus Veränderungen des beobachtbaren Verhaltens erschlossen“*. Aber wie funktioniert Lernen?

Wachsende Bedeutung bei der Beantwortung dieser Frage kommt dabei einer neueren Lerntheorie, dem Konstruktivismus, zu. Dieser seit Ende der 1980er Jahre stark in den Fokus geratene Ansatz (KEIMES 2003, S. 5) versteht den Lernprozess als einen individuellen Vorgang der aktiven Wissenskonstruktion. *„Wissen wird demnach nicht einfach angeeignet oder durch Instruktion übernommen, sondern selbstaktiv und individuell unterschiedlich konstruiert“* (HELLBERG-RODE 2002, S. 1). *„Dementsprechend ist Wissen nicht vermittelbar, sondern muss immer wieder individuell konstruiert, reorganisiert und erweitert*

werden“ (ebd.). STORK (1995, S. 19) ergänzt für den Chemieunterricht: „[...] *chemisches Wissen kann man nicht vererben wie eine Million Dollar. Die Lernenden müssen es in der Tat für sich rekonstruieren*“. Zwar *„findet ohne Zweifel auch eine Weitergabe gesellschaftlich präsenten Wissens statt“* (ebd.), jedoch müssen *„Übernahme“* und *„Eigenaktivität“* zusammenwirken.

Eine kurze Übersicht über Annahmen der konstruktivistisch orientierten Lerntheorie gibt WOLFF (1997, S. 107):

- *„Es kann nur das verstanden und gelernt werden, was sich mit bereits vorhandenem Wissen verbinden lässt.*
- *Die eingesetzten Konstruktionsprozesse sind individuell verschieden; deshalb sind auch die Ergebnisse von Lernprozessen nicht identisch.*
- *Wissen ist immer "subjektives" Wissen, das sich selbst für Lernende, die im gleichen sozialen Kontext lernen, beträchtlich unterscheiden kann. Auch deshalb sind die Ergebnisse von Lernprozessen individuell verschieden.*
- *Neues Wissen impliziert die Umstrukturierung bereits vorhandenen Wissens. Der soziale Kontext, die soziale Interaktion sind (insofern) beim Lernenden von ausschlaggebender Bedeutung.*
- *Weil die Konstruktion von neuem Wissen an bereits vorhandenes Wissen angebunden ist, müssen Lernprozesse in reiche und authentische Lernumgebungen eingebettet werden. [...]*
- *Von besonderer Bedeutung ist das Prinzip der Selbstorganisation. Der Mensch als in sich geschlossenes System organisiert sich selbst und organisiert damit für sich die Welt.*
- *Selbstorganisation verbindet sich mit Eigenverantwortlichkeit. Der Mensch ist für das eigene Lernen verantwortlich, weil er damit sein Überleben als System sichert.“*

Nach GERSTENMAIER & MANDL 1995 sollte ein Konstruktivismus bejahender Unterricht folgende Elemente aufweisen:

- Umgang mit realistischen Problemen und authentischen Situationen
- multiple Kontexte für einen flexiblen Transfer auf andere Problemstellungen
- Problemerschließung aus unterschiedlichen Perspektiven
- kooperatives Problemlösen in Lerngruppen
- gemeinsames Lernen in der Interaktion zwischen Lernenden und Experten

Folgende Lernformen treten in einem konstruktivistisch orientierten Unterricht auf: eigenaktives konstruktives Lernen, situatives Lernen, soziales und kooperatives Lernen, selbstgesteuertes Lernen und unterstütztes Lernen (MÖLLER 1999, S. 132f und 2000, S. 141f).

Sowohl GERSTENMAIER & MANDL 1995 als auch MÖLLER (ebd.) sehen demnach in Kooperativem Lernen ein geeignetes Element, um Lernen aus konstruktivistischer Sicht zu fördern. EILKS & STÄUDEL (2005b, S. 4) betonen hierbei speziell die Rolle der Kommunikation beim Lernen und die Eignung kooperativer Lernformen. Weiterhin verweisen sie auf zwei zusätzliche Aspekte, die für kooperatives Lernen sprechen und insbesondere für den deutschen Chemieunterricht relevant sind:

Zum einen besitzt das Fach Chemie den Auftrag zur allgemeinen Bildung (vgl. MINISTERIUM FÜR SCHULE UND WEITERBILDUNG, WISSENSCHAFT UND FORSCHUNG DES LANDES NRW 1999, S. XI). Zur allgemeinen Bildung gehören neben fachlichen auch kommunikative, soziale und kooperative Kompetenzen. EILKS & STÄUDEL (2005b, S. 4) betonen: *„Kooperatives Lernen zielt auf die eigenständige Strukturierung und Bewältigung von Gruppenprozessen, gegenseitiges Erklären und das Wahrnehmen von Verantwortung für das Lernen anderer. Es schult dabei ganz zentral diese Ziele einer allgemeinen Bildung, die zunehmend von beruflichen Ausbildungsstätten, Hochschulen oder letztlich der Arbeitswelt eingefordert werden“*. Eine naturwissenschaftliche Bildung im Sinne der Scientific Literacy fordert auch die Unabhängigkeit von Lehrer, Lehrbuch und anderen schulgebundenen Medien (WOLLWEBER 2000, S. 185).

Zum anderen kann durch kooperatives Lernen Chemie als Unterrichtsfach für die Schüler attraktiver gemacht werden. Als Gründe führen EILKS & STÄUDEL (2005b, S. 5) hohe Schüleraktivität, die Möglichkeit zur Einflussnahme sowie mögliche Erfolgserlebnisse an. Dies ist insofern von Bedeutung, als dass gerade dieses Fach von Schülern als wenig populär angesehen wird bzw. es im Verlauf der Schulzeit dazu wird (ebd., S. 4).

Letztlich kann der Einsatz von kooperativem Lernen mit der Verfassung – hier am Beispiel des Landes Nordrhein-Westfalen – legitimiert werden. Dort heißt es in Artikel 7: *„Ehrfurcht vor Gott, Achtung vor der Würde des Menschen und Bereitschaft zum sozialen Handeln zu wecken, ist vornehmstes Ziel der Erziehung.“* (MINISTERIUM FÜR SCHULE UND WEITERBILDUNG, WISSENSCHAFT UND FORSCHUNG DES LANDES NRW (1993), S. 10)

Zur Häufigkeit des Einsatzes kooperativer Lernformen machen LAZAROWITZ & HERTZ-LAZAROWITZ (1998, S. 466) folgende Angaben: Sie sollten einen Anteil von wenigstens 30 % am Unterricht haben.

Trotz dieser Argumente werden kooperative Unterrichtsformen bislang nur wenig im deutschen Chemieunterricht eingesetzt (FISCHER et al. 2003 und MELLE & PÖPPING 2002). Ak-

tuellere Zahlen liegen nicht vor. Allerdings darf eine leichte Zunahme angenommen werden, da durch Lehrerfortbildungen die Bekanntheit dieser Methoden gesteigert werden konnte und diese mittlerweile verbindlich im Lehrplan festgeschrieben sind (vgl. MINISTERIUM FÜR SCHULE UND WEITERBILDUNG, WISSENSCHAFT UND FORSCHUNG DES LANDES NRW 1999). Laut Richtlinien für die Sekundarstufe II für das Fach Chemie sollen Lernprozesse *„Gelegenheit für kooperative Arbeitsformen geben“*, um eine Nachhaltigkeit des Lernens zu fördern. Speziell die Teamfähigkeit soll durch kooperative Unterrichtsformen herausgebildet werden. Die Bedeutsamkeit von kooperativem Lernen ist folglich bekannt. Dennoch ist der Einsatz als eher gering einzustufen. Als Gründe vermuten GREEN & GREEN 2005 (S. 19) einen höheren Geräuschpegel, die nicht durchweg aufgabenbezogenen Aktivitäten der Schüler sowie die Weigerung von manchen Schülern, mit „unbeliebten“ Mitschülern zusammenzuarbeiten. Zudem würden Aufgaben oft genug individuell bearbeitet und alle Ergebnisse zur Präsentation lediglich zusammengestellt (vgl. BISKUP 1994, S. 14), was dem Grundgedanken kooperativen Lernens nicht hinreichend entspricht. Der wichtigste Faktor, der über den Einsatz von kooperativen Unterrichtsmethoden entscheidet, ist der Lehrer selbst: Vielen Lehrkräften fällt es scheinbar schwer, einfach nur zuzuschauen (EILKS & STAUDEL 2005a). Dies kann einerseits durch negative Erfahrungen, mangelndes Vertrauen in die Methode oder dem Gefühl, dass die Schüler am Ende nicht genug gelernt haben, bedingt sein. Gerade für den letzten Punkt soll diese Arbeit Klarheit verschaffen.

Wie gut Lernen selbst gänzlich ohne Lehrer funktionieren kann, zeigt ein Beispiel aus der Schweiz: Dort erhielten rund 60 Schüler wegen Lehrermangels ein halbes Jahr lang in den Fächern Mathematik, Deutsch, Englisch, Französisch und Sport zum Schuljahresbeginn *„lediglich einen Auftrag mit Aufgaben und Lernzielen – die Schüler konnten dann wahlweise zu Hause, in der Mediothek oder in leer stehenden Klassenräumen arbeiten. Die Fachlehrer kamen nur einmal in der Woche zur Sprechstunde, darüber hinaus waren sie per E-Mail erreichbar“* (Der Spiegel 2005, S. 63). Bei Prüfungen schnitten diese 60 17-jährigen Schüler genauso gut ab wie ihre herkömmlich unterrichteten Mitschüler.

2.5 Vorteile und Vorurteile

GREEN & GREEN (2005, S. 12) nennen folgende Gründe, die für den Einsatz von kooperativen Lernformen sprechen:

- *„Schüler begreifen 95 % von dem, was sie gelernt haben erst dann, wenn sie es anderen vermitteln.“*
- *„Einer Gruppe kann man schwierigere Aufgaben als Einzelnen zumuten.“*
- *„Wenn Schüler verschiedener kultureller Herkunft zusammenarbeiten, beginnen sie ihre Unterschiede zu verstehen, und wie sie konstruktiv zusammenarbeiten können.“*
- *„Die Fähigkeit zu kritischem Denken nimmt zu, das Interesse und die Behaltensleistung in Bezug auf den Unterrichtsstoff verbessern sich.“*
- Allgemeiner: *„Was Sie heute in einer Gruppe tun, können Sie morgen als Einzelner.“*
- *„Lernende helfen sich gegenseitig und bilden so eine fördernde Gemeinschaft, die das Leistungsniveau des Einzelnen hebt“ (Synergieeffekt).*

Hinzu kommt eine gewisse Unterstützung gegen den „Burnout“ von Lehrern: Je weniger der Lehrer als „Disziplinator“ in Erscheinung treten muss, desto geringer ist seine psychische Belastung. WODZINSKI (2004, S. 5) sieht einen weiteren Vorteil in der Möglichkeit zu differenzieren. Gerade heterogene Lerngruppen würden vom Kooperativen Lernen profitieren. Dies gelte umso mehr, als dass sich der traditionelle Unterricht eher am oberen Drittel der Klasse orientiert.

Die meisten Erkenntnisse sind mittlerweile zumindest für das amerikanische und kanadische Schulsystem wissenschaftlich belegt (vgl. JOHNSON & JOHNSON 1999 und Abschnitt 2.6 dieser Arbeit). WEIDNER (1998, S. 19) konnte eine größere Motivation und Zufriedenheit unter den Schülern feststellen.

Trotz dieser auf den ersten Blick einleuchtenden und erstrebenswerten Eigenschaften von kooperativen Lernformen existieren ihnen gegenüber vielfach Vorurteile aus Sicht von Lehrern. Die teilnehmenden Lehrkräfte befürchten z. B., *„dass die Schüler dabei nichts lernen, der Zeitbedarf im Unterricht größer und die Vorbereitung sehr aufwendig ist“*. Ebenso kritisch sind sie bei der Einschätzung, ob sich „schwierige“ Klassen für diese Unterrichtsform überhaupt eignen. Viele hätten außerdem mit dem „Rucksack-Effekt“ Erfahrung gemacht: Ein Schüler arbeitet und die anderen profitieren.

2.6 Forschungsstand

2.6.1 Internationale Studien zum kooperativen Lernen

Kooperative Lernformen sind besonders in den USA und in Kanada vielfältig untersucht worden; JOHNSON et al. (2000) nennen über 900 Studien (z. B. COHEN 1994; JOHNSON & JOHNSON 1989; SHARAN 1980; SLAVIN 1977), in denen dieses Unterrichtsprinzip Gegenstand der Forschung war. Sie weisen jedoch ausdrücklich darauf hin, dass in der Mehrzahl der Studien nicht die Effektivität direkt getestet wurde.

Dennoch soll an dieser Stelle eine kurze Auflistung der wichtigsten Forschungsergebnisse bezüglich kooperativen Lernens erfolgen. EILKS et al. (2005, S. 9) fassen die Meta-Studien von LAZAROWITZ et al. (1994) und LAZAROWITZ & HERTZ-LAZAROWITZ (1998, S. 16-18) wie folgt zusammen, verweisen allerdings auf die Uneinheitlichkeit der Studien:

Kooperatives Lernen

- führt häufig zu einem höheren kognitiven Lernerfolg.
- führt zu einer positiveren Einstellung zum betreffenden Unterrichtsfach und zum Lerngegenstand.
- trägt zu einem besseren sozialen Klima innerhalb der Lerngruppe bei.
- ermöglicht individuelleres Lernen, trägt zur Entwicklung der Kommunikations- und Kooperationsfähigkeiten bei und führt in vielen Fällen zu einer positiven Entwicklung im Selbstwertempfinden der Schülerinnen und Schüler.
- hat ein hohes Potenzial für innere Differenzierung.
- führt zu einer stärkeren Einbindung der Schüler und zu mehr Aktivität im Lernprozess.

Eine ähnliche Übersicht geben GREEN & GREEN (2005, S. 33-37).

SLAVIN (1995) prüfte mehrere Duzend Studien, die vier Wochen oder länger dauerten und unterschiedliche kooperative Methoden untersuchten. Insgesamt schnitten kooperativ unterrichtete Schüler in Wissenstests um eine Viertel-Standardabweichung besser ab. Dabei hing der positive Effekt maßgeblich von der verwendeten Methode ab³.

Die aktuellste Meta-Studie stammt von JOHNSON⁴ und Mitarbeitern aus dem Jahr 2000⁴. Dort wurden 164 Studien aus den Jahren 1970 bis 1999 analysiert, in denen acht verschie-

³ <http://college.hmco.com/education/pbl/tc/coop.html#3> (Letzter Zugriff am 5.4.2007)

⁴ <http://www.co-operation.org/> (Letzter Zugriff am: 5.4.2007)

dene kooperative Lernformen mit wettbewerblichem bzw. individualistischem Lernen hinsichtlich des Lernerfolges verglichen wurden. Es zeigt sich eine deutliche Überlegenheit für das kooperative Arbeiten im Vergleich zu wettbewerblichem Lernen, jeweils mit hohen Effektstärken. Gegenüber individuellem Lernen konnten jedoch kaum Unterschiede (vgl. auch JOHNSON & JOHNSON 1999, S. 194) nachgewiesen werden. Kritisch anzumerken ist dabei der Informationsverlust, der entsteht, wenn JOHNSON et al. (2000) den Frontalunterricht („traditional instruction“) entweder mit „wettbewerblich“ oder „individualistisch“ kodiert haben, weil eine eindeutige Zuordnung nicht immer möglich ist.

Andere Studien, z. B. von LORD (1999, S. 22), beschäftigen sich mit kooperativem Lernen an der Hochschule. Zwar konnte er einen Vorteil für das Gruppenlernen aufzeigen, jedoch entspricht der stark dozierende Unterrichtsstil in der Kontrollgruppe kaum dem fragend-entwickelnden Unterricht an deutschen Schulen. Ein Vergleich mit deutschem „traditionellem“ Frontalunterricht scheint demnach zumindest sehr schwierig.

Untersuchungen im Fach Chemie sind unter den internationalen Studien kaum zu finden. Anzuführen ist BOWEN (2000, S. 118 ff), der 15 Untersuchungen im Fach Chemie aus den Jahren 1990 bis 1998 in einer Metastudie zusammenfasste. Insgesamt nahmen 437 High School-Schüler und fast 1100 College-Studenten teil. Im Mittel lagen die Leistungen der kooperativ unterrichteten Probanden um 14 Perzentil-Punkte höher als der Median (50 % Perzentil) der traditionell unterrichteten Schüler/Studenten. Die Effektstärke betrug 0.37.

2.6.2 Internationale Studien zum Gruppenpuzzle

Von den 164 Studien, die JOHNSON et al. (2000) in ihre Metastudie einbezogen, befassten sich lediglich 14 mit dem Gruppenpuzzle, und davon fünf mit dem Vergleich Gruppenpuzzle versus individualistischer Unterricht. Die Effektstärken variierten von 0.13 (individuelles Lernen) bis 0.29 (wettbewerbliches Lernen), folglich kann das Gruppenpuzzle Schülern einen Lernvorteil verschaffen.

Aus weiteren empirischen Untersuchungen geht hervor, dass durch Anwendung des Gruppenpuzzles die Aggressionsbereitschaft in der Klasse verringert und die Wertschätzung der Schüler untereinander erhöht werden kann. Tendenziell sind größere Lernerfolge gegenüber dem lehrerzentrierten Unterricht möglich (vgl. FREY-EILING & FREY 1999, S. 52 ff).

Auch dem Problem der permanenten Überforderung von einem Großteil der Klasse und der daraus möglicherweise resultierenden Schwächung des Selbstbewusstseins kann das

Gruppenpuzzle begegnen: „Das Gruppenpuzzle ist die einzige Unterrichtsmethode, die uns in den letzten 30 Jahren begegnet ist, welche nachweislich das Selbstvertrauen der Lernenden stärkt“ (FREY-EILING & FREY 1999, S. 52).

Bei allen Vorteilen, die das Gruppenpuzzle zu besitzen scheint, sollen aber auch etwaige Nachteile nicht unerwähnt bleiben: Da das Gruppenpuzzle ein „Sich-Zurücklehnen“ im Unterricht konstruktionsbedingt nicht gestattet, ist es für die Schüler anstrengender, was sich durch eine abnehmende Motivation, anderen etwas beizubringen, bei wiederholtem Einsatz zeigt⁵. Der größte Nachteil dieser Methode liegt aber in dem hohen Aufwand, der mit der Konzeption der Materialien verbunden ist. Hinzu kommt eine ausführliche Erprobung bis hin zur endgültigen Fassung. Dadurch ist die erstmalige Umsetzung der Methode für den Lehrer im Vergleich zu anderen Unterrichtsmethoden zeitintensiver (vgl. FREY-EILING & FREY 1999), wenngleich eine gewisse Kompensation während der Durchführung stattfinden kann.

2.6.3 Nationale Studien zum kooperativen Lernen

Auch in Deutschland ist das Gruppenpuzzle mittlerweile Inhalt evaluativer Studien. Eine der ersten führten EPPLER/WINTER/HUBER bereits 1985⁶ im Deutschunterricht einer Haupt- und Realschule durch, allerdings ohne Kontrollgruppe. BORSCH et al. (2002) evaluierten seine Effektivität im Sachunterricht der Primarstufe, JÜRGEN-LOHMANN et al. (2001) untersuchten die Auswirkungen an Teilnehmern von zwei Hochschulseminaren. Für den Sek. II-Bereich liegen Erkenntnisse von z. B. BERGER & HÄNZE (2004) aus der Jahrgangsstufe 12 im Fach Physik vor. Zumindest BORSCH et al. (2002) bestätigen einen Vorteil des Gruppenpuzzles bezüglich des deklarativen Wissens. An der Hochschule und im Physikunterricht war der Wissenszuwachs in etwa vergleichbar. BERGER & HÄNZE (2005, S. 152) verglichen das Gruppenpuzzle mit einem Lernzirkel. Dabei schnitten Schüler der Lernzirkelgruppe signifikant besser im Wissenstest ab als die Gruppenpuzzlegruppe.

Bis 2003 existierten für die Chemiedidaktik in Deutschland kaum systematische Untersuchungen über die Auswirkungen von kooperativen Lernformen (EILKS 2003a, S. 53). Eine aktuelle Erhebung von WALPUSKI 2006 beschäftigt sich zwar mit kooperativen Lernfor-

⁵ <http://widawiki.wiso.uni-dortmund.de/index.php/Gruppenpuzzle> (Letzter Zugriff am: 5.4.2007)

⁶ vgl. auch HUBER, G. L. (1985), Pädagogisch-psychologische Grundlagen für das Lernen in Gruppen, Studienbrief 1: Lernen in Schülergruppen, Tübingen: Deutsches Institut für Fernstudien.

men, jedoch unter dem Aspekt der strukturierten Instruktion. Dabei ging es um die Optimierung von experimentellem Arbeiten in Kleingruppen und weniger um die Effektivität im Vergleich zu einer herkömmlich unterrichteten Kontrollgruppe. Seine Arbeit basiert auf einer Untersuchung von RUMANN (2005a), der das Konzept der Gruppenrecherche mit Interaktionsboxen an insgesamt acht Klassen zweier Gymnasien erprobt hatte. Dabei erzielten kooperativ unterrichtete Schüler einen signifikanten Leistungsvorteil (RUMANN 2005b, S. 59) gegenüber der Kontrollgruppe. Allerdings war die Effektstärke des Leistungstests sehr gering (vgl. SUMFLETH et al. 2004, S. 291).

Beschreibungen über Lernerfolge wurden daher vor allem auf Basis von Erfahrungen einzelner Lehrer bzw. im Rahmen Partizipativer Aktionsforschung erstellt (z. B. MARKIC & EILKS 2005). Auch in anderen Fächern findet empirische Unterrichtsforschung in Deutschland bisher nur in geringem Umfang statt (vgl. HELMKE 2006, S. 42). Insofern steht die quantitativ-empirische Forschung über kooperatives Lernen im Chemieunterricht noch am Anfang.

3 Das Gruppenpuzzle

Im Folgenden soll die Methode des Gruppenpuzzles näher vorgestellt werden.

3.1 Einführung

Das Gruppenpuzzle, auch Jigsaw Classroom oder Jigsaw-Methode genannt, wurde 1971 vom Sozialpsychologen ELLIOT ARONSON und Mitarbeitern entwickelt, um auf die Aufhebung der Rassentrennung in Schulen der Stadt Austin, Texas, zu reagieren (vgl. ARONSON et al. 1978 und Homepage⁷). Erstmals trafen Schüler unterschiedlicher ethnischer Gruppierungen aufeinander, die vor allem sehr divergierende Lernvoraussetzungen mitbrachten. Hinzu kamen ein extremer Leistungsdruck und Konkurrenzkampf der Schüler untereinander. Als Folge dessen entwickelten sich massive Probleme in Form von Gewalt in der Schule. Eine Analyse des Unterrichts durch ARONSON und Mitarbeiter ergab, dass der stark konkurrenzbetonte Unterricht ursächlich für die Probleme war und zugunsten eines kooperativeren Unterrichts verändert werden musste: Im Kontext dieser Umstände entstand das Gruppenpuzzle. Nach acht Wochen Einsatz des Gruppenpuzzles waren bereits deutliche Fortschritte zu erkennen: Im Vergleich zu traditionell unterrichteten äußerten Gruppenpuzzle-Schüler weniger Vorurteile und Vorbehalte gegenüber ihren Kollegen, waren selbstbewusster und gingen lieber zur Schule.

3.2 Allgemeiner Ablauf

Namensgebend für das Gruppenpuzzle (jigsaw = engl.: Laubsäge) ist die Aufteilung („Zerlegen“) eines Oberthemas in kleinere Teilthemen (sog. Puzzlestücke). Diese werden in insgesamt drei Phasen von den Schülern arbeitsteilig erarbeitet und vermittelt, so dass am Ende jeder Schüler einen vollständigen Überblick (zusammengesetztes Puzzle) über alle Themen hat:

- Erarbeitungsphase: Jeder Schüler erarbeitet selbstständig sein Thema in Form vorstrukturierter Lernmaterialien in sog. Stammgruppen á ca. 3-5 Schüler.
- Expertenphase: Alle „Experten“ für das jeweilige Thema setzen sich zusammen und klären offene Fragen. Außerdem soll jeder Schüler einen strukturierten

⁷ <http://www.jigsaw.org/history.htm> (Letzter Zugriff am: 11.4.2007)

Vortrag ausarbeiten, mit dem er sein Wissen in der nächsten Phase seinen Mitschülern vermitteln möchte.

- **Vermittlungsphase:** Zurück in den Stammgruppen vermitteln sich die Schüler ihr Teil-Wissen gegenseitig. Somit sind die Schüler für ihren Lernerfolg und den ihrer Mitschüler verantwortlich.

Zum detaillierten Ablauf dieses Gruppenpuzzles siehe auch Kapitel 3.3.4 sowie Abbildung 1.

Ablauf des Gruppenpuzzles

Phase	Gruppentyp	Gruppeneinteilung
Erarbeitungsphase ca. 30 min	Stammgruppe à 3 SuS	
Expertenphase ca. 2 x 15 min	Expertengruppe (max. 5 SuS)	
Vermittlungsphase ca. 30 + 45 min	Stammgruppe	

Abbildung 1: Allgemeiner Ablauf des Gruppenpuzzles beispielhaft für 26 Schüler. In Grau jeweils doppelt besetzte Themen.

3.3 Das Gruppenpuzzle Seifen

Für die im Rahmen dieser Arbeit durchgeführte Untersuchung wurde ein Gruppenpuzzle zum Thema Seifen erstellt.

3.3.1 Die Voraussetzungen

Das Gruppenpuzzle Seifen wurde für die Jahrgangsstufe 10 an Gymnasien konzipiert. Ausschlaggebend für das Thema Seifen war, dass die Schüler in der 10. Jahrgangsstufe hierüber relativ wenig Vorwissen haben, was für die Untersuchung von Vorteil war (vgl. TEPNER/MELLE/ROEDER 2005c, S. 151). Lediglich Wasserstoffbrückenbindungen sowie das Lösen von Stoffen in Wasser sollten bekannt sein. Ein weiterer Vorteil bestand in der leichten Entsorgbarkeit der ungefährlichen und zugleich preiswerten „Chemikalien“. Aus lerntheoretischer Sicht relevant war zudem die Berücksichtigung von Alltags-, lebensweltlichen und alternativen Vorstellungen der Schüler, die STORK (1995 S. 20) im Sinne konstruktivistischen Chemieunterrichts fordert. In seinem Artikel rät er weiterhin zu „kognitiven Konflikten“, um den „Übergang von lebensweltlichen zu wissenschaftlichen Konzepten“ (ebd.) zu fördern, und bezieht sich dabei vor allem auf BODNER (1986), GLYNN, YEANY und BRITTON (1991), DUIT (1993) sowie GLYNN und DUIT (1995). Berücksichtigung findet dieses z. B. durch den Pfefferversuch sowie die verschmutzten Armaturen im Badezimmer.

Die Aufteilung in nur drei Unterthemen (s. o.) erfolgte, weil erstens die Gruppengröße aufgrund des Untersuchungsdesigns teilweise recht gering war (11 – 19 Schüler, Aufteilung der Klasse, vgl. Kapitel 5.4.1) und ansonsten die Gruppenanzahl für die Untersuchung zu klein geworden wäre. Zweitens sollte das Gruppenpuzzle in 3 x 45 Minuten durchzuführen sein. Es wurde darauf geachtet, dass alle Unterthemen etwa gleich schwer waren und pro Thema mindestens ein Versuch durchgeführt werden musste. Die Schülerexperimente wurden vorgegeben, da eine eigenständige Entwicklung durch die Schülergruppe zwei Nachteile mit sich bringt: Erstens ist diese Methode sehr viel zeitaufwendiger. Zweitens birgt die selbstständige Entwicklung ein hohes Fehlerpotenzial, z. B. durch falsche Ansätze, fehlerhafte Durchführung oder inkorrekte Interpretationen, so dass ein korrektes Endergebnis nicht immer sichergestellt werden kann (vgl. SUMFLETH et al. 2004, S. 293-294). Ferner stand die Entwicklung eines Experiments bei dieser Untersuchung nicht im Vordergrund. Die Experimente fungierten als Einstiegs- bzw. Bestätigungsexperiment für die Experten und als Anschauungsexperiment für die Novizen.

3.3.2 Expertenmaterial

Die Lernmaterialien wiesen eine klare Strukturierung auf, durch eine Kopfzeile, in der Name und Klasse angegeben werden mussten, wurde eine erhöhte Verbindlichkeit erzielt. Besonders wichtige Sachverhalte wurden unterstrichen bzw. als Merksatz formuliert. Die Veranschaulichung durch Bilder/Zeichnungen lockerte den Text, der in maximal 20 Minuten zu erarbeiten sein sollte, auch optisch auf (siehe Abbildung 2, vgl. TEPNER & MELLE 2005). Zusätzlich befanden sich die Lernziele sowie eine Auflistung der wichtigen Begriffe auf dem Expertenmaterial. Bei der Bearbeitung des Stoffes sollten die Schüler plastische Lernziele vor Augen haben, sie „müssen ganz klar wissen, was Sie von ihnen am Schluss erwarten“ (FREY-EILING & FREY 1999, S. 54). Bei der Formulierung der Lernziele diente ein Gruppenpuzzle der ETH Zürich als Vorlage (vgl. BAUMANN & BLÄTTLER 1994).

Mit Hilfe der wichtigen Begriffe sollten die Schüler sich auf ihren Vortrag vorbereiten, die Reihenfolge konnte ihnen als Strukturierungshilfe dienen. Insgesamt wurde auf ein funktionelles Layout sowie eine angemessene, verständliche Sprache Wert gelegt (vgl. BALLSTAEDT 1993).

Datum:	Thema 1	Vorname:
Klasse:	Expertenmaterial	Erster Buchstabe des Nachnamens:

Aufbau und Löseverhalten von Seife/Waschwirkung

Experiment:
Bedecke eine Wasseroberfläche mit Hilfe eines Pfefferstreuers leicht mit gemahlener Pfeffer und gib anschließend ein kleines Stück Kernseife in die Mitte der Wasseroberfläche. Beobachte!
Im Folgenden wollen wir dieses kleine Schauspiel erklären. Dazu müssen wir jedoch zunächst einiges über den chemischen Aufbau von Seife wissen:

Das Seifenteilchen:
Ein Seifenteilchen besteht aus einem langen, unpolaren Teil und einem kurzen, polaren Anteil. Das Aussehen ähnelt dem eines Streichholzes (siehe Abb. 1). Der ungeladene, unpolare Schwanz löst sich nicht in polarem Wasser, daher ist dieser Bereich „hydrophob“ = wassermeidend. Der negativ geladene, polare Kopf löst sich hingegen gut in Wasser, er ist „hydrophil“ = wasserliebend. Allein er sorgt dafür, dass sich das gesamte Seifenteilchen im Wasser löst. Merke dir folgende Löslichkeits-Regel:

Abb. 1: Streichholzmodell

Polare Stoffe lösen sich in polaren Stoffen, unpolare Stoffe in unpolaren

Verhalten von Seife an der Wasseroberfläche:
Wird Seife in Wasser gegeben, so richtet sich nach dem Streichholzmodell der hydrophile, polare Kopf zum Wasser hin aus, der hydrophobe, unpolare Streichholzstiel wendet sich hingegen vom Wasser ab. Daher stehen die Seifenteilchen senkrecht nebeneinander auf der Wasseroberfläche (siehe Abb. 2). Es bildet sich eine monomolekulare Schicht aus, die viel Platz braucht.

Abb. 2: Seifenteilchen an der Wasseroberfläche

Erklärung der Versuchsbeobachtungen:
Beim Ausbreiten der Seifenteilchen werden die Pfefferkörnchen „beiseite geschoben“, da sich die Kernseife (zunächst einmal) nur an der Wasseroberfläche verteilt.

2

Verhalten von Seife im Wasser:
Erst wenn die Seifenteilchen die gesamte Wasseroberfläche besetzt haben und kein Platz mehr an der Oberfläche ist, bilden sie im Wasser so genannte Mizellen. Das sind Kugeln, bei denen alle unpolare Schwänze nach innen und alle polare Streichholzköpfe nach außen zeigen (siehe Abb. 3). Insgesamt ist die Mizelle außen polar und damit in Wasser löslich.

Abb. 3: Vereinfachte Darstellung einer Kugelmizelle

Was macht die Seife mit dem Schmutz?
Um dreckige Wäsche sauber zu waschen, muss der Schmutz im Wasser gelöst werden. Wie ihr sicherlich wisst, ist z. B. Öl nicht in Wasser löslich, sonst könnte man sich ölige Finger auch ganz einfach mit Wasser waschen. Der Grund liegt darin, dass öliger Schmutz unipolar, also hydrophob ist. Und genau deshalb nimmt ihr zum Waschen Seife: Gemäß der Löslichkeitsregel lagert sich der unpolare Schwanz eines Seifenteilchens an den Schmutz. Gleichzeitig ragt der polare Kopf in das Wasser. Da sich ganz viele Seifenteilchen auf dem Öl bzw. Schmutzteilchen niederlassen, überziehen sie es schließlich mit einer neuen polaren Oberfläche, so dass sich auch unpolare Stoffe wie Schmutz in Wasser lösen (siehe Abb. 4).

Abb. 4: Gelöster Schmutz in Wasser

Lernziele:

- Du weißt, wie ein Seifenteilchen aufgebaut ist.
- Dir ist klar, was passiert, wenn man ein bisschen Seife in Wasser gibt.
- Du kannst erklären, was Mizellen sind und ab wann sie sich bilden.
- Du weißt, warum sich öliger Schmutz bzw. Öl nicht in Wasser löst.
- Du hast verstanden, warum Schmutz von Seife in Wasser gelöst wird.

Wichtige Begriffe:
polar/unipolar, hydrophob/hydrophil, Streichholzmodell, Kopf, Schwanz, Löslichkeits-Regel, senkrecht, monomolekulare Schicht, viel Platz, Pfefferföhen werden beiseite geschoben, Mizelle, öliger Schmutz bzw. Öl, neue polare Oberfläche.

Abbildung 2: Expertenmaterial - Beispiel für Thema 1.

3.3.3 Arbeitsblätter

Zu jeder Phase (Erarbeitungs-, Experten-, Vermittlungsphase) wurden Arbeitsblätter ausgeteilt. Diese enthielten eine kurze Erklärung der jeweiligen Phase und die dazugehörigen Arbeitsaufträge. Das Arbeitsblatt „Erarbeitungsphase“ enthielt zusätzlich eine Lernkontrolle, das Arbeitsblatt „Vermittlungsphase“ nochmals Lernziele/wichtige Begriffe und eine kurze Versuchsvorschrift (siehe Abbildung 3).

Datum:	Thema 1	Vorname:
Klasse:	Arbeitsblatt Vermittlungsphase	Erster Buchstabe des Nachnamens:

Aufbau und Löseverhalten von Seife/Waschwirkung

Geht in eure alten Gruppen zurück. In der Vermittlungsphase erklärt jeder Schüler seinen Mitschülern das von ihm erarbeitete Thema. Bei deinen Erklärungen sollst du nicht nur die unten aufgeführten Lernziele berücksichtigen, sondern auch folgende Begriffe verwenden. Benutze auch Stift und Papier zur Erklärung. Sei dir bewusst, dass nur du für den Lernerfolg der anderen verantwortlich bist und die anderen für deinen. Am Ende der Vermittlungsphase sollen alle Schülerinnen und Schüler alle Lernziele sowie die durchgeführten Versuche verstanden haben.

Lernziele:

- Ihr wisst, wie ein Seifenteilchen aufgebaut ist.
- Euch ist klar, was passiert, wenn man ein bisschen Seife in Wasser gibt.
- Ihr könnt erklären, was Mizellen sind und ab wann sie sich bilden.
- Ihr wisst, warum sich öliger Schmutz bzw. Öl nicht in Wasser löst.
- Ihr habt verstanden, warum Schmutz von Seife in Wasser gelöst wird.

Wichtige Begriffe:
polar/unpolar, hydrophob/hydrophil, Streichholzmodell, Kopf, Schwanz, Löslichkeits-Regel, senkrecht, monomolekulare Schicht, viel Platz, Pfefferteilchen werden beiseite geschoben, Mizelle, öliger Schmutz bzw. Öl, neue polare Oberfläche.

Arbeitsaufträge:

1. Erklärt euch gegenseitig in ca. 15 min eure Themen. Verwendet dabei die wichtigen Begriffe und achtet auf die Lernziele. Es ist sinnvoll, die Erklärungen der Mitschüler schriftlich festzuhalten.
2. Führt gemeinsam den Pfeffer-Versuch und folgendes Experiment durch und beantwortet die dazugehörigen Fragen.

Experiment:
In zwei Reagenzgläsern befindet sich jeweils ein Öl/Wasser-Gemisch. In das leerere Reagenzglas gibst du soviel Seifenlösung hinzu, bis fast keine Luft mehr drin ist. Verschließe mit einem Stopfen, schüttele beide kurz (Daumen auf Stopfen!) und lasse stehen.

Fragen:

- Welche Beobachtungen erwartest du?
- Wie kannst du dieses Verhalten erklären?
- Würde sich Seife auch in Öl (ohne Wasser!) lösen? Begründe! Falls ja: Wie würden die Mizellen dann aussehen?

Abbildung 3: Arbeitsblatt Vermittlungsphase - Beispiel für Thema 1.

Die Schülerversuche wurden durch den Untersuchungsleiter vorbereitet und zur jeweiligen Untersuchung mitgebracht. Als hilfreich erwiesen hat sich dafür eine Kiste (siehe Abbildung 4), die sämtliche Versuche inklusive aller „Chemikalien“ sowie eine Inhalts-

übersicht enthielt, anhand derer sehr schnell die Vollständigkeit überprüft werden konnte. Gleiches galt für den Frontalunterricht, in dem die Materialien ebenfalls zur Verfügung gestellt wurden.



Abbildung 4: „Einsatzkiste“ mit Schülerversuchen.

3.3.4 Durchführung

Für das im Rahmen der Untersuchung eingesetzte Gruppenpuzzle wurde das Oberthema „Seifen“ in drei Unter-Themen („Aufbau und Löslichkeit von Seife/Waschwirkung“, „Herabsetzung der Oberflächenspannung durch Seife“ und „Was ist Seife und die Nachteile von Seife“) aufgeteilt. Jeder Schüler erarbeitete selbstständig sein Thema in Form vorstrukturierter Lernmaterialien in sog. Stammgruppen á 3-5 Schüler („Erarbeitungsphase“). In der darauf folgenden „Expertenphase“ setzten sich alle „Experten“ für das jeweilige Thema zusammen und klärten offene Fragen. Der Lehrer durfte nur im Notfall gefragt werden. Außerdem sollte jeder Schüler einen strukturierten Vortrag ausarbeiten, mit dem er sein Wissen seinen Mitschülern in der nächsten Phase vermittelte. Eine Lernkontrolle, transparente Lernziele sowie Protokollbögen (siehe Anhang), welche durch die Schüler während der Expertenphase auszufüllen waren, sollten sicherstellen, dass jeder Schüler dazu in der Lage war.

In der letzten Phase, der „Vermittlungsphase“, gingen die Schüler wieder zurück in ihre Stammgruppen und lehrten sich ihr Teil-Wissen gegenseitig. Am Ende sollte jeder Schüler einen möglichst vollständigen Überblick über das Thema Seifen haben.

Das ausgehändigte Expertenmaterial durfte nicht an die Mitschüler weitergegeben werden. Eine abschließende Besprechung der Aufgaben oder eine Zusammenfassung im Unter-

richtsgespräch fand beim Gruppenpuzzle nicht statt, um Effekte des Gruppenpuzzles nicht durch den Eingriff der Lehrperson zu verfälschen.

Die Abfolge Stammgruppe – Expertengruppe – Stammgruppe wurde hier wegen des Vorteils gewählt, dass jede Gruppenpuzzlephase für die Schüler zeitlich und räumlich abgegrenzt war (vgl. TEPNER/MELLE/ROEDER 2005b, S. 82). Ein vierjähriger Modellversuch der BAG (2005, S. 59) hat zudem gezeigt, dass *„Bewegung im Unterricht die Konzentration und die Arbeitsleistung fördert“*.

4 Stichprobe

4.1 Stichprobenmerkmale und Auswahl der Stichprobe

Aufgabe einer guten Stichprobe ist es, die Grundgesamtheit (Population) möglichst gut zu repräsentieren. Als Population bezeichnet BORTZ (2005, S. 86) „[...] *alle potenziell untersuchbaren Einheiten oder „Elemente“, die ein gemeinsames Merkmal (oder eine gemeinsame Merkmalskombination) aufweisen*“. Im vorliegenden Fall sind dieses alle Chemie-Schüler der 10. Klasse an nordrhein-westfälischen Gymnasien.

Eine Stichprobe „*stellt eine Teilmenge aller Untersuchungsobjekte dar, die die untersuchungsrelevanten Eigenschaften [hier z. B. Lernzuwachs] der Grundgesamtheit [Population] möglichst genau abbilden soll*“ (BORTZ 2005, S. 86). Am besten wählt man diese per Zufall aus – z. B. per Los; man spricht von „Zufallsstichproben“ (BORTZ 2005, S. 86f). Da dieses in der Praxis jedoch häufig nicht möglich ist, z. B. weil pro Person ein zu großer Aufwand notwendig wäre, greift man aus organisatorischen Gründen auf zufällig auszuwählende Teilmengen zurück, die bereits vorgruppiert sind, z. B. Schulkassen. Eine derartige, in diesem Fall aus mehreren Schulklassen bestehende Stichprobe wird als „Klumpenstichprobe“ bezeichnet (BORTZ 2005, S. 87) und auch in dieser Untersuchung verwendet.

Streng genommen handelt es sich um zwei Stichproben aus zwei Grundgesamtheiten (Populationen), wenngleich der Begriff Grundgesamtheiten in der Literatur nicht einheitlich gehandhabt wird. Daher wird in dieser Arbeit auf BORTZ (2005, S. 86 & 139) bzw. BORTZ/DÖRING (2003, S. 131) Bezug genommen.

Die Grundgesamtheit „Chemie-Schüler der 10. Klasse an nordrheinwestfälischen Gymnasien“ muss nach BORTZ (ebd.) weiter differenziert werden in

- a) Chemie-Schüler der 10. Klasse an nordrheinwestfälischen Gymnasien, die frontal unterrichtet werden und
- b) Chemie-Schüler der 10. Klasse an nordrheinwestfälischen Gymnasien, die nach der Gruppenpuzzlemethode unterrichtet werden.

Aus beiden Populationen werden dann Stichproben gezogen, die voneinander unabhängig sind. Als unabhängig bezeichnet man Stichproben dann, wenn kein Zusammenhang zwischen den Objekten (Probanden) der einen Stichprobe und den Objekten (Probanden) der

anderen Stichprobe bestehen, so dass eine Beeinflussung der Ergebnisse durch die jeweils andere Stichprobe entfällt (vgl. BORTZ 2005 S. 143). Vereinfachend kann man sagen, dass unabhängige Stichproben immer dann vorliegen, wenn diese unterschiedliche Probanden enthalten (ZÖFEL 2002, S. 90).

Daher soll im Folgenden die FU (Frontalunterricht)- bzw. die GP (Gruppenpuzzle)-Gruppe getrennt ausgewiesen werden. Die Aufteilung der Schüler in die jeweilige Gruppe wird in Abschnitt 5.4 beschrieben.

Im Rahmen der Untersuchung fand aus mehreren Gründen eine Einschränkung auf die Schulform Gymnasium und die Jahrgangsstufe 10 statt. Allerdings wurde die Studie im Jahr 2005 um drei weitere Schulformen ergänzt, weswegen letztere im Folgenden als „externe“ Hauptuntersuchung bezeichnet werden soll (zu den Bezeichnungen „intern“ und „extern“ siehe auch Abschnitt 5.4).

Es gibt mehrere Gründe für die Wahl der Jahrgangsstufe 10: Erstens sollten die Schüler ein möglichst geringes chemisches Vorwissen aufweisen, um Vorwissenseffekte zu minimieren und einen ausreichend großen Wissenszuwachs zu ermöglichen. Auf der anderen Seite sollten sie bereits selbstständig auch anspruchsvollere Texte erarbeiten können. Drittens sollte die Beliebtheit des Faches Chemie erfasst werden, was nach der 9. Klasse und kurz vor den Wahlen in der Oberstufe (an Gymnasien) von besonderer Bedeutung ist. Die Jahrgangsstufe 10 bot sich auch deshalb besonders an, da hier ein Thema aus der Organischen Chemie, das geringere Vorkenntnisse erfordert, obligat ist. Schließlich existieren in der Oberstufe meist kleinere Kurse, die für das eingesetzte Untersuchungsdesign nicht geeignet sind.

Das Gymnasium war deshalb besonders geeignet für die Untersuchung, da im Gegensatz zu Real- und Hauptschulen in NRW an jeder Schule gleich viel Chemie unterrichtet worden sein sollte. Durch verschiedene Vertiefungskurse z. B. an der Realschule lässt sich dieses nicht für alle Schüler gleichermaßen sicherstellen. Hinzu kommt bei Gesamtschulen und Berufskollegs das stärkere Leistungsgefälle, was einen Vergleich der verschiedenen Schüler und auch der Schulen untereinander erschwert.

Die Auswahl der Stichprobe erfolgte auf zwei Arten: Einerseits wurden Teilnehmer von Lehrerfortbildungen in Dortmund direkt angesprochen, ob sie Interesse an einer Mitarbeit hätten. Andererseits wurden für die erste Hauptuntersuchung 29 Lehrer per E-Mail, deren Adressen aus einem Verteiler des Chemielehrerfortbildungszentrums stammten, ange-

schrieben. Zusätzlich erfolgte eine Rekrutierung über „Mundpropaganda“ der bereits involvierten Lehrer. Insofern kann von einer zufälligen Auswahl der Probanden und damit der gesamten Stichprobe nur bedingt gesprochen werden. Da allerdings unmotivierte Lehrer mindestens genauso wenig repräsentativ sind wie motivierte, wovon bei dem o. a. Klientel ausgegangen werden kann, und andererseits die Schüler die eigentlichen Probanden darstellen, erscheint diese Stichprobenauswahl legitim. Hinzu kommt die Überlegung, dass „gute“ Lehrer auch eher „guten“ Frontalunterricht erteilen, was das Ergebnis dieser Untersuchung eher zu Ungunsten des Gruppenpuzzles verschieben dürfte. Trotzdem können die Befunde streng genommen nur auf eine Teilmenge der Grundgesamtheit generalisiert werden.

4.2 Beteiligte Schulen

An der Voruntersuchung nahmen zwei Gymnasien in NRW teil (siehe Tabelle 2), davon eins mit naturwissenschaftlichem Schwerpunkt. Auch die Schulen der ersten und zweiten Hauptuntersuchung (intern) lagen in NRW, allerdings ziemlich weit auseinander (siehe Abbildung 5), so dass sich der durch regionale Besonderheiten bedingte Fehler minimieren dürfte. Ein solcher Fehler tritt auf, wenn die per Zufall auszuwählende Stichprobe typische gemeinsame Merkmale aufweist, hier z. B. Schulen ausgewählt worden wären, die alle auf dem Land bzw. in der Stadt liegen. Insofern stellt die Auswahl der Stichprobe einen Kompromiss zwischen größtmöglicher Variation und räumlicher Nähe dar.

Die weiter entfernten Schulen in Niedersachsen (Sulingen) und Hessen (Wehretal) wichen aus organisatorischen Gründen vom Untersuchungsdesign ab, siehe dazu Abschnitt 5.4.2.

Anmerkung zu Tabelle 2: Ein Gymnasium in Dortmund (1) ist mit einer Lehrkraft zweimal aufgeführt, da diese sowohl im Jahr 2004 als auch 2005 teilgenommen hat. Aufgrund des Untersuchungsdesigns, das ein Unterrichten von Kontroll- als auch Interventionsgruppe durch dieselbe Lehrkraft vorsieht, kann davon ausgegangen werden, dass eine mehrmalige Teilnahme auf die Ergebnisse keinen Einfluss hat.

Tabelle 2: Beteiligte Schulen und Untersuchungsorte. Abkürzungen: Gym = Gymnasium, HS = Hauptschule, BK = Berufskolleg, RS = Realschule, mp = matched pairs, VU = Voruntersuchung, HU = Hauptuntersuchung. Zu „extern“ und „matched pairs“ siehe Abschnitt 5.4.

Schulort	Schulform	Klassen	Lehrer	Design	Projektteil
1. Lüdenscheid	Gym.	2	1	mp	VU
2. Meschede	Gym.	1	1	mp	VU
3. Kaarst	Gym.	2	1	mp	HU I
4. Marl	Gym.	3	2	mp	HU I
5. Dortmund 1	Gym.	2	1	mp	HU I
6. Hagen	Gym.	2	1	mp	HU II
7. Werne	Gym.	2	1	mp	HU II
8. Dortmund 2	Gym.	2	2	mp	HU II
9. Dortmund 2	Gym.	1	1	mp	HU II
10. Sulingen	Gym.	2	2	mp	HU extern
11. Reken	HS	3	1	1 x mp 2x parallel	HU extern
12. Kleve	BK	3	1	1 x mp 2x parallel	HU extern
13. Wehretal-Reichensachsen	RS	2	1	parallel	HU extern
14. Unna	BK	1	1	mp	HU extern



Abbildung 5: Karte mit Untersuchungsorten. Quelle: map24

4.3 Stichprobengröße

Insgesamt standen für die Untersuchung 23 Vergleichspaare (siehe Tabelle 8) mit insgesamt 667 Schülern zur Verfügung, von denen allerdings nur 648 Schüler teilgenommen haben. Zu erklären ist die Differenz mit Klassen- bzw. Schulabgängern, die vor Beginn der Untersuchung zwar eingeplant – da in den Klassenlisten aufgeführt – waren, jedoch an keinem der Tests partizipiert haben. Diese „Soll-Ist-Abweichung“ liegt bei ca. 2.85 % und wird vor allem durch die vier Klassen des Berufskollegs verursacht. Bei ihnen war die Quote der Abgänger besonders hoch.

Zu diesem organisatorischen „Schülerschwund“ kommen noch krankheitsbedingte Ausfälle hinzu, was die Zahl der tatsächlich auswertbaren Probanden noch einmal um ca. 1 % verringert, abhängig von der Testart. Für Residuen z. B. sind sowohl Vor- als auch Nachtest erforderlich, siehe Abschnitt 7.1.

Eine Übersicht über die Stichprobengrößen der einzelnen Untersuchungen bietet Tabelle 3.

Tabelle 3: Umfang der Stichproben. In Klammern die laut Klassenlisten teilnehmenden Schüler. Abkürzungen: m = männlich, w = weiblich.

	VU	HU I	HU II _{intern}	HU II _{extern}
Lehrer	2 m: 1, w: 1	4 m: 1, w: 3	5 m: 2, w: 3	6 m: 6, w: 0
Klassen	3	7	7	11
Schüler	74 (74) m: 35 w: 39	158 (161) m: 86 w: 72	181 (182) m: 81 w: 100	235 (250) m: 104 w: 131
durchschnittliches Alter der Schüler [Jahre]	15.8	16.1	15.9	16.7

4.4 Verteilung von Alter und Geschlecht der Schüler

Wie aus Tabelle 3 hervorgeht, waren die beteiligten Schüler ca. 16 Jahre alt, der Altersdurchschnitt in der externen Hauptuntersuchung lag etwas höher, da vier Klassen zweier Berufskollegs teilgenommen haben. Teilweise waren die Schüler hier deutlich älter als 16 Jahre, besaßen allerdings vergleichbare chemische Vorkenntnisse, wie aus den Vortester-

gebnissen hervorgeht (siehe Tabelle 8). Beide Geschlechter waren mit Ausnahme der externen Hauptuntersuchung etwa gleichstark vertreten: 202 Jungen vs. 211 Mädchen. Die ungleiche Jungen-Mädchen-Relation der externen Hauptuntersuchung (siehe Tabelle 3) wird ausschließlich durch die drei Klassen des Berufskollegs Kleve verursacht. Der Jungen-Anteil lag hier bei lediglich 24 %.

4.5 Lehrkräfte

Alle beteiligten Lehrer konnten eine mindestens 5jährige Schulpraxis aufweisen, der Altersdurchschnitt lag bei ca. 45 Jahren. Ungefähr die Hälfte der Lehrkräfte kannte die Unterrichtsmethode Gruppenpuzzle vor Beginn der Untersuchung, das Thema Seifen hatten alle bis auf wenige Ausnahmen bereits zuvor in anderen Lerngruppen unterrichtet. Das Lehrer-Lehrerinnen-Verhältnis betrug 10:7.

4.6 Einordnung

Bis auf zwei Schulen (Dortmund 2 und Sulingen) haben alle Lehrer ihre Schulen und damit ihre Schüler als „guten Durchschnitt“ bezeichnet. Die Schule in Sulingen ist naturwissenschaftlich orientiert, z. B. nahm an der Untersuchung eine Klasse teil, der als NW-Klasse in der Jahrgangsstufe 9 drei statt zwei Stunden Chemie pro Woche erteilt wurden. Das zweite Dortmunder Gymnasium steht in der Trägerschaft des Erzbistums Paderborn und weist einige Besonderheiten auf, die möglicherweise die Ergebnisse beeinflussen könnten, allerdings in Kontroll- und Interventionsgruppe gleichermaßen:

- kaum Unterrichtsausfall durch zusätzlichen Bereitschaftsdienst der Lehrer;
- zusätzliche Förderung des Experimentalunterrichts durch eine weitere Stunde Chemie in der Klasse 10;
- Naturwissenschaften werden blockweise erteilt, d. h. durchgängig von Klasse 8 – 10;
- regelmäßige Teilnahme an Wettbewerben;
- schnelle Reaktion auf Unregelmäßigkeiten und nicht angepasstes Verhalten durch Mitteilungen an die Eltern und durch Elterngespräche;
- mehr Elternsprechtage als üblich;
- Schulvertrag mit Eltern, dass diese ihrer Erziehungspflicht nachkommen.

Die Lehrer der Berufskollegklassen ordneten ihre Schüler als für Berufskollegs typisch, im Vergleich zu Gymnasien aber deutlich schwächer ein. Speziell das kaufmännisch orientier-

te Berufskolleg in Unna weist laut Lehrerauskunft eine stark heterogene Schülerklientel mit sehr unterschiedlicher Vorbildung auf. Nicht ganz so stark gilt Selbiges für das Berufskolleg Kleve. Seine Teilnehmer waren Absolventen des Bildungsganges Sozial- und Gesundheitswesen. Grundsätzlich war die Motivation an den anderen Schulformen bedingt durch die Bewerbungen für einen Ausbildungsplatz mit Hilfe des Abschlusszeugnisses sehr gut. Aufgrund der laut Lehrerauskunft (s. o.) durchschnittlichen Schulen und der großen Schülerzahl an 13 relativ gut verteilten Standorten ist von einer repräsentativen Stichprobe (Gymnasialklassen in NRW) auszugehen.

5 Untersuchungsdesign

5.1 Projektverlauf

Im April 2003 wurden 35 Lehrer angeschrieben, von denen sich sieben bereit erklärten, an der Untersuchung mitzuwirken. Eine Lehrkraft musste kurzfristig aus gesundheitlichen Gründen absagen, die anderen Absagen wurden mit unpassender Themenwahl bzw. mangelnden 10er Klassen begründet.

Vor Beginn der Voruntersuchung wurden die konzipierten Meinungstests auf ihre Lesbarkeit, Verständlichkeit sowie das Layout überprüft (Voruntersuchung 0), was mit Hilfe einer 10. Klasse eines Gymnasiums in Lünen im Januar 2004 erfolgte. Eine Übersicht über den Projektverlauf gibt Abbildung 6, zu Einzelheiten siehe Abschnitt 5.4.

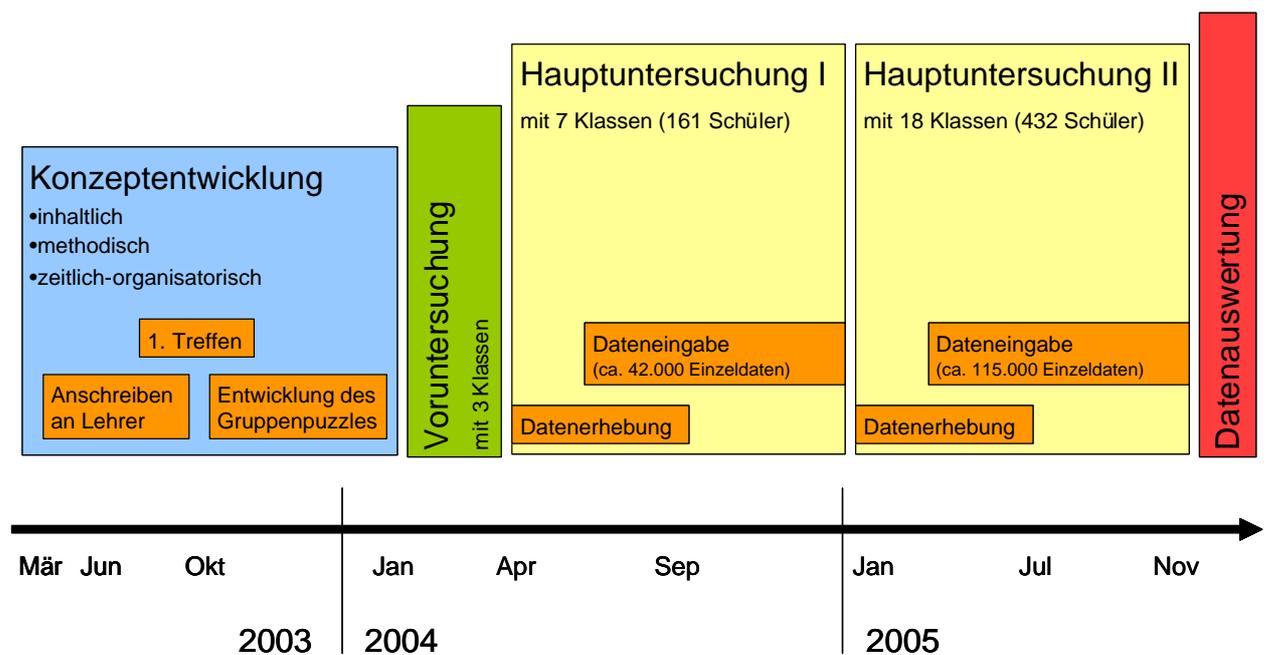


Abbildung 6: Übersicht über den Projektverlauf von 2003 – 2005.

5.2 Konzeption

Bei der durchgeführten Studie handelt es sich um eine Methodenevaluation. Unter Evaluation versteht ROEDER (1998, S. 252) „eine wissenschaftliche Wirkungsanalyse, an deren Ende eine Entscheidung darüber getroffen wird, ob eine Maßnahme durchgeführt werden soll oder nicht“.

Die experimentelle Untersuchung wurde als Retroaktionsplan (vgl. MASENDORF 1997) konzipiert. Darunter versteht man die Veränderungsmessung einer Experimentalgruppe im Vergleich zu einer Kontrollgruppe (siehe Abbildung 7).

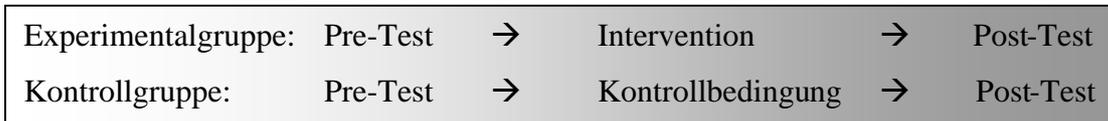


Abbildung 7: Allgemeines Schema des Untersuchungsablaufs

ROEDER (1998) hält folgende Punkte bei einem Evaluationsexperiment für beachtenswert:

- a) Das Prinzip der isolierten Variation
- b) Kontroll- und Experimentalgruppe
- c) Vor- und Nachtest
- d) Niveaubereinigte Veränderungsmessung (Residuen)

Isolierte Variation bedeutet, dass der zu untersuchende Faktor (Unterrichtsmethode) möglichst unabhängig von Störvariablen (z. B. Intelligenz, Motivation, sozialer Status) untersucht wird, indem man nur diesen Faktor (Unterrichtsmethode) variiert und alle anderen Faktoren (z. B. Vorkenntnisse, Lehrer, Geschlecht, Inhalte) konstant hält. Eine Minimierung dieser unerwünschten Einflüsse gelingt durch Randomisierung (vgl. BORTZ/DÖRING 2003, S. 58; zum Randomisierungsverfahren siehe Abschnitt 5.4). Darunter versteht man die zufällige Zuteilung der Untersuchungspersonen zu den Untersuchungsbedingungen (Kontroll- bzw. Interventionsgruppe, vgl. BORTZ 2005), so dass diese intervenierenden Variablen neutralisiert werden. Insofern handelt es sich beim vorliegenden Versuchsdesign um eine experimentelle Untersuchung, da hier zufällig zusammengestellte Gruppen miteinander verglichen werden. Nach BORTZ (2005) ist die Randomisierung eine unabdingbare Bedingung gemäß der drei zentralen Gütekriterien: *Objektivität*, *Reliabilität* und *Validität* einer Untersuchung (siehe Abschnitt 5.3).

Im Gegensatz dazu vergleicht eine quasiexperimentelle Studie natürliche Gruppen (vgl. BORTZ/DÖRING 2003, S. 57), so dass relevante Störvariablen nicht ausgeschaltet werden können (vgl. HUBER 2005, S. 80).

Vor- und Nachtest einer Untersuchung stellen sicher, dass etwaige Effekte tatsächlich auf die Interventionsmaßnahme zurückzuführen sind und nicht durch im Vorfeld bereits bestehende Unterschiede verursacht werden. Dementsprechend erfolgte die Auswertung über

eine niveaubereinigte Veränderungsmessung (Residuen), worauf in Abschnitt 7.1 näher eingegangen werden soll.

5.3 Zentrale Gütekriterien und deren Einhaltung

5.3.1 Objektivität

Nach BORTZ/DÖRING (2003, S. 194) ist ein Test objektiv, „*wenn verschiedene Testanwender bei denselben Personen zu den gleichen Resultaten gelangen, d. h. ein objektiver Test ist vom konkreten Testanwender unabhängig*“. Dabei unterscheiden sie in drei Unterformen:

1. Durchführungsobjektivität
2. Auswertungsobjektivität
3. Interpretationsobjektivität

1. Durchführungsobjektivität

Im Sinne einer hohen Durchführungsobjektivität wurde jeder Test bzw. Fragebogen durch den Untersuchungsleiter persönlich erklärt, ausgeteilt und eingesammelt. Bei der Durchführung wurde darauf geachtet, dass beiden Gruppen die gleiche Instruktion gegeben wurde, um den „Versuchsleiter-Erwartungseffekt“ zu minimieren (vgl. HUBER 2005, S. 184). Diese Erklärungen fanden sowohl vor den Tests als auch vor den Interventionen statt und waren schriftlich fixiert. Während der Tests durften keine weiteren Fragen gestellt werden. Zusätzlich war eine exakte Anleitung zur Durchführung auf jedem Test vorhanden.

Die Tests von Interventions- und Kontrollgruppe wurden zeitgleich bzw. in der gleichen Unterrichtsstunde durchgeführt. Ferner konnte ein Abschreiben der Schüler untereinander weitestgehend verhindert werden (Platz zwischen Schülern frei gelassen, doppelte Aufsicht). Die HU II_{extern} fand zum größten Teil ohne den Untersuchungsleiter statt (siehe Abschnitt 5.4.2).

2. Auswertungsobjektivität

Nach PERLETH 2007⁸ beinhaltet *„die Forderung nach Auswertungsobjektivität [...], daß die Testergebnisse unabhängig von der Person des Auswerters sein sollen“*. Durch das Verwenden von geschlossenen Antworten sowohl im Wissens- als auch im Meinungstest konnte dieser Forderung genüge getan werden.

Anders bei den Kodierungen der Schülerbeobachtungen: Hier werden Verhaltensweisen mit einem Kategoriensystem, welches auf subjektiven Beobachtungen basiert, erfasst. Zwar erhielten die Beobachter im Vorfeld eine „Schulung“, bei der ihnen gesagt wurde, wie sie vorzugehen haben, jedoch bleibt ein gewisser Unsicherheitsfaktor. Dieser konnte auch durch eine Probe-Beobachtung, bei der zwei Beobachter die gleiche Gruppe observierten, so dass ihre Ergebnisse miteinander verglichen werden konnten, nicht gänzlich beseitigt werden. Dieses stellte jedoch kein Problem dar, da eine weitergehende Auswertung der Daten nicht erfolgte (vgl. Kapitel 7.7).

3. Interpretationsobjektivität

Da bei dieser Untersuchung Rohwerte (Scores) als Maß für eine Leistung bzw. Meinung verwendet werden, bleibt kein Spielraum für subjektives Interpretieren. Für beide Testarten gilt: Je höher der Score, desto besser der jeweilige Schüler bzw. desto positiver seine Meinung. Auf diese Weise sind Vergleiche innerhalb der Population ohne Interpretation möglich.

5.3.2 Reliabilität

BORTZ/DÖRING (2003, S. 195) bezeichnen einen Test als reliabel, wenn er *„bei wiederholter Anwendung bei denselben Personen zu exakt den gleichen Ergebnissen führt“*. Die Reliabilität gibt folglich den *„Grad der Genauigkeit an, mit dem das geprüfte Merkmal gemessen wird“* (ebd.), also praktisch seine Zuverlässigkeit.

Zur Berechnung der Reliabilität stehen vier Methoden zur Verfügung:

1. Retest-Reliabilität
2. Paralleltest-Reliabilität
3. Testhalbierungs-Reliabilität
4. Interne Konsistenz

⁸ http://www.phf.uni-rostock.de/institut/ipp/lehmaterialien/diagnostikvorlesung/vd1_3_qualitaetsaspekte.htm#VD1_3_Qualitaetsaspekte
(Letzter Zugriff am 29.7.07)

Um die Retest-Reliabilität zu bestimmen, muss derselbe Test derselben Stichprobe nach einer vordefinierten Zeit noch einmal vorgelegt werden. Das Bestimmtheitsmaß B gibt an, wie viel Prozent der Gesamtunterschiedlichkeit der Testergebnisse auf „wahre“ Merkmalsunterschiede zurückzuführen ist (vgl. BORTZ/DÖRING 2003). Im vorliegenden Fall wird ein relativ kurzer Test mit „*inhaltlich interessanten Items*“ verwendet, so dass die Wahrscheinlichkeit von Erinnerungseffekten sehr hoch ist, welche das Testergebnis beeinflussen würden. Zwar nimmt diese Wahrscheinlichkeit mit der Zeit ab, insgesamt ist der (zeitliche) Aufwand jedoch sehr groß. Daher eignet sich die Retest-Reliabilität zur Qualitätsbestimmung für diese Untersuchung nicht.

Ebenso groß ist der untersuchungstechnische Aufwand bei der Ermittlung der Paralleltest-Reliabilität: Hier werden zwei inhaltlich ähnliche Paralleltests entwickelt und den Probanden direkt aufeinander folgend zur Bearbeitung gereicht. Ein Vergleich der beiden Ergebnisse gibt Auskunft über die Reliabilität der Tests. BORTZ/DÖRING (2003, S. 197) sehen diesen hohen Aufwand nur gerechtfertigt, „*wenn für praktische Zwecke tatsächlich zwei (oder auch mehr) äquivalente Testformen benötigt werden*“. Zwei Testformen a und b wurden bei dieser Untersuchung nicht benötigt, da durch die verringerte Gruppenstärke ohnehin kaum die Möglichkeit zum Abschreiben gegeben ist. Insofern wäre dieser Aufwand nicht gerechtfertigt gewesen.

Bei der Testhalbierungs-Reliabilität berechnet man pro Untersuchungsteilnehmer zwei Testwerte, die auf je einer (zufällig ausgewählten) Hälfte der Items basieren. Die Korrelation der beiden Testhälften-Testwerte entspricht der Testhalbierungs-Reliabilität. Die Art wie die beiden Testhälften generiert werden, beeinflusst die Reliabilität stark (vgl. BORTZ/DÖRING 2003).

Um diesem Nachteil zu entgehen, verwendet man die Interne Konsistenz. Hierbei wird ein Test nicht nur in zwei Hälften zerlegt, sondern in „viele kleinste Teile“; er behandelt die einzelnen Items also wie viele kleine Paralleltests. Das gebräuchlichste Maß für die Interne Konsistenz ist Cronbachs Alpha. Dieser Koeffizient entspricht formal der mittleren Testhalbierungs-Reliabilität eines Tests für alle möglichen Testhalbierungen (vgl. BORTZ/DÖRING 2003).

Um eine möglichst hohe Interne Konsistenz der verwendeten Tests zu gewährleisten, wurde bereits in der Voruntersuchung Cronbachs Alpha bestimmt. Zusätzlich fand eine Itemanalyse statt: Items mit schlechter Trennschärfe wurden umformuliert bzw. aus dem Test

entfernt. Die Distraktorenanalyse erfolgte mittels Chi²-Test. Distraktoren, die zu attraktiv bzw. zu unattraktiv waren, wurden verändert, so dass Cronbachs Alpha für die Hauptuntersuchung auf akzeptable Werte angehoben werden konnte (siehe hierzu Kapitel 6.1). Gute Tests besitzen einen Alpha-Koeffizienten von über 0.8 (vgl. BORTZ/DÖRING 2003, S. 198).

5.3.3 Validität

Unter Validität (Gültigkeit) eines Tests verstehen BORTZ/DÖRING (2003, S. 199) „*wie gut ein Test in der Lage ist, genau das zu messen, was er zu messen vorgibt*“. Zum Beispiel sollte der verwendete Wissenstest tatsächlich Wissen über das Thema Seifen abfragen und nicht allgemein über Chemie; und der Meinungstest sollte auch wirklich die Attraktivität der letzten drei Stunden erfassen und nicht die Person des Lehrers! Beide Beispiele gehören zur so genannten Inhaltsvalidität:

1. Inhaltsvalidität

Sie ist gegeben, „*wenn der Inhalt der Test-Items das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend erfasst*“ BORTZ/DÖRING (2003, S. 199). Dies wurde versucht zu berücksichtigen, indem fast alle Items des Wissenstests zuvor Gelerntes abfragten, also Reproduktionswissen. Die Testentwicklung erfolgte parallel zur Entwicklung des Gruppenpuzzles bzw. der Frontalunterrichtseinheit. Aus jedem Themengebiet wurden etwa gleich viele Items generiert. Ein sehr geringer Anteil an Transferwissen unterstützt die Forderung nach Inhaltsvalidität.

Die Konstruktion der Meinungstest-Items für eine möglichst hohe Inhaltsvalidität war aufwendiger, kann aber auf die wesentlichen Punkte reduziert werden:

- immer komplett ausgeschriebene Items (Bsp.: „Der Unterricht **der letzten 3 Stunden** hat mir Spaß gemacht“)
- Statements wurden sehr exakt formuliert
- keine Abhängigkeit von Personen, sondern nur von der Sache („Ich fand, dass der Lehrer in den letzten 3 Stunden **gut erklärt** hat“ anstatt „Ich fand den Lehrer gut“)
- auf die Zielperson ausgerichtet (Ich-Form) „**Ich** fand die...“
- keine doppelten Verneinungen (siehe hierzu auch Kapitel 6.2.1.).

2. Kriteriumsvalidität

Die so genannte Kriteriumsvalidität lässt sich messen, indem man z. B. den Wissenstest mit einem korrespondierenden Kriterium bzw. Merkmal korreliert. Im vorliegenden Fall waren das Schulnoten, welche sehr gut mit den Testwerten (Residuen) übereinstimmten (siehe Tabelle 13). Bei den Meinungstests konnte die Kriteriumsvalidität nicht kontrolliert werden.

3. Konstruktvalidität

Schließlich ist die Konstruktvalidität anzuführen. BORTZ/DÖRING (2003, S. 200) bezeichnen einen Test als *„konstruktvalide, wenn aus dem zu messenden Zielkonstrukt [z. B. Wissen über das Thema Seifen] Hypothesen ableitbar sind, die anhand der Testwerte bestätigt werden können“*, wie im Fall des Wissenstests. Hier konnte z. B. gezeigt werden, dass das Abschneiden im Wissenstest unabhängig vom Geschlecht ist.

Darüber hinaus wurde zur Prüfung der Konstruktvalidität eine Faktorenanalyse durchgeführt (vgl. PERLETH 2007⁹ und SCHÖLER et al. 2004, S. 8). Sie gibt Aufschluss darüber, wie „stimmig“ die Items des Post-Wissenstests zueinander sind. Als wichtigstes Ergebnis liefert SPSS¹⁰ einen „Screeplot“ (oder auch „Eigenwertediagramm“, siehe Abbildung 8), der *„die Größe der in Rangreihe gebrachten Eigenwerte als Funktion ihrer Rangnummern darstellt“* (BORTZ 2005, S. 544). Unter Eigenwert versteht BROSIUS (2002, S. 689) *„eine Kennzahl dafür, wie „sauber“ die Diskriminanzfunktion zwischen den verschiedenen Kategorien der abhängigen Variablen zu trennen in der Lage ist“*. Ein hoher Eigenwert deutet demnach auf eine hohe Trennschärfe des Erklärungsmodells hin (vgl. BROSIUS 2002, S. 690).

Mit Hilfe eines Screeplots lässt sich somit die Faktorenzahl bestimmen. Dieser sog. „Scree-Test“ ist ein rein optisches Analyseverfahren und betrachtet nur diejenigen Faktoren als bedeutsam, die *„vor dem Knick [„Elbow“] liegen“* (CATTEL 1966, zitiert in BORTZ 2005, S. 544), also vor dem Punkt, bei dem sich die Steigung der Eigenwerte stark verändert (BÜHL & ZÖFEL, 2005, S. 481), *„da von diesem Punkt an der Erklärungsgehalt weiterer Faktoren entsprechend gering ist“* (BROSIUS, 2002, S. 740).

⁹ http://www.phf.uni-rostock.de/institut/ipp/lehrmaterialien/diagnostikvorlesung/vd1_3_qualitaetsaspekte.htm#VD1_3_Qualitaetsaspekte
(Letzter Zugriff am 10.4.07)

¹⁰ SPSS: Statistical Package for the Social Sciences

Für den Wissenstest der Hauptuntersuchung spricht der Verlauf der Eigenwerte somit klar für eine einfaktoruelle Lösung, da sich die Eigenwerte bei weiteren Faktoren asymptotisch der Abszisse annähern (siehe Abbildung 8). Ergo: Der Test ist homogen konstruiert und fragt nur ein Themengebiet (Dimension, hier: Seifen) ab.

Gleiches gilt für die beiden Meinungstests (ScreepLOTS, siehe Anhang).

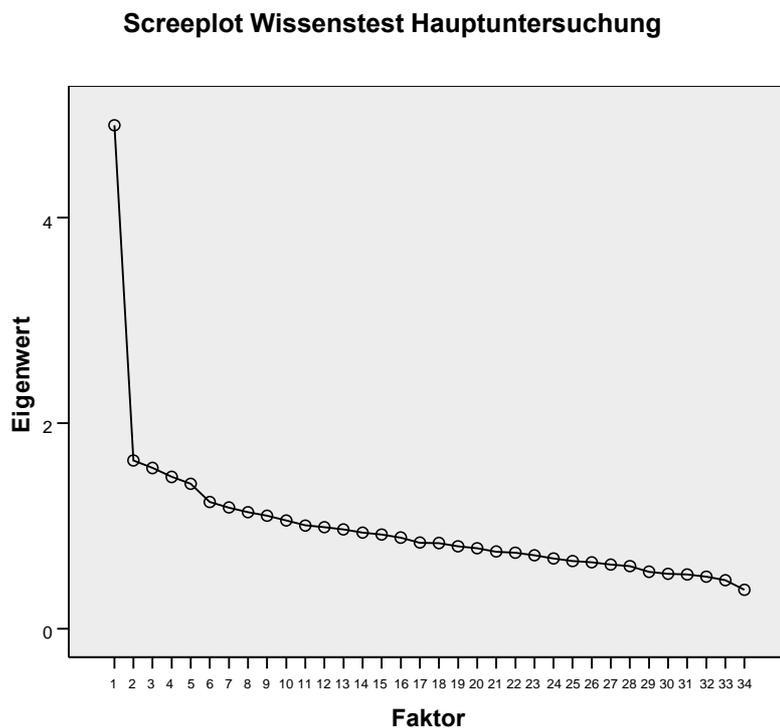


Abbildung 8: ScreepLOT für den Wissenstest der Hauptuntersuchung.

PERLETH (2007¹¹) fasst zusammen: „Nur ein Test mit hoher Validität kann daher wirklich sinnvoll interpretiert werden. Deshalb stellt die Optimierung der Validität bei der Konstruktion von Messverfahren eines der wichtigsten (und zugleich schwierigsten) Ziele dar“. Da dies nicht in jedem Fall möglich ist, kann stattdessen von einer guten Reliabilität auf die Validität geschlossen werden (vgl. BORTZ/DÖRING 2003).

¹¹http://www.phf.uni-rostock.de/institut/ipp/lehmaterialien/diagnostikvorlesung/vd1_3_qualitaetsaspekte.htm#VD1_3_Qualitaetsaspekte (Letzter Zugriff am 10.4.07)

5.4 Details zum Untersuchungsdesign

5.4.1 Allgemeines

Die beteiligten Lerngruppen wurden in zwei Hälften geteilt; sie erhielten nacheinander von der gewohnten Lehrkraft Unterricht zum Thema Seifen, wobei dieses inhaltsgleich bei der ersten Gruppe „herkömmlich“ im Frontalunterricht („FU-Gruppe“, also im weitesten Sinne darlegend-informativ bzw. fragend-entwickelnd) und bei der Experimentalgruppe nach der Gruppenpuzzle-Methode geschah („GP-Gruppe“, siehe Abbildung 9). Dabei durfte der Frontalunterricht durchaus Gruppenarbeit-Elemente enthalten, jedoch nicht überwiegend.

Wie wichtig das Unterrichten durch dieselbe Lehrkraft in beiden Gruppen war, zeigt eine Studie der AERA (2005, S. 71). Aus ihr geht hervor, dass der Lernfortschritt der Schüler zu einem bedeutsamen Teil von der Lehrperson abhängt. Eine Abweichung von diesem Design hätte somit gravierende Auswirkungen auf die Ergebnisse der Untersuchung.

Die Lehrer konzipierten den Frontalunterricht auf der Basis der Materialien des Gruppenpuzzles, so dass die behandelten Inhalte in beiden Gruppen identisch waren. In jeder Stunde war der Untersuchungsleiter anwesend und führte Stundenprotokolle. Dadurch konnte sichergestellt werden, dass alle Lerninhalte dargeboten wurden (siehe hierzu auch TEPNER/MELLE/ROEDER 2005a, S. 214 ff). Im abweichenden Fall wurde dieses in der Auswertung berücksichtigt (siehe Kapitel 5.4.3 und 8.6).

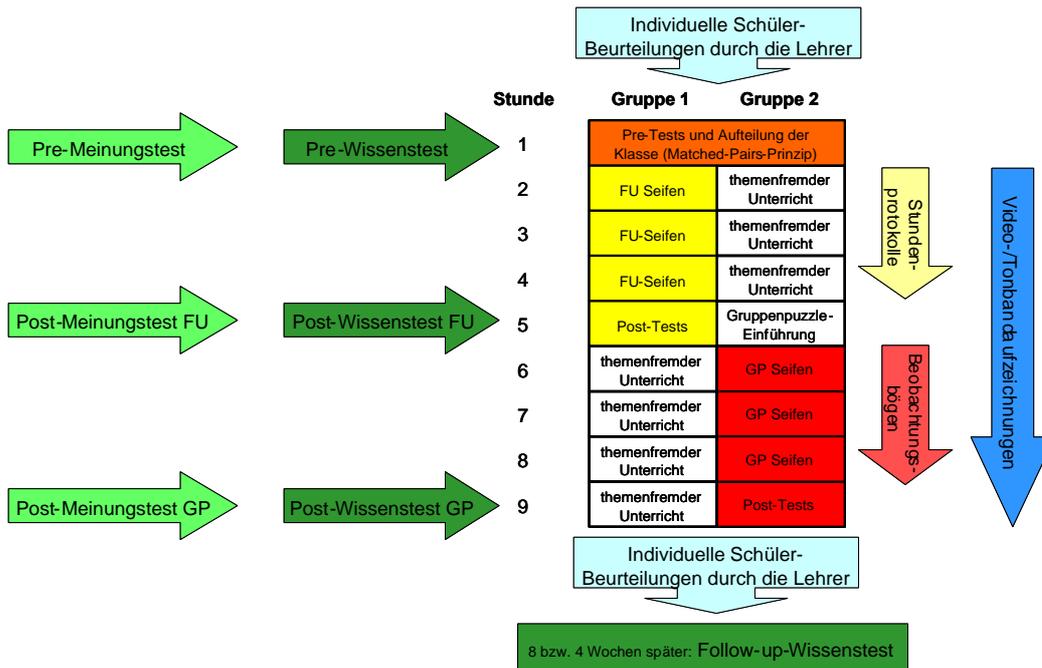


Abbildung 9: Untersuchungsdesign im Detail

Die Aufteilung der Klassen in die beiden Hälften erfolgte randomisiert nach dem Matched-Pairs-Prinzip. Dabei wurde auf der Grundlage von individuellen Schülerbeurteilungen (siehe Abschnitt 6.3), welche die Lehrer im Vorfeld angefertigt hatten, ein „Ranking“ der Leistungsfähigkeit erstellt. Per Zufallsgenerator wurde der beste Schüler einer Gruppe zugelost, der darauf folgende Schüler wurde automatisch der anderen Gruppe zugeordnet usw. So entstanden zwei gleich große und vor allem gleich leistungsstarke Gruppen. Diese für einen objektiven Vergleich wichtige Randomisierung kompensiert etwaige Unterschiede, wie sie z. B. zwischen Parallelklassen bestehen können, hervorgerufen durch die individuelle Lerngenese einer Klasse. So weisen Parallelklassen selten die gleichen kognitiven, sozialen oder motivationalen Voraussetzungen auf. Zugunsten einer besseren Vergleichbarkeit wurde das durch das Design veränderte soziale Gefüge des Klassenverbands und die geringere Schülerzahl – bedingt durch die halbe Klassenstärke – in Kauf genommen.

Die jeweils gerade nicht an der Untersuchung teilnehmende Gruppe wurde von studentischen Hilfskräften der Universität themenfremd unterrichtet. Unmittelbar vor bzw. nach Durchführung der Unterrichtsreihe wurden Wissens- und Meinungstests eingesetzt. Mit letzteren sollte die Attraktivität der Unterrichtsmethode erfasst werden.

In jeder Stunde wurde mit Ton- bzw. Videoaufzeichnungen der Redeanteil der Lehrkraft erfasst. Im Anschluss an die Untersuchung wurden nochmals individuelle Schülerbeurtei-

lungen speziell unter Berücksichtigung der letzten Unterrichtseinheit durch die Lehrer angefertigt.

Die Dauer eines Untersuchungsabschnitts pro Klasse betrug neun Schulstunden. Ca. acht bzw. später ca. vier Wochen¹² nach dem jeweiligen Post-Test wurde die Nachhaltigkeit des erworbenen Wissens mit Follow-Up-Tests überprüft. In zwei Fällen konnten zwei (kleine) Klassen zusammengelegt werden, so dass die Teil-Stichproben geringfügig größer ausfielen (siehe Tabelle 8).

Für die Schüler wichtig war die anonyme Durchführung der Meinungstests. Sie mussten lediglich ihren Vornamen sowie den Anfangsbuchstaben des Nachnamens angeben. Die jeweiligen Fachlehrer bekamen die Tests nicht zu sehen.

Anders verhielt es sich bei den Leistungstests: Hierbei wurden die Ergebnisse des Post-Tests von den Lehrern für die Zensurenggebung herangezogen. Die Schüler wurden durchweg darum gebeten, möglichst wenig für den Test zu üben, was weitgehend befolgt wurde (siehe Kapitel 8.1.6). Während des Tests existierten kaum Möglichkeiten zum Betrügen, was einerseits an der meist geringen Klassenstärke (Plätze frei gelassen), aber auch an der Anwesenheit von zwei Lehrern lag. Um den Faktor Stress auszuschalten, konnten die Schüler so lange schreiben wie sie wollten; sie waren durchweg nach 15-20 Minuten fertig.

5.4.2 Änderungen von Hauptuntersuchung I nach Hauptuntersuchung II

Die zweite Hauptuntersuchung im Jahre 2005 wurde als echte Replikation konzipiert, d. h. alle Faktoren blieben unverändert. Bezeichnet wird sie im Folgenden mit HU II_{intern}.

Für die parallel dazu durchgeführte HU II_{extern} ergeben sich folgende Änderungen:

- Erweiterung auf andere Schulformen (Haupt-, Realschule, Berufskolleg).
- 6 Klassen im Parallelklassendesign.
- In 10 Klassen wurde auf die permanente Anwesenheit des Untersuchungsleiters verzichtet (Einführung und Pre-Test wurden allerdings weiterhin durch den Untersuchungsleiter durchgeführt). Dadurch konnte die Untersu-

¹² Bedingt durch äußere Einflüsse wie z. B. Ferien, Praktika, Auslandsaufenthalte, variierte der Abstand, blieb jedoch für IG und KG immer gleich.

chung flexibler eingesetzt werden, da nun auch weiter entfernte Schulen partizipieren konnten.

- In einer Klasse musste die GP-Gruppe ohne die ständige Anwesenheit des Lehrers arbeiten.
- Erfassen der Übungsdauer der Schüler für den Test¹³.

Außerdem wurde für die Replikation das Signifikanzniveau auf 0.01 gesenkt, um noch konservativer zu testen (siehe nächster Abschnitt).

5.4.3 Ausrichtung

Das gesamte Design der Untersuchung ist konservativ ausgelegt worden, d. h. gegen die Wunschhypothese. Zum Beispiel durften die Lehrkräfte im FU am Ende der Stunde die Begriffe nochmals – z. B. in Form eines Blitzlichtes – wiederholen. Und nicht vom Lehrer dargebotene Lerninhalte wurden im Test als richtig gewertet (siehe Kapitel 8.6). Zusätzlich war die Gruppengröße der Kontrollgruppe zwar genauso groß wie die der Interventionsgruppe, im Vergleich zum alltäglichen Unterricht jedoch dürften sich Vorteile für die FU-Gruppe ergeben, da ein großer Klassenverband positiver auf diese geringe Gruppenstärke reagiert als das ohnehin in Kleingruppen praktizierte Gruppenpuzzle. Empirische Befunde hierzu jedoch sind widersprüchlich:

„Während die nationalen Studien keine eindeutigen Effekte der Klassengröße auf die Leistungsentwicklung von Schülern feststellen, belegen neuere internationale Studien [z. B. GRADDY & STEVENS 2003] recht eindeutig, dass eine kleine Klasse [einen positiven] Einfluss auf die Schülerleistungen bewirken kann; insbesondere dann, wenn Schüler der unteren Jahrgangsstufen dauerhaft in einer kleinen Klasse mit deutlich weniger als 20 Schülern verbleiben“, (IM BRAHM, 2006, Kap. 4). Eine Ursache für den z. T. ausbleibenden Lernerfolgsvorsprung in kleineren Klassen sieht IM BRAHM in der fehlenden Fähigkeit der Lehrer, auf diese neue Situation zu reagieren. Anstatt den Unterricht auf die kleinere Klasse abzustimmen, verwenden sie die gleichen Methoden und Standards wie in größeren Klassen.

¹³ Die Übungsdauer der Schüler wurde erst ab der 3. Klasse der 2. Hauptuntersuchung erfasst.

Vier Lehrer haben im FU unerwünschterweise zusätzliche Übungen als Hausaufgabe aufgegeben. Dennoch wurde der Test normal gewertet, obwohl im Prinzip mehr Lernzeit zur Verfügung stand. Allerdings wurde einer Lehrerin untersagt, das GP-Material der FU-Gruppe als Hausaufgabe zu geben.

Ein weiterer Vorteil für die Frontalunterricht-Gruppe mag darin bestanden haben, dass der Zeitraum zwischen Vor- und Nachtest kürzer war und sich evtl. einige Schüler besser an die Fragen erinnern konnten. Andererseits könnten sich Schüler der GP-Gruppe evtl. die Testfragen der FU-Gruppe haben mitteilen lassen, als letztere gemerkt hat, dass Vor- und Nachtest identisch sind. Dieses kann zwar nicht völlig ausgeschlossen werden, gerade bei Multiple-Choice-Tests kommt es jedoch auf exakte Formulierungen an, so dass der hierdurch entstandene Fehler als gering einzustufen ist.

Bei den wenigen Parallelklassen wurde stets die etwas schlechtere Klasse zur Interventionsgruppe.

6 Vorstellung der eingesetzten Erhebungsinstrumente

Alle Tests wurden selbst entwickelt, wobei inhaltliche sowie strukturelle Aspekte mit Prof. Dr. ROEDER, Lehrstuhl für Psychologie, Universität Dortmund, diskutiert wurden.

6.1 Wissenstest

Der Wissenstest setzte sich aus insgesamt 34 Fragen zum Thema Seifen zusammen, die im Multiple-Choice-Verfahren von den Schülern beantwortet werden mussten (siehe Abbildung 10). Abgefragt wurden sowohl deklaratives als auch zu einem kleinen Teil Transferwissen. Die Dauer des Tests betrug ca. 15 Minuten. Pre-Test und Post-Test waren identisch, wobei letzterer notenrelevant war. Die ersten drei Fragen waren etwas leichter, da sie als Shock-Absorber (auch Eisbrecherfragen genannt) fungierten.

Datum:	Was wisst ihr über Seifen?	Vorname:
Klasse:		Erster Buchstabe des Nachnamens:

Dieser Test dürfte dir in ähnlicher Form bekannt vorkommen. Nun zeige, was du über Seife gelernt hast! Kreuze die richtige Aussage bzw. Antwort an. Es ist jeweils nur eine Lösung richtig. Raten ist in jedem Fall besser als nichts anzukreuzen. Viel Erfolg!!!

Ein Seifenteilchen ist aufgebaut wie ein Streichholz, ...

- ...der Kopf ist unipolar.
- ...der Schwanz ist geladen.
- ...der Kopf ist positiv geladen.
- ...der Schwanz ist unipolar.
- ...der Kopf ist ungeladen.

Kernseife ist das...

- ...Kaliumsalz einer Fettsäure.
- ...Calciumsalz einer Fettsäure.
- ...Natriumsalz einer Fettsäure.
- ...Kaliumsalz einer Fettsäure.
- ...Magnesiumsalz einer Fettsäure.

Zwischen Molekülen treten immer dann Wasserstoffbrückenbindungen auf, wenn...

- ...in diesen Molekülen Sauerstoff mit einem stark elektronegativen Atom verbunden ist.
- ...in diesen Molekülen zwei gleich stark elektronegative Atome vorkommen.
- ...Wasserstoff ein Elektron komplett an den Sauerstoff abgibt.
- ...die Moleküle geladen sind.
- ...in diesen Molekülen Wasserstoff mit einem stark elektronegativen Atom verbunden ist.

Die Ursache für den grauen Belag in nicht geputzten Waschbecken ist eine Reaktion von...

- ...Seifengigonen mit Kalk zu Kalkseife.
- ...Seifengigonen mit Magnesiumgigonen zu Kalkseife.
- ...Seifengigonen mit Natriumgigonen zu Kalkseife.
- ...Seifengigonen mit Calciumgigonen zu Kalkseife.
- ...Seifengigonen mit Calciumgigonen zu Kalkseife.

1 Blatt bitte wenden

Abbildung 10: Wissenstest der Hauptuntersuchung, S. 1.

Testkonstruktion

Auf die allgemeine Konstruktion wurde bereits in Abschnitt 5.3 eingegangen. Für den Wissenstest mussten zusätzlich Decken- bzw. Bodeneffekte berücksichtigt werden. Diese auch als *ceiling effect* bzw. *floor effect* bezeichneten Effekte treten bei nicht an die Schülerklientel abgestimmten Tests auf (vgl. LIENERT und VON EYE 1994). Ersterer beschreibt die Rechtsverteilung der Messwerte, welche auftritt, wenn fast alle Schüler Testwerte nahe ihrer Obergrenze erhalten, der Test also zu leicht ist. Eine linksgipflige Verschiebung der Messwerte zeugt von einem zu schwierigen Test, was ebenfalls zu vermeiden ist. Nach LIENERT & RAATZ (1998) sollten die Schwierigkeitsindizes eine Lösungswahrscheinlichkeit von 20-80 % aufweisen, was durch die Voruntersuchung sichergestellt wurde (siehe Kapitel 5.3.2). Aus diesem Grund wurden in der Voruntersuchung Aufgaben, die eine Lösungswahrscheinlichkeit von über 80 % besaßen, herausgenommen bzw. umformuliert.

Follow-up-Tests

Follow-up-Tests dienen vor allem dazu, die Nachhaltigkeit des erworbenen Wissens zu messen. Gleichzeitig können Rückschlüsse auf die Vergessensrate gezogen werden. Da in der Literatur kaum verbindliche Angaben für die Dauer zwischen Intervention und Follow-up-Test zu finden sind, darf der Einsatz als Vorstudie gelten. Die Ergebnisse sollen weniger Hinweise auf die Effektivität der Methode als auf den optimalen Zeitpunkt des Testeinsatzes liefern. Zu diesem Zweck wurden verschiedene Abstände zwischen Post-Test und Follow-up-Test gewählt. Sie variierten von 24-77 Tagen (siehe Abbildung 27).

6.2 Meinungstests

Für die Untersuchung wurden insgesamt drei Meinungstests entwickelt: Ein Test zur allgemeinen Einstellung zum Fach Chemie (pre-post), zur Unterrichtseinheit (also die Einschätzung zu den letzten drei Stunden, nur post) und speziell zum GP Seifen (post). Mit Ausnahme des letzten wurden alle Tests sowohl in der Kontroll- als auch Interventionsgruppe vorgelegt, und zwar im Anschluss an den Wissenstest. Es folgt eine kurze Vorstellung der drei Tests.

6.2.1 Einstellung zum Fach Chemie

Der Test bestand aus insgesamt 24 Items, die ungefähr zur Hälfte positiv formuliert sind um die „Ja-Sage-Tendenz“ zu verringern. Zusätzlich wurden sie in eine Zufallsreihenfolge

gebracht um Positionseffekte (Reaktion auf die Abfolge der Items) zu verhindern. Positionseffekte sind „*Störvariablen, die von der Position einer experimentellen Bedingung in der Reihenfolge bestimmt*“ sind (HUBER 2005, S. 173). HUBER (2005) nennt folgende Beispiele: Ermüdung, Übungseffekt, Ablenkung, Durst, schwindende Angst etc. Analog zum Wissenstest enthielt der Meinungstest drei Shock-Absorber am Anfang. Das Layout der Bögen war an der Universität Dortmund im Fachbereich Psychologie standardisiert, um Vergleiche zwischen verschiedenen Studien anstellen zu können (siehe Abbildung 11).

	sehr zu- befried	zu- befried	mittel	unzu- befried	sehr unzu- befried
Ich fand den Unterricht der letzten 3 Stunden angenehm	<input type="checkbox"/>				
Ich habe in den letzten 3 Stunden das meiste verstanden	<input type="checkbox"/>				
Ich fand den Unterricht der letzten 3 Stunden abwechslungsreich	<input type="checkbox"/>				
Ich war im Unterricht der letzten 3 Stunden entspannt	<input type="checkbox"/>				
Ich fand den Unterricht der letzten 3 Stunden stressig	<input type="checkbox"/>				
Ich fühlte mich in den letzten 3 Stunden von meinen Mitschülern ernst genommen	<input type="checkbox"/>				
Mir hat der Chemieunterricht der letzten 3 Stunden gefallen	<input type="checkbox"/>				
So wie in den letzten 3 Stunden sollten auch andere Themen unterrichtet werden	<input type="checkbox"/>				
Ich bin froh, dass diese Unterrichtseinheit nun vorbei ist	<input type="checkbox"/>				
Die Unterrichtsmethode der letzten 3 Stunden fand ich gut	<input type="checkbox"/>				
In den letzten 3 Stunden war ich unaufmerksam	<input type="checkbox"/>				
Ich habe in den letzten 3 Stunden wenig Neues gelernt	<input type="checkbox"/>				

4

Blatt bitte wenden

Abbildung 11: Layout des Meinungstest am Beispiel „Attraktivität der Unterrichtseinheit“

Die Länge der Items betrug maximal 70 Zeichen (ohne Leerzeichen). In der Voruntersuchung sind die Formulierungen auf Verständlichkeit überprüft worden. Eine Itemanalyse

stellte sicher, dass jedes Item eine genügende Trennschärfe besitzt. Folgende Punkte wurden in Anlehnung an ROEDER (2006, S. 637 ff), BÜHNER (2004, S. 62 ff) und BORTZ & DÖRING (2003, S. 253 ff) bei der Itemformulierung beachtet:

- Lediglich eine Dimension pro Test
- pro Dimension mindestens 20 Items
- Items randomisieren (Positionseffekte)
- Items zu 50 % positiv, 50 % negativ formuliert (Ja-Sage-Tendenz)
- Item 1-3 als Shock-Absorber
- unabhängig voneinander formulierte Items
- präzise Formulierungen
- keine Statements, die sich auf mehr als eine Weise interpretieren lassen
- Sprache des Befragten verwenden
- Statements in Ich-Form
- Vermeidung extremer Formulierungen wie „immer“, „alle“, „keiner“, „niemals“, etc.
- keine direkten Verneinungen
- keine Statements, die entweder von fast jedem oder fast niemandem bejaht werden können
- kurze Statements (nur selten mehr als zwanzig Wörter)
- nur ein vollständiger Gedanke pro Statement

6.2.2 Attraktivität der Unterrichtseinheit

Die im vorherigen Kapitel beschriebenen Konstruktionsmerkmale gelten auch in diesem Test. Die Schüler sollten die jeweils letzten drei Stunden Chemieunterricht beurteilen. Anzumerken ist, dass die Items so formuliert waren, dass der Begriff „Gruppenpuzzle“ nicht vorkam. Stattdessen wurden durchweg von den „letzten 3 Stunden“ gesprochen, um den Test für IG und KG gleich zu halten.

6.2.3 Attraktivität des Gruppenpuzzles

Die Konstruktionsmerkmale gelten auch hier, lediglich die Anzahl der Items wurde in HU II von 19 auf 23 Items erhöht (siehe Anhang).

6.3 Individuelle Schülerbeurteilungen

Im Vorfeld der Untersuchung an der jeweiligen Schule wurden von den Lehrern individuelle Schülerbeurteilungen angefertigt. Dabei kamen Polaritätenprofile (oder „semantische Differenziale“, vgl. auch ROEDER 2006, S. 639) mit einer siebenstufigen bipolaren Skala zum Einsatz (siehe Anhang). Polaritätenprofile eignen sich besonders für die Messung von Stereotypen (vgl. BORTZ/DÖRING 2003). Die eigens für die Untersuchung konzipierten Fragebögen bestanden aus zwei Teilen mit den beiden Dimensionen: Mitarbeit im Unterricht (16 Items) und Verhalten (13 Items). Erstere umfasste leistungsbezogene Eigenschaften, die für die Notengebung relevant sind („meldet sich von alleine – muss aufgefordert werden“, „fachlich kompetent – inkompetent“, „selbstständig – unselbstständig“). Der zweite Teil (nicht-leistungsbezogene Eigenschaften) erfasste Schülerverhalten, das nicht direkt in die Notengebung einfließen sollte (z. B. freundlich – frech, geduldig – ungeduldig, selbstbewusst – ängstlich).

Ziel der Individuellen Schülerbeurteilungen war es, die Leistungsfähigkeit eines jeden Schülers einzuschätzen und mit den Ergebnissen ein Ranking zu erstellen, mit dem die Klasse in Interventions- bzw. Kontrollgruppe aufgeteilt werden konnte. Die nicht-leistungsbezogenen Eigenschaften (Verhalten) gingen in dieses Ranking nicht mit ein. Mit ihrer Hilfe sollten lediglich Aussagen über die Beurteilbarkeit von sog. „Soft Skills“ gemacht werden. Nach Durchführung der Intervention erstellten die Lehrer erneut Beurteilungen über jeden Schüler, um sie mit den Einschätzungen vor der Untersuchung zu vergleichen.

Weiterhin sollte der Frage nachgegangen werden, ob dieser zweifellos recht hohe Aufwand gerechtfertigt ist. Dabei sollte untersucht werden, in wieweit die individuellen Schülerbeurteilungen mit den Schulnoten der Schüler korrelieren und ob in Zukunft nicht der Einfachheit halber Schulnoten zur Aufteilung der Klasse herangezogen werden können.

6.4 Audioaufzeichnungen

Sowohl im Frontal- als auch im Gruppenpuzzle-Unterricht wurde die Stimme der Lehrkraft mit Hilfe eines kleinen MP3-Sticks, welcher um den Hals gehängt wurde, aufgezeichnet. Damit sollten einerseits die Redeanteile der Lehrer erfasst werden, andererseits auch die Stunde protokolliert werden, was von Vorteil war, wenn beispielsweise eine Videoauf-

zeichnung misslang. Eine Erfassung der Redeanteile erfolgte lediglich in Klassen, in denen der Untersuchungsleiter persönlich anwesend war.

6.5 Videoaufzeichnungen

Jede der vom Untersuchungsleiter persönlich protokollierten Stunden wurde gefilmt. Die Videoaufzeichnungen waren von Anfang an als Protokollierungsinstrument angelegt: Mit ihnen sollten evtl. auftretende „Ungereimtheiten“ im Nachhinein analysiert und aufgeklärt werden. Zusätzlich dienten sie als „Sicherung“, falls eine Audioaufnahme misslang. Da die Analyse und Kodierung von Unterrichtsvideos sehr aufwendig und für das Ziel dieser Arbeit nicht notwendig ist, sollte eine tiefer gehende Analyse nicht erfolgen. Für diese Untersuchung käme erschwerend hinzu, dass beim Gruppenpuzzle jede Gruppe für sich gefilmt werden und die Kamera in unmittelbarer Nähe der Schüler gestanden haben müsste, was als zusätzlicher „Fremdkörper“ das natürliche Lernverhalten der Schüler gestört hätte. Daher empfiehlt sich die Aufstellung bereits während der Einführungsstunde, um die Schüler daran zu gewöhnen.

Zur Aufzeichnung wurde eine mini-DV-Kamera verwendet, die auf einem Stativ im Rücken der Schüler aufgestellt war. So waren die Gesichter der Schüler in den meisten Fällen nicht zu erkennen und eine Einverständniserklärung der Schüler bzw. Eltern nicht erforderlich. Im Vorfeld wurde den Schülern allerdings der Sinn dieser Aufnahmen erläutert.

6.6 Stundenprotokolle

Zu jeder Frontalunterrichtsstunde wurde durch den Untersuchungsleiter ein Protokoll erstellt, auf dem festgehalten wurde, ob die Lehrkraft die entsprechenden Unterrichtsinhalte dargeboten hat, wo es Schwierigkeiten gab und wie der Lernerfolg eingeschätzt wurde (siehe Anhang). Fehlten Lerninhalte, wurden diese nachgearbeitet (vgl. auch 5.4.3).

6.7 Beobachtungsbögen

Während des Gruppenpuzzles wurden die Gruppen in jeder Phase von studentischen Hilfskräften beobachtet und das Schülerverhalten protokolliert. Schwerpunkt der Beobachtungen war das Abhaken der „wichtigen Begriffe“, so dass – analog zu den Stundenprotokollen im Frontalunterricht – sichergestellt werden konnte, dass alle Schüler die Lerninhalte

zumindest gehört haben. Ab dem zweiten Teil der HU II wurden die Beobachtungen durch schülereigene Protokollbögen ersetzt.

Die Beobachtungsbögen waren so konzipiert, dass nur direkt beobachtbare Verhaltensweisen erfasst wurden (siehe Anhang). Um die Versuchsbedingungen gleich zu halten, wurden die studentischen Hilfskräfte aufgefordert, den Schülern nicht zu helfen.

6.8 Klassen- und Lehrerfeedback

Mit Hilfe des Klassenfeedbacks sollten Meinungen/Kommentare der Schüler erfasst werden, die mit dem geschlossenen Fragebogen nicht eingeholt werden konnten. Die Befragung fand unmittelbar nach dem letzten Wissenstest in der jeweiligen Gruppe statt. Bei der FU-Gruppe war der Lehrer nicht anwesend.

In der ersten Hauptuntersuchung wurde zusätzlich ein Lehrerfragebogen im Pre-Post-Testdesign eingesetzt, um den Erfolg der Methode aus Sicht der Lehrer zu erfassen. Da dieser für die Lehrer eine weitere zeitliche Belastung darstellte und daher nur sporadisch ausgefüllt wurde, es sich außerdem zeigte, dass daraus keine wesentlichen neuen Erkenntnisse gewonnen werden konnten, wurde ab der HU II auf einen Einsatz verzichtet. Im Folgenden soll dieser nicht weiter diskutiert werden.

7 Auswertung

7.1 Wissenstest

7.1.1 Einfluss der Unterrichtsmethode auf den Lernzuwachs

Wie bereits in Abschnitt 5.4 beschrieben, wurde derselbe Wissenstest vor und unmittelbar nach Durchführung der Unterrichtseinheit eingesetzt (Pre-Post-Testdesign). Die Eingabe der Daten geschah per Hand in den Windows Editor. Dieses sehr einfache Programm hat die Vorteile, dass es einerseits auf jedem beliebigen Rechner verfügbar ist, andererseits der Cursor nach jeder Eingabe automatisch weiter springt. Somit entfällt das Drücken der Pfeil-Taste nach jeder Eingabe, wie es z. B. in SPSS erforderlich ist. Ein weiterer Vorteil bestand darin, dass die Daten sehr leicht für andere Programme, z. B. „Analiese“ (von B. ROEDER) und SPSS 14, zugänglich sind.

Eingabefehler konnten durch das o. g. Eingabeverfahren sowie die Beschränkung auf fünf Antwortmöglichkeiten auf ein Minimum reduziert werden. Zur Kontrolle wurde eine zufällige Auswahl an Rohdaten im Jahre 2006 von einer unabhängigen Person nochmals eingegeben. Die Anzahl der Eingaben aus 10 Wissens- und 10 Meinungstests betrug 587, was ca. 0.6 % der Gesamt-Eingaben an Wissens- und Meinungstest entspricht. Dabei lag die Fehlerquote der Ersteingabe bei 0.0 %.

Die Auswertung der Wissenstests im Pre-Post-Testdesign erfolgte in einer niveaubereinigten Veränderungsmessung über Residuen. „*Das Regressionsresiduum kennzeichnet die Abweichung eines empirischen y -Wertes vom vorhergesagten y -Wert.*“ (BORTZ 2005, S. 207). Begründen lässt sich die Verwendung von Residuen wie folgt: Eine Möglichkeit die Effektivität einer Unterrichtsmethode zu bestimmen ist mit Hilfe des Lernzuwachses. Dazu könnte man den Pre-Testwert vom Post-Testwert subtrahieren. Dabei wird allerdings nicht berücksichtigt, dass ein guter Schüler mit z. B. 80 % richtiger Antworten im Vortest kaum noch einen Lernzuwachs erreichen kann. Ein schlechter Schüler hingegen, der nur 10 % wusste, kann theoretisch noch viel dazulernen. Zusätzlich existiert ein so genannter „Regressionseffekt“ (vgl. ROEDER 1980, S. 154). Darunter versteht man in der Statistik die Tatsache, dass extreme Messwerte der Erstmessung in der Zweitmessung zu mittleren streben, ohne dabei reale Veränderungen abzubilden. Dieser „Rückschritt zur Mitte“ ist also das Phänomen, dass im Vortest besonders schlechte Schüler im Post-Test per se eine Ver-

besserung erzielen – auch ohne Intervention. Gründe hierfür können einerseits die Bekanntheit der Testfragen sein, aber auch die Tagesform. So gemessene Effekte stellen Artefakte dar.

Die Regressionstransformation berücksichtigt diese Probleme, indem sie die Differenz in Abhängigkeit der Korrelation Vortest/Nachtest misst, was man als Residuum bezeichnet (siehe Abbildung 12).

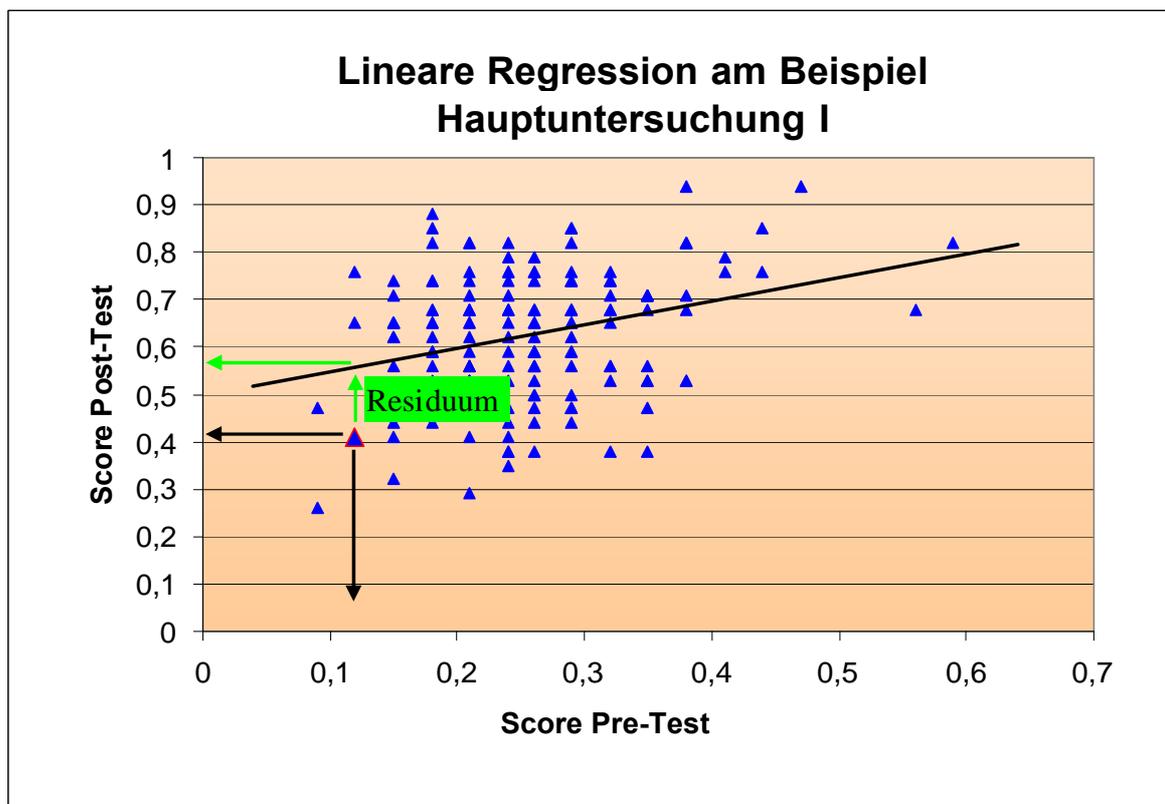


Abbildung 12: Veranschaulichung der Residuen. Aufgetragen sind die Ergebnisse des Nachtests eines oder mehrerer Schüler(s) gegen die entsprechenden Vortestergebnisse. Die Gerade symbolisiert die statistischen Erwartungswerte. Exemplarisch hervorgehoben ist ein Schüler mit 12 % im Vortest (schwarze Pfeile), für den im Nachtest 58 % (grüne Pfeile) zu erwarten gewesen wären. Unter Scores versteht man die erreichten Punktwerte (Lösungswahrscheinlichkeiten) der Schüler im Test. Nimmt man die Scores mal 100, so erhält man die richtig beantworteten Fragen in Prozent.

Die Residuen wurden abhängig von der betrachteten Stichprobe jeweils separat berechnet und sind z-standardisiert, um die Vergleichbarkeit mit anderen Untersuchungen zu gewährleisten (vgl. BORTZ 2005, S. 216 & S. 45).

Eine lineare Regression kann angenommen werden, da die grafische Auswertung darauf hindeutet; die statistische Linearitätsprüfung bestätigt den Befund (siehe Anhang). Gegen-

über anderen Modellen bietet die lineare Regression den Vorteil, dass sich mit SPSS standardisierte Residuen berechnen lassen, die einen Vergleich mit anderen Untersuchungen zulassen. Weiterhin ergaben Vergleichsrechnungen mit insgesamt elf Modellen (vgl. TEPNER, O., i. V.), dass sich zwar gewisse Unterschiede in der Passung der Modelle ergeben, diese sich jedoch bei der Durchführung der U-Tests irrelevant werden, so dass das verwendete Modell keine Auswirkungen auf die Ergebnisse der Untersuchung hat. Dies ist nicht verwunderlich, da für IG (Interventionsgruppe) und KG (Kontrollgruppe) dasselbe Modell Verwendung findet.

Als Reliabilitätsmaß wurde Cronbachs Alpha auf Basis der dichotomisierten Rohdaten (richtig bzw. falsch) errechnet. Letztere wurden allerdings zugunsten einer sehr konservativen Auswertung modifiziert. Dabei erhielten FU-Schüler, deren Lehrer den Lernstoff – z. B. aus zeitlichen Gründen – nicht durchgenommen hatten, für die entsprechende Frage im Test einen Punkt. Insgesamt waren drei Lerngruppen betroffen, davon eine in der Voruntersuchung. Für die gesamte Untersuchung wurden in beiden Gruppen zudem alle fehlenden Werte (missing data) als Fehler gewertet.

Neben Residuen werden bei den Ergebnissen zusätzlich Mittelwerte angegeben, um einen Vergleich zu den momentan noch populäreren herkömmlichen Auswertungsverfahren über Mittelwertvergleiche zu ermöglichen. Zudem erleichtern sie die Vorstellbarkeit.

t-Test versus U-Test

Sollen wie im vorliegenden Fall zwei voneinander unabhängige Stichproben aus zwei Grundgesamtheiten (alle Chemieschüler der Klasse 10 an NRW-Gymnasien, die frontal unterrichtet werden, versus alle Chemieschüler der Klasse 10 an NRW-Gymnasien, die nach der Gruppenpuzzlemethode unterrichtet werden) verglichen werden, so überprüft der t-Test für unabhängige Stichproben die Nullhypothese, dass die beiden Stichproben aus Populationen (Grundgesamtheiten) stammen, deren Mittelwerte identisch sind (vgl. BORTZ 2005). Die Nullhypothese lautet:

$$H_0: \mu_1 - \mu_2 = 0$$

Voraussetzung für den Einsatz dieses parametrischen Testverfahrens ist die Annahme, dass die Grundgesamtheiten, aus denen die Stichproben entnommen wurden,

- a) Varianzhomogenität aufweisen,

- b) normalverteilt und
- c) intervallskaliert sind (vgl. z. B. BROSIUS 2002, S. 455 f).

„Für eine Intervallskala gilt, dass gleich große Merkmalsunterschiede durch äquidistante Zahlen abgebildet werden, d. h. identische Meßwertunterschiede zwischen Objektpaaren entsprechen identischen Merkmalsunterschieden“ (BORTZ/DÖRING 2003, S. 72). Da die erhobenen Daten in Scores umgerechnet wurden, darf diese Bedingung als erfüllt gelten und stellt somit kein Hindernis für den Einsatz des t-Tests dar. Auch das Problem der Varianzhomogenität ist weniger relevant, da der t-Test auch bei heterogenen Varianzen Anwendung finden kann (vgl. BORTZ 2005). SPSS gibt für diesen Fall eine gesonderte Statistik aus (vgl. BROSIUS 2002).

Bedeutsamer ist die Voraussetzung der Normalverteilung der beiden Grundgesamtheiten auf Basis der Residuen, da diese auf Unterschiede überprüft werden sollen. Zwar geht aus Monte-Carlo-Studien hervor, dass der t-Test auf Verletzungen seiner Voraussetzungen relativ robust¹⁴ reagiert (vgl. BONEAU 1971, GLASS et al. 1972, SAWIŁOWSKY u. BLAIR 1992, alle zit. in BORTZ 2005), jedoch steigt die Anzahl der Fehlentscheidungen, wenn weder Varianzgleichheit noch Normalverteilung vorliegen.

Ist die Varianz der kleineren Stichprobe größer als in der größeren Stichprobe, so entscheidet der t-Test eher zugunsten der H_1 , also progressiv. Genau andersherum verhält es sich, wenn die Varianz der kleineren Stichprobe kleiner ist als in der größeren Stichprobe; der Test entscheidet eher konservativ (vgl. BORTZ 2005). Da insbesondere progressive Fehlentscheidungen zu vermeiden sind, wird in der Auswertung der U-Test nach Mann und Whitney als nicht parametrisches Testverfahren eingesetzt, wohl aber mit dem t-Test verglichen. Die Formel eines parameterfreien Tests baut nicht auf Rohdaten, sondern auf Rangplätzen auf, die diesen Werten zugeordnet sind (vgl. ZÖFEL 2002).

Für den U-Test spricht außerdem, dass er selbst bei vorliegender Normalverteilung eine Effizienz von 95 % besitzt, bei großen Fallzahlen sogar 95.5 % (vgl. ZÖFEL 2002). Unter Effizienz versteht ZÖFEL (2002) das für den Signifikanznachweis erforderliche Verhältnis der Stichprobenumfänge beim parametrischen und dem entsprechenden nicht parametrischen Test.

¹⁴ Robust bedeutet, dass er trotz Verletzung seiner Voraussetzungen richtig entscheidet.

Sofern die Daten mindestens bei einer Stichprobe nicht normalverteilt sind, empfiehlt ZÖFEL (2002, S. 91) daher stets die Verwendung von parameterfreien Tests, „um ein schwer interpretierbares Durcheinander verschiedener Tests zu vermeiden“. Zwar wurde bei der Testkonstruktion darauf geachtet, den Test mittelschwer zu gestalten, um normalverteilte Daten zu erhalten (Stichwort: 20 % – 80 %-Regel, siehe Abschnitt 6.1). Die Ausweitung der Untersuchung auf andere Schulformen jedoch birgt das Risiko einer schiefgipfligen Verteilung, was zwangsläufig zum Einsatz des U-Tests führt. Abgesehen davon waren die Daten der HU I (GP-Gruppe, siehe Tabelle 7) nicht normalverteilt, wie der Kolmogorov-Smirnov-Test ($\alpha = 0.2$) ergab. Das hierbei angehobene Signifikanzniveau von 20 % (vgl. BORTZ/LIENERT 2003, S. 227 und BORTZ/DÖRING 2003, S. 603) macht Sinn, da beim Test auf Normalverteilung H_0 die Wunschhypothese ist; deren Beibehaltung gilt es zu erschweren.¹⁵

Signifikanzniveau

Für die Voruntersuchung (VU) und die erste Hauptuntersuchung (HU I) gelten Signifikanzniveaus von 5 %, ab der zweiten Hauptuntersuchung in 2005 (HU II) gilt das schärfere 1 %-Niveau. In der Auswertung werden die üblichen Symbole verwendet (siehe Tabelle 4). Aufgrund der Ergebnisse der Voruntersuchung wurde ab der HU I beim Wissenstest einseitig getestet.

Tabelle 4: Für die statistische Auswertung verwendete Signifikanzniveaus

Signifikanzniveau	Bezeichnung	Symbol
$p < 0.05$	signifikant	*
$p < 0.01$	hoch signifikant	**
$p < 0.001$	höchst signifikant	***

Zusätzlich zu den Irrtumswahrscheinlichkeiten wird die faktisch erzielte Effektstärke (-größe) mit angegeben. Nach FALLER (2004) macht dies Sinn, da die Signifikanz eines Untersuchungsergebnisses in hohem Maße von der Größe der untersuchten Stichprobe

¹⁵ Beim Test auf Normalverteilung handelt es sich um einen Vorrasssetzungstest, so dass die H_0 zur Wunschhypothese (Annahme: Die Daten der beiden Stichproben sind normalverteilt) wird. Da für diesen Fall keine spezifische Alternativhypothese (H_1) vorliegt, muss der β -Fehler (Wahrscheinlichkeit, die H_0 zu akzeptieren, obwohl sie eigentlich falsch ist) indirekt klein gehalten werden, was durch Vergrößern des α -Fehlers (richtige H_0 wird zugunsten der H_1 abgelehnt) geschieht (BORTZ 2005, S. 122).

abhängt. So können selbst minimale, unbedeutende Unterschiede aufgrund einer großen Stichprobe signifikant werden, während bei kleinen Stichproben sogar inhaltlich bedeutende Effekte die Signifikanz verfehlen können. Er empfiehlt daher, bei der Darstellung der Ergebnisse auch immer die Effektstärke, welche vom Einfluss der Stichprobengröße bereinigt ist, mit anzugeben.

Die Formel für die Effektstärke d des t-Tests für unabhängige Stichproben lautet:

$$d = \frac{M_{IG} - M_{KG}}{S_{gesamt}},$$

nach BÜHNER (2004, S. 112), BORTZ (2005, S. 143), BORTZ/DÖRING (2003, S. 604). Hierbei steht M für die Residuen-Mittelwerte der Interventions- bzw. Kontrollgruppe (ex-post Bestimmung von d). Mit der angegebenen Formel lässt sich streng genommen lediglich die Effektstärke im Anschluss an den t-Test bestimmen; dementsprechend sind Effektstärken nur zu Vergleichszwecken mit angegeben – analog zum t-Test. Für den U-Test wird üblicherweise keine Effektstärke angegeben, da dieser bei Nicht-Normalverteilung der Stichproben verwendet wird und dementsprechend keine Standardabweichungen berechnet werden können.

Die Gesamt-Standardabweichung S setzt sich aus den gepoolten Standardabweichungen beider Gruppen zusammen und wurde mit SPSS errechnet. BORTZ/DÖRING (2003, S. 604) klassifizieren die Effektstärke folgendermaßen:

$d = 0.20 \rightarrow$ schwacher Effekt

$d = 0.50 \rightarrow$ mittlerer Effekt

$d = 0.80 \rightarrow$ starker Effekt

Durch diese Standardisierung der Ergebnisse mit Hilfe der Effektstärken wird also eine relative Unabhängigkeit vom Messinstrument erreicht, so dass auch sehr unterschiedliche Untersuchungen miteinander verglichen werden können.

Eine Angabe der Konfidenzintervalle der Mittelwerte, wie sie FALLER (2004) empfiehlt, ist nicht notwendig, da die Unterschiede zwischen den Stichproben per Mann-Whitney-U-Test auf Signifikanz getestet werden.

Für die Zusammenhangsanalysen – etwa Übungsdauer mit Residuen – wurde der Korrelationskoeffizient nach Spearman (ρ , r_s oder ρ) bestimmt, da die zu Grunde liegenden Daten nicht normalverteilt sind. Ab einem $p(z) < 0.05$ gilt die Korrelation als signifikant.

Zum Vergleich wurden ebenfalls „Scores“ angegeben, was den Lösungswahrscheinlichkeiten der Schüler entspricht. Nimmt man die Scores mal 100, so erhält man die richtig beantworteten Fragen in Prozent.

Die Ergebnisse der U-Tests wurden auf Basis der Rohdaten von B. ROEDER mit dem Programm „Analiese“ verifiziert. Dadurch konnten Fehler, etwa bei der Scoreberechnung oder Umkodierung, ausgeschlossen werden.

7.1.2 Abhängigkeit der Lernleistung von der Art der Wissensvermittlung

Drei Möglichkeiten der Wissensvermittlung kommen bei dieser Untersuchung in Frage: Der Frontalunterricht wurde so ausgelegt, dass lediglich der Lehrer die notwendigen Informationen liefert. Beim Gruppenpuzzle kann die Darbietung der Lerninhalte durch Mitschüler und durch Eigenleistung („selbst“) erfolgen, sofern man Experte für das entsprechende Thema ist. Zum Zweck der Unterscheidung sind während der Untersuchung die Gruppenpuzzle-Expertenthemen erfasst worden. Bei der Auswertung wurde so vorgegangen, dass zuerst für alle vier Themengebiete separat Scores berechnet wurden. Die Themen waren:

Thema 1 = Aufbau und Löseverhalten von Seife/Waschwirkung

Thema 2 = Herabsetzung der Oberflächenspannung durch Seife

Thema 3 = Was ist Seife und die Nachteile von Seife

(Thema 4 = Transfer)

Für jedes Themengebiet wurde festgelegt, ob das Thema selbst beigebracht wurde (mit 1 kodiert) bzw. von Mitschülern (mit 2 kodiert) vermittelt wurde. FU-Schülern wurde grundsätzlich eine „3“ für „vom Lehrer vermittelt“ eingetragen. Da es sich bei Thema 4 um Transferwissen handelte, d. h. Wissen, das nicht explizit im Expertenmaterial des Gruppenpuzzles vorhanden war, wäre eine analoge Kodierung zwar möglich gewesen, jedoch hätte dieses zu einer zu kleinen Stichprobe geführt, so dass Thema 4 nicht in diese Auswertung eingeht.

Nach Berechnung der Residuen für jede Untersuchung (VU, HU I, HU II) wurde mit ihnen ein Trendtest durchgeführt. Dieser basiert auf der Annahme, dass die Vermittlungsart zu Effekten führt, was sich in auf- bzw. absteigenden Mittelwerten der Residuen äußert. Es wurde vermutet, dass selbst angeeignetes Wissen zu einem höheren Lernzuwachs führt und dass die Vermittlung durch Schüler in Kleingruppen effektiver ist als die vom Lehrer – auch wenn sie evtl. nicht so gut erklären. Formulieren lässt sich das als folgende Hypothese:

$$H_1: \text{MW Residuen}_{\text{Lehrer}} < \text{MW Residuen}_{\text{Mitschüler}} < \text{MW Residuen}_{\text{Selbst}}$$

Allgemeiner ausgedrückt:

$$H_1: \mu_1 < \mu_2 < \mu_3$$

Eine Hypothese wie diese bezeichnet man als „monotone Trendhypothese“ (vgl. BORTZ 2005). Zur Überprüfung wurde der Trendtest nach Jonckheere-Terpstra eingesetzt. Dieser hat den Vorteil, dass er sich auch für ungleich große Stichproben eignet. Bei diesem werden alle Paare von Stichproben mit einer U-Test-ähnlichen Teststatistik verglichen (vgl. BORTZ/LIENERT 2003, S. 163) Liegt die Irrtumswahrscheinlichkeit unter $\alpha = 0.05$, so ist die H_0 abzulehnen. Dies bedeutet, die Stichproben folgen einem Trend.

Da nicht unbedingt große Unterschiede bei nicht selbst angeeigneten Themengebieten zu erwarten sind, wurde zusätzlich grober kategorisiert in „selbst beigebracht“ und „fremd vermittelt“, worunter im zweiten Fall Lehrer und Mitschüler zusammengefasst (gepoolt) worden sind. „Selbst“ und „Fremd“ wurden per U-Test auf etwaige Differenzen getestet. Die verschiedenen Items wurden den Themen folgendermaßen zugeteilt (siehe Tabelle 5).

Tabelle 5: Aufteilung der Wissenstest-Items für die Hauptuntersuchungen.

Thema	Nr. der Items	N
1	1, 5, 6, 11, 13, 14, 20, 21, 23, 31, 32	11
2	8, 10, 12, 19, 24, 25, 29, 34	8
3	2, 4, 9, 15, 16, 17, 18, 26, 28, 33	10
4	3, 7, 22, 27, 30	5

Die Richtigkeit der Zuteilung sowie die Kodierungen wurden zweifach überprüft.

7.1.3 Wechselwirkungen zwischen Unterrichtsmethode und Chemiezensur

Ziel dieser Teiluntersuchung war es, zu überprüfen, ob das Gruppenpuzzle für gute und schlechte Schüler gleichermaßen wirkt. Die Klassifizierung gelang mit Hilfe der letzten Zeugnisnote im Fach Chemie, welche im Vorfeld erfasst wurde.

Zum Verfahren:

Da zwei unabhängige Variablen (Chemienote und Methode) sich gegenseitig beeinflussen können, wurde statt eines einfachen U-Tests eine zweifaktorielle Varianzanalyse (ANOVA¹⁶) gerechnet (BORTZ 2005, S. 290). Dieser so genannte Globaltest vereint mehrere Signifikanztests und berücksichtigt Wechselwirkungen höherer Ordnung. Beim einfachen U-Test hätten für jede Gruppe getrennt Residuen berechnet und mit diesen Stichproben jeweils ein U-Test durchgeführt werden müssen. Bei zwei oder mehr Signifikanzanalysen unterschiedlicher Stichproben ist die Wahrscheinlichkeit, Unterschiede zu finden, größer als bei nur einer Analyse (vgl. STATSOFT 2007¹⁷ und ROEDER 1974a, S. 287, BORTZ 2005, S. 250, BROSIUS 2002, S. 484). Die Gefahr falscher Schlussfolgerungen würde somit bestehen.

Verwendet wurde die univariate Varianzanalyse nach dem Allgemeinen Linearen Modell (ALM). Dieses Verfahren überprüft, wie „*eine abhängige Variable* [Residuen pre-post] *von zwei unabhängigen Variablen* (=Faktoren) [Gruppe, Chemienote] *beeinflusst wird*“ (vgl. BORTZ 2005, S. 290). Vorteil dieses Verfahrens ist, dass ebenfalls Wechselwirkungen zwischen den unabhängigen Variablen detektiert werden können. Zum Beispiel kann nicht nur überprüft werden, ob der Lernzuwachs mit den Noten korreliert, sondern auch, ob das Gruppenpuzzle für die unterschiedlichen Leistungstypen innerhalb der Stichprobe gleichermaßen wirkt.

Da Varianzanalysen eine Normalverteilung der Residuen (für FU und GP getrennt berechnet) voraussetzen (BORTZ 2005, S. 284), wird im Vorfeld der Kolmogorov-Smirnov-Test auf Normalverteilung durchgeführt. Ab einem $p(z) \geq 0.2$ darf die H_0 (Wunschhypothese) angenommen werden, d. h. die Daten sind bis auf die GP-Gruppe der HU I normalverteilt (siehe Anhang). Letztere besitzt ein $p(z)$ von 0.066.

¹⁶ ANOVA = ANalysis Of VAriance

¹⁷ [http://www.statsoft.com/textbook/esc.html#What%20is%20"statistical%20significance"%20](http://www.statsoft.com/textbook/esc.html#What%20is%20%22statistical%20significance%20) (Letzter Zugriff am: 6.4.2007)

Eine weitere Voraussetzung für Varianzanalysen ist die Unabhängigkeit der Residuen (BORTZ 2005, S. 284 und 286). Hiervon ist in dieser Arbeit auszugehen, da die verschiedenen Treatments (GP, FU) zufällig zugeordnet wurden und verschiedene Stichproben untersucht worden sind (vgl. BORTZ 2005 ebd.).

Generell gilt jedoch, „*dass die Voraussetzungen der Varianzanalyse mit wachsendem Umfang der untersuchten Stichproben an Bedeutung verlieren*“ (BORTZ 2005, S. 286). Zudem sind „*Abweichungen von der Normalität zu vernachlässigen, wenn die Populationsverteilungen schief sind*“ (BORTZ 2005, S. 287), wie im vorliegenden Fall (rechtsgipflig, siehe Anhang, Abschnitt 12.3.2).

Weiterhin setzen Varianzanalysen eine Varianzhomogenität der Stichproben voraus, welche nicht immer vorlag. Im Fall von nicht gegebener Varianzhomogenität empfehlen BÜHL & ZÖFEL (2005, S. 403), „*die Signifikanzschranke nicht bei $p = 0.05$, sondern bei $p = 0.01$ anzusetzen*“. Zusätzlich beeinflussen heterogene Varianzen „*den F-Test nur unerheblich, wenn die [wie im vorliegenden Fall] Stichproben gleich groß sind*“ (BORTZ 2005, S. 287).

Folgende Signifikanzniveaus sollen bei der zweifaktoriellen Varianzanalyse gelten:

- Gruppen sind unterschiedlich für $p < 0.01$, H_0 ist abzulehnen (Haupteffekte)
- Lernzuwachs ist unabhängig (z. B. von der Chemienote) für $p \geq 0.2$, H_0 (Wunschhypothese) darf angenommen werden

Gleichzeitig wird die Effektstärke für Varianzanalysen, η^2 , angegeben, die nach BORTZ (2005, S. 259) klassifiziert wird:

$$\eta^2 = 0.01 \rightarrow \text{schwacher Effekt}$$

$$\eta^2 = 0.06 \rightarrow \text{mittlerer Effekt}$$

$$\eta^2 = 0.14 \rightarrow \text{starker Effekt}$$

SPSS gibt jeweils das partielle η^2 aus. Die grafische Auswertung erfolgte mit Microsoft Excel.

Anmerkung: Dem Autor ist bewusst, dass dieses Verfahren bei der vorliegenden Datenstruktur nach streng statistischen Maßgaben nicht hätte angewendet werden dürfen. In Ermangelung eines adäquaten nichtparametrischen Verfahrens zur Berechnung von Wechsel-

selwirkungen und um eine gewisse Vergleichbarkeit mit anderen Studien herzustellen, soll es hier dennoch zum Einsatz kommen.

7.1.4 Abhängigkeit des Lernerfolgs vom Geschlecht

Analog zu Punkt 7.1.3 wurde eine zweifaktorielle Varianzanalyse mit grafischer Unterstützung durchgeführt. Dadurch sollten sowohl Geschlechtsunterschiede per se als auch in Abhängigkeit von der Unterrichtsmethode aufgedeckt werden, so dass geklärt werden kann, ob die Intervention auf ein Geschlecht stärker wirkt (zum Verfahren: siehe Abschnitt 7.1.3.).

7.1.5 Abhängigkeit des Lernerfolgs von der Schulform

Das Verfahren entspricht dem zuvor aufgeführten (Varianzanalyse), um evtl. Einflüsse der Schulform auf den Lernzuwachs nachzuweisen. Auch hier bietet die mehrfaktorielle ANOVA zusätzlich den Vorteil, Auswirkungen der Methode auf unterschiedliche Schulformen zu untersuchen. Zu diesem Zweck wurde eine dreifaktorielle Varianzanalyse gerechnet, welche mögliche Wechselwirkungen zwischen Methode, Geschlecht und Schulform berücksichtigt. In die Auswertung ein ging lediglich die externe Hauptuntersuchung, da lediglich hier andere Schulformen beteiligt waren.

7.1.6 Einfluss der Übungsdauer auf den Lernerfolg

Hierzu wurde ebenfalls eine zweifaktorielle Varianzanalyse durchgeführt, die um eine Korrelationsanalyse (Spearman's rho) ergänzt wurde. Einerseits sollten dadurch evtl. Unterschiede in den Übungszeiten der Gruppen erfasst werden, andererseits auch die Frage beantwortet werden, ob die Methode auf verschiedene „Übetypen“ (Schüler die mehr bzw. weniger für den Test geübt haben) unterschiedlich stark wirkt. Eine dreifaktorielle Varianzanalyse (Methode, Geschlecht, Übedauer) konnte aufgrund des zu kleinen Datenpools nicht berechnet werden¹⁸.

7.1.7 Nachhaltigkeit des erworbenen Wissens

Mit Hilfe von Follow-up-Tests lassen sich zwei Fragen beantworten:

¹⁸ Die Übungsdauer der Schüler wurde erst ab der 2. Hauptuntersuchung erfasst.

1. Welche Gruppe hat nach einem bestimmten zeitlichen Abstand mehr Wissen?
2. Welche Gruppe vergisst das Gelernte schneller?

Zur Beantwortung der ersten Frage wurden Residuen aus Pre-Test (unabhängige Variable) und Follow-up-Test (abhängige Variable) erzeugt und mit dem U-Test nach Mann und Whitney auf evtl. Unterschiede überprüft. Um feststellen zu können, ob die Schüler nach 6-8 Wochen bereits signifikant viel vergessen haben, wurde der Wilcoxon-Test für 2 verbundene Stichproben durchgeführt. In diesen nichtparametrischen Test gingen Residuen aus Pre- und Post-Test und Residuen aus Pre- und Follow-up-Test ein. Eventuelle Wechselwirkungen zwischen Methode und Abstand wurden mittels zweifaktorieller Varianzanalyse untersucht (siehe hierzu Kapitel 8.1.3). Dabei wurden die Abstände Post-Test – Follow-up-Test mit SPSS mediandichotomisiert. Darunter versteht man die Aufteilung des Datenpools in zwei Gruppen:

1. kurzer Abstand \leq Median,
2. langer Abstand $>$ Median

(Zur Verwendung des Medians siehe Abschnitt 7.2.1.)

7.2 Meinungstest

7.2.1 Einstellung zum Fach Chemie

Die Rohdaten wurden zunächst per Hand in den Windows Editor eingegeben und dann in SPSS kopiert. Mit Hilfe einer für jeden der drei Untertests geschriebenen Syntax wurden negativ formulierte Items umkodiert. Die Daten wurden auf Basis der Mediane dichotomisiert. Dieses Verfahren hat den Vorteil, dass Wissens- und Meinungstest erstens gleich behandelt werden (vgl. ROEDER¹⁹). Zweitens kann einer Schiefverteilung der Antworten bei bestimmten Items besser begegnet werden, da jede Antwort gemessen an ihrem Median als Zustimmung bzw. Ablehnung (Score 0/1) gewertet wird (siehe hierzu auch [ad 1](#), unten). Die Mediandichotomisierung erfolgte für jedes Item separat – ebenfalls mit SPSS. Hierbei wurden Werte, die kleiner als der exakte Median (zwei Nachkommastellen, zum Verfahren siehe BORTZ 2005, S. 37) waren, mit 1, und Werte, die über dem exakten Median lagen,

¹⁹ seminarserver.fb14.uni-dortmund.de/roeder/Methodologie/MessAlgo.pdf (Letzter Zugriff am: 9.4.2007)

mit 0 kodiert. Pre- und Post-Test wurden gepoolt. Alle Items eines jeden Schülers wurden zu einem Pre-Test-Score bzw. Post-Test-Score zusammengerechnet (Mittelwertbildung).

Die Vorteile dieses zweifellos aufwendigeren Verfahrens gegenüber der üblichen (aber nicht zulässigen!, s. u.) Likert-Skalen-Mittelwertbildung sind:

1. Vermeidung der Gewichtung der Antworten,
2. Bessere Berücksichtigung extremer Verteilungen,
3. Möglichkeit der Medianberechnung auch für nicht kardinalskalierte Daten.

ad 1

Die Standardabweichung eines Items, also die Streuung der Antworten, beeinflusst die Gewichtung eines Items zum Gesamtscore. Beispiel: Wurde bei einem Item von fast allen Probanden „mittel“ angekreuzt, so erhalten alle Probanden gleichermaßen den gleichen Zahlenwert zu ihrem Score – es findet demnach keine Differenzierung mehr statt (was sich nebenbei in einer äußerst niedrigen Trennschärfe äußert!). Das Item liefert folglich keinen Beitrag mehr zur Schätzung der Dimension und müsste demnach aussortiert werden was – im Nachhinein erst bemerkt – zu sehr kleinen Itemzahlen führen kann (ein Item liefert allgemein höhere Beiträge zu einem Score, wenn es die volle Skala eines Tests ausnutzt).

Derartige Extremfälle sind allerdings selten. Häufiger wird man Items vorfinden, die zu 80 % von allen Probanden gleich angekreuzt worden sind, und eine sehr geringe Trennschärfe aufweisen und demnach aussortiert werden müssten. Durch die Mediandichotomisierung erreicht das betreffende Item jedoch wieder eine respektable Trennschärfe.

ad 2

Ist ein Item derart extrem formuliert, dass es von fast allen Testpersonen als unzutreffend bewertet wird, so resultiert daraus eine sehr schiefgipflige Verteilung der Antworten. Die Mediandichotomisierung berücksichtigt dieses, da sie unabhängig von der Skalenbeschriftung nur in „*pro und contra dimensionem*“ (vgl. ROEDER²⁰) einteilt.

Hinzu kommt der Fall der Zweigipfligkeit: Beurteilt z. B. eine Hälfte der Probanden ein Item entweder als sehr gut (1) und die andere Hälfte lehnt es völlig ab (5), so entsteht praktisch ein Berg mit zwei Spitzen bei 1 und 5 auf der Likert-Skala. Der Mittelwert von drei

²⁰ seminarserver.fb14.uni-dortmund.de/roeder/Methodologie/MessAlgo.pdf (Letzter Zugriff am: 2.12.2006)

würde suggerieren, dass das Item „mittelmäßig“ gefalle. Tatsächlich jedoch repräsentiert der Mittelwert keine der beiden Antworten.

Daraus resultiert ein weiterer Vorteil der Medianberechnung: seine universelle Einsetzbarkeit. Zwar kann eine Normalverteilung der Messwerte durch eine saubere Testkonstruktion stark begünstigt werden, möchte man aber wie bei dieser Untersuchung den gleichen Test auf eine unbekannte Stichprobe (hier: andere Schulformen) anwenden, so lässt diese Form der Auswertung auch mögliche rechts- oder linksgipflige Verteilungen zu. Dieses gilt für Meinungstests in noch stärkerem Maße, da Meinungen noch schwerer vorherzusagen sind als Wissenstests (wenngleich aus diesen Gründen extra eine Voruntersuchung durchgeführt wurde).

ad 3

Mittelwertberechnungen lassen sich streng genommen nur bei kardinalskalierten (metrischskaliert²¹) Daten rechtfertigen (vgl. BORTZ 2005, S. 26, 37). Gerade bei Meinungstests jedoch kann nicht unbedingt Äquidistanz zwischen „sehr zutreffend“ und „zutreffend“ und zwischen „mittel“ und „unzutreffend“ angenommen werden. Dafür ist die Sprache in sich viel zu variabel. Aus diesem Grund empfehlen STEINHAUSEN und LANGER (1977, Kap. 3.2.2, zitiert in BORTZ 2005, S. 568) ordinalskalierte Daten „*künstlich zu dichotomisieren*“. Diese Mediandichotomisierung nivelliert unterschiedliche Abstände der Skalenstufen durch Aufteilung in „größer“ bzw. „kleiner“ Median.

Zwar entsteht durch die Dichotomisierung ein Informationsverlust, jedoch ist dieses Verfahren methodologisch sauberer (ROEDER²²). Wie eigene Vergleichsrechnungen ergeben haben, weicht dieses Verfahren in seinen Ergebnissen kaum vom Likert-Median-Vergleich ab, allerdings nur, wenn ausreichend viele Items pro Test (ca. 20) verwendet wurden, die Items sauber konstruiert (alle Distraktoren gleich attraktiv) und die Daten normalverteilt (volle Ausnutzung der Skala) sind. Zumindest der letzte Punkt dürfte eher selten realisierbar sein, vor allem wenn der Test nicht nur für eine sehr spezielle Zielgruppe konstruiert wurde (z. B. für andere Schulformen). Aus diesen Gründen sollten Wissens- und Meinungstest gleichbehandelt werden.

²¹ Merkmale, die sowohl intervall- als auch verhältnisskaliert sind, nennt man auch *kardinalskaliert* bzw. *metrisch skaliert* (vgl. BORTZ 2005., S. 22).

²² seminarserver.fb14.uni-dortmund.de/roeder/Methodologie/MessAlgo.pdf (Letzter Zugriff am: 2.12.2006)

Um eine unabhängige Untersuchung der Stichproben zu gewährleisten, wurden für jede Untersuchung (VU, HUI, HU II_{intern}, HUII_{extern}) eigene exakte Mediane berechnet. Die exakten Mediane basieren auf der Annahme, dass es sich bei der 5-stufigen Likert-Skala gar nicht um reale Stufen handelt, sondern mehr um eine „gleitende“ Meinung, die mit Hilfe der Stufen erfasst wird. Insofern macht diese Art von „Zwischenstufen“ durchaus Sinn (siehe hierzu auch Punkt 3).

Im Falle des ersten Meinungstests (Einstellung zum Fach Chemie) wurden aus den Scores zusätzlich Residuen berechnet, um die Meinungsänderung zu messen.

Obwohl die herkömmliche Auswertung nach Likert-Skalen mittelwerten für Meinungstests aus den o. g. Gründen nicht zulässig ist, wurden sie aus Gründen der Anschaulichkeit aufgeführt, allerdings nicht in der Auswertung berücksichtigt!

7.2.2 Attraktivität der Unterrichtseinheit

Die Auswertung erfolgte analog zu der des ersten Meinungstests (siehe Kap. 7.2.1) mit dem Unterschied, dass aufgrund des nicht vorhandenen Pre-Tests keine Residuen berechnet wurden.

7.2.3 Attraktivität des Gruppenpuzzles

Für diesen Teil-Test existiert keine Kontrollgruppe; daher soll auf Basis der Mediane lediglich eine Detailanalyse der Items durchgeführt werden. Diese genügt zwar nicht wissenschaftlichen Anforderungen, liefert aber Informationen über das Gruppenpuzzle. Vor allem Gründe, weshalb es aus Sicht der Schüler evtl. beliebter ist und warum es evtl. so gut funktioniert hat, lassen sich so eruieren. Es werden beide Hauptuntersuchungen (nur Gymnasien) zusammengefasst.

7.3 Individuelle Schülerbeurteilungen

Das Verfahren zur Erstellung des Rankings entspricht im Wesentlichen dem des Meinungstests (siehe Kap. 7.2 und 6.3). Daher soll hier nur ein kurzer Abriss erfolgen:

1. Manuelle Eingabe der Rohdaten in den Windows Editor nach Dimensionen (Mitarbeit – Verhalten) und für jede Klasse getrennt
2. Übertragung in SPSS

3. Umkodierung der negativ formulierten Items
4. Berechnung des exakten Medians für jedes Item in einer Klasse
5. Median-Dichotomisierung: Werte $<$ Median werden mit 1 kodiert, Werte $>$ Median werden mit 0 kodiert
6. Mittelwertbildung (Score) über alle Items für jeden Schüler
7. Bildung eines nach Schulklassen gegliederten Rankings auf Basis der Scores

Da dieses Verfahren mit SPSS sehr umständlich ist und am Anfang der Untersuchung wenig Zeit zur Verfügung stand, wurden die Berechnungen und die Erstellung der Scores mit Hilfe des speziell dafür entwickelten Statistikprogramms „Analiese“ von B. ROEDER durchgeführt, jedoch exemplarisch mit SPSS überprüft. Die Auswertung, in diesem Fall die Erstellung der Scores, geschah im Vorfeld der Untersuchung, da die Daten zur Randomisierung benötigt wurden. Cronbachs Alpha wurde auf Basis der umkodierten (negativ zu positiv) Rohdaten berechnet.

Weiterhin sollten die Einschätzungen der Lehrer nach der Untersuchung mit denen vor der Untersuchung verglichen werden. Eine erste Sichtung ergab, dass die Lehrer kaum Änderungen vorgenommen und teilweise sogar die gleichen Bögen nochmals abgegeben haben. Daher soll sich in der Auswertung auf Unterschiede in den beiden Gruppen beschränkt werden: Bei der Analyse sowohl von Pre- und Post-Test als auch der Residuen kam der U-Test nach Mann & Whitney zum Einsatz. Dabei wurde auf das aufwendige Verfahren der Mediandichotomisierung verzichtet. Stattdessen wurden Likert-Mittelwerte verwendet.

7.4 Audioaufzeichnungen

Die MP3-Dateien der Sticks wurden auf die Festplatte kopiert und mit Hilfe des Behavior Recording Systems (BRS, B. ROEDER) aufbereitet. Dazu musste jede Stunde in Echtzeit abgehört und kodiert werden: Immer wenn der Lehrer redete, wurde auf der Computertastatur eine 1 gedrückt, sobald er schwieg, wurde die 2 gedrückt. Es wurden somit Häufigkeiten, wie oft eine Lehrkraft redete bzw. schwieg, und deren Dauern (Intervalle) festgehalten.

Insgesamt gingen sechs Aufzeichnungen in die Auswertung ein, wovon eine Lehrerin mit zwei Klassen vertreten war. Die Kodierung übernahm eine einzige Person, um Fehler zu minimieren.

Analysiert wurden die Daten per Konfigurationsfrequenzanalyse (KFA, vgl. BORTZ 2005, S. 175f). Die KFA setzt als parameterfreies Verfahren keine Normalverteilung oder Linearität der untersuchten Zusammenhänge voraus und findet dann Anwendung, wenn mehrere nominalskalierte Merkmale (hier: FU/GP-Gruppe, redet/schweigt) mit mehreren Stufen verglichen werden sollen (vgl. ROEDER 1974b, S. 819). Die untersuchten Merkmale werden also auf ihre stochastische Unabhängigkeit überprüft (vgl. BORTZ 2005). Man könnte meinen, dass ein einfacher Vergleich der Redezeiten ausreichend statistische Aussagekraft hätte. Im vorliegenden Fall jedoch sind die Daten stark linksgipflig verteilt, was eine Interpretation äußerst schwierig gestaltet. Daher wurde die Rededauer in zwei Gruppen geteilt (Dichotomisierung): *Low* und *High*. *Low* bedeutet kleiner als der gemeinsame Median, *High* bedeutet größer.

Folgende Merkmale wurden miteinander verglichen: FU- bzw. GP-Gruppe (Subject), redet-schweigt (Action) und die Dauer (Duration: Low-High). Überprüft werden sollte, ob Lehrer, die nach der Gruppenpuzzle-Methode unterrichten, überzufällig häufig hohe Rededauern haben. Die Nullhypothese wäre also: Die drei Merkmale sind stochastisch voneinander unabhängig (weitere Erläuterungen hierzu: siehe Kapitel 8.4).

Die entsprechenden Berechnungen waren mit SPSS nicht möglich, weshalb die Analyse und die Erstellung der Grafiken von B. ROEDER mit dem Statistikprogramm „Analiese“ durchgeführt wurden.

Alternativ hätte man auch die Rededauer logarithmieren können und damit eine mehrfaktorielle Varianzanalyse durchführen können. Da die logarithmierte Dauer jedoch noch schwieriger zu interpretieren ist (B. ROEDER, persönliche Mitteilung), wurde die hierarchische KFA gewählt. Das Signifikanzniveau wurde dabei auf $\alpha = 0.01$ festgelegt. Zum Vergleich wurde mit SPSS und Excel die mittlere Rededauer pro Unterrichtsstunde berechnet; sie geht in die statistische Analyse nicht ein.

7.5 Videoaufzeichnungen

Untersuchungen mit Schülern haben den Nachteil, dass sie relativ anfällig sind für Einflüsse von außen. Gerade im Schulalltag können leicht unvorhergesehene Dinge – etwa Krankheitsfälle, Raumänderungen oder spontane Feiern am letzten Schultag der Abiturjahrgänge – eine Untersuchung beeinflussen. Aus diesem Grund wurden Videos von jeder Unterrichtsstunde erstellt. Mit ihrer Hilfe sollten evtl. Besonderheiten/Missstände im Nachhinein aufgedeckt bzw. erklärt werden können.

Die einzelnen Stunden wurden jeweils auf einer 60 minütigen mini-DV Kasette aufgezeichnet und zusätzlich auf VHS (HU I) bzw. DVD (HU II) archiviert. Insgesamt wurden 15 Klassen á 9 Stunden gefilmt, wovon 90 Stunden archiviert worden sind. In die Archivierung nicht ein gingen die Einführungsstunden sowie der jeweilige Test.

7.6 Stundenprotokolle

Die Stundenprotokolle wurden lediglich auf potenzielle Besonderheiten durchgesehen. Direkt im Anschluss an die jeweilige Stunde wurden dem Lehrer fehlende Inhalte mitgeteilt, so dass diese in der folgenden Stunde nachgeholt werden konnten. Fehlende Schüler wurden zwar erfasst, jedoch nicht bei der Auswertung berücksichtigt, da

1. in beiden Gruppen etwa gleich viele Schüler fehlten und
2. nach Angaben der Lehrer alle Unterrichtsinhalte von den Schülern nachgearbeitet werden mussten.

7.7 Beobachtungsbögen

Analog zu den in Abschnitt 7.6 vorgestellten Stundenprotokollen wurden die während des Gruppenpuzzles angefertigten Beobachtungsbögen auf Vollständigkeit der Erklärungen der Schüler durchgesehen. Eine tiefer gehende Analyse des Schülerverhaltens konnte im Rahmen dieser Arbeit nicht mehr durchgeführt werden und war für die Klärung der Hypothesen nicht erforderlich.

7.8 Klassen- und Lehrerfeedback

Die offenen Antworten wurden auf Auffälligkeiten durchgesehen – eine statistische Auswertung erfolgte nicht.

8 Ergebnisse und Diskussion

Wie aus den Kapiteln 6 und 7 hervorgeht, wurde eine Vielzahl an Erhebungs- und Auswertungsinstrumenten eingesetzt. Der Schwerpunkt der Auswertung soll jedoch auf den beiden Haupthypothesen (Wissenszuwachs, Attraktivität, siehe Kapitel 1) liegen.

8.1 Wissenstest

8.1.1 Einfluss der Unterrichtsmethode auf den Lernzuwachs

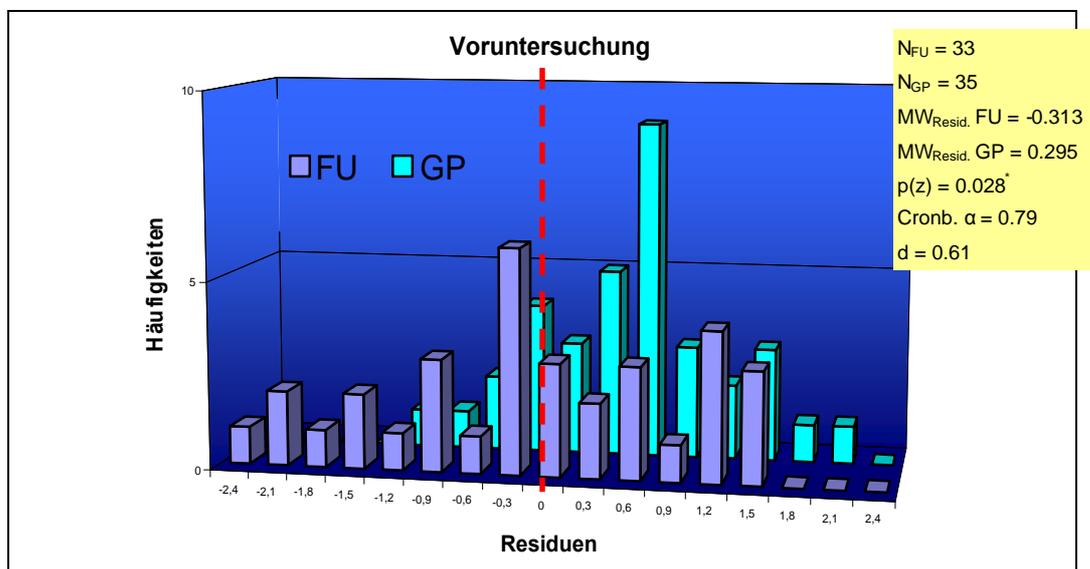


Abbildung 13: Häufigkeitsverteilung der Residuen für die Voruntersuchung. Die gestrichelte Linie kennzeichnet den standardisierten Erwartungswert. Die z-standardisierten Residuen basieren auf den Scores von Vor- und Nachtest. Zum Einsatz kam der U-Test nach Mann und Whitney.

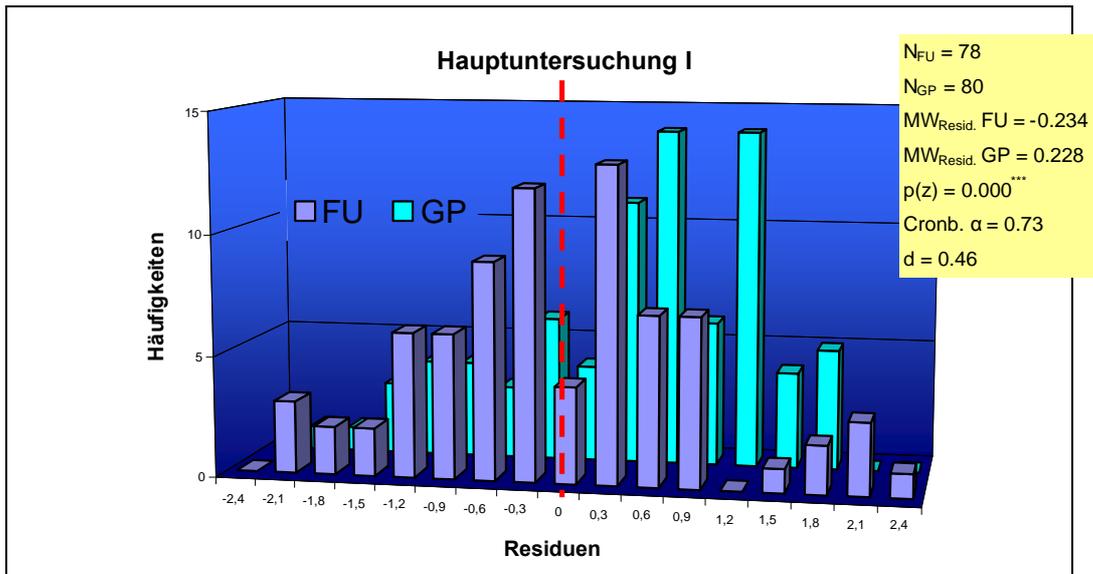


Abbildung 14: Häufigkeitsverteilung der Residuen für die Hauptuntersuchung I. Die gestrichelte Linie kennzeichnet den standardisierten Erwartungswert. Die z-standardisierten Residuen basieren auf den Scores von Vor- und Nachtest. Zum Einsatz kam der U-Test nach Mann und Whitney.

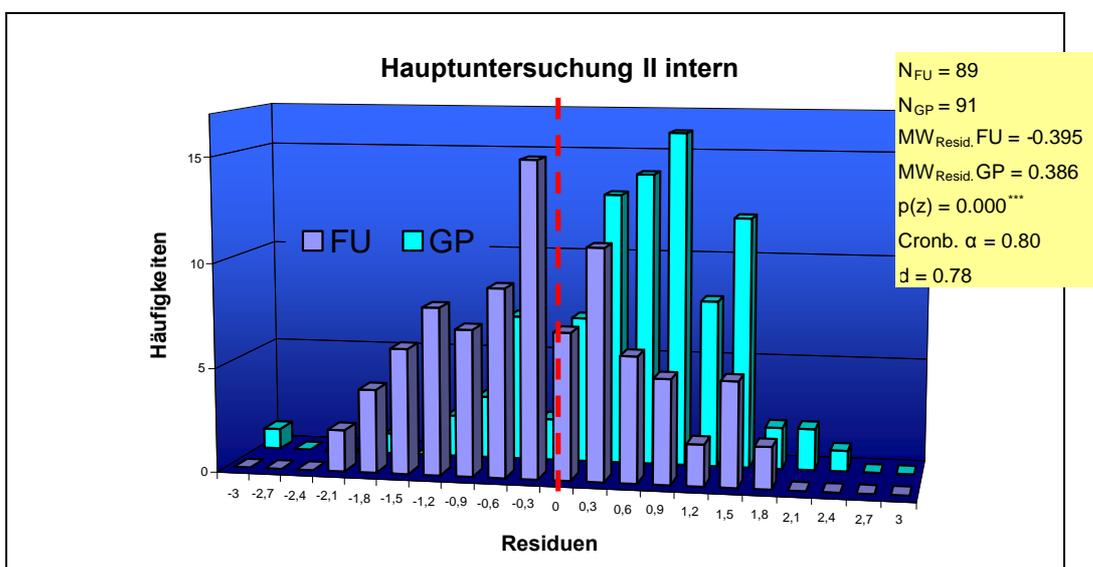


Abbildung 15: Häufigkeitsverteilung der Residuen für die Hauptuntersuchung II_{intern}. Die gestrichelte Linie kennzeichnet den standardisierten Erwartungswert. Die z-standardisierten Residuen basieren auf den Scores von Vor- und Nachtest. Zum Einsatz kam der U-Test nach Mann und Whitney.

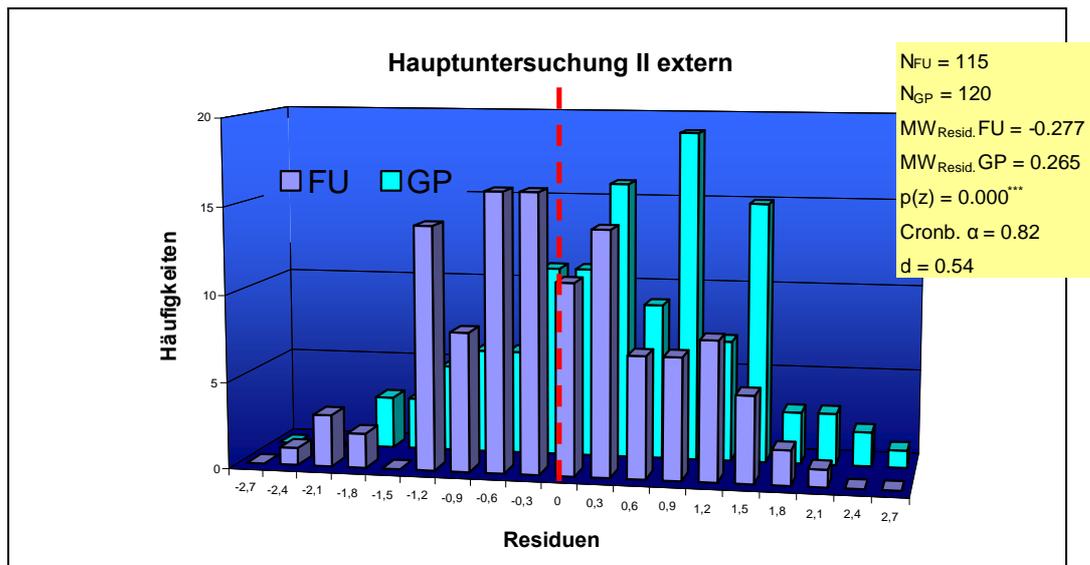


Abbildung 16: Häufigkeitsverteilung der Residuen für die Hauptuntersuchung II_{extern}. Die gestrichelte Linie kennzeichnet den standardisierten Erwartungswert. Die z-standardisierten Residuen basieren auf den Scores von Vor- und Nachtest. Zum Einsatz kam der U-Test nach Mann und Whitney.

In allen Hauptuntersuchungen zeichnet sich das gleiche Bild ab: Schüler, die nach der Gruppenpuzzle-Methode unterrichtet wurden, erreichen signifikant höhere Lernzuwächse als die entsprechende FU-Gruppe, erkennbar an der Rechts-Verschiebung der türkisen Balken in Abb. 13-16. Während die Residuen der Interventionsgruppe (GP) von 0.23 bis 0.39 Einheiten über dem mittleren Erwartungswert liegen, weichen die der Kontrollgruppe (FU) 0.23 bis 0.4 Einheiten nach unten ab. Das entspricht stellenweise mehr als einer halben Standardabweichung. Die Unterschiede sind auf dem 5 %-Niveau signifikant (VU und HU I); für die HU II gilt ein Signifikanzniveau von 0.01 (echte Replikation), das ebenfalls unterschritten wird (siehe Tabelle 6). Cronbachs Alpha liegt bei mindestens 0.73. Damit konnten die Ergebnisse der Voruntersuchung bestätigt werden.

Wegen der besseren Vorstellbarkeit sind im Folgenden die Scores der Wissenstests aller Untersuchungen aufgeführt.

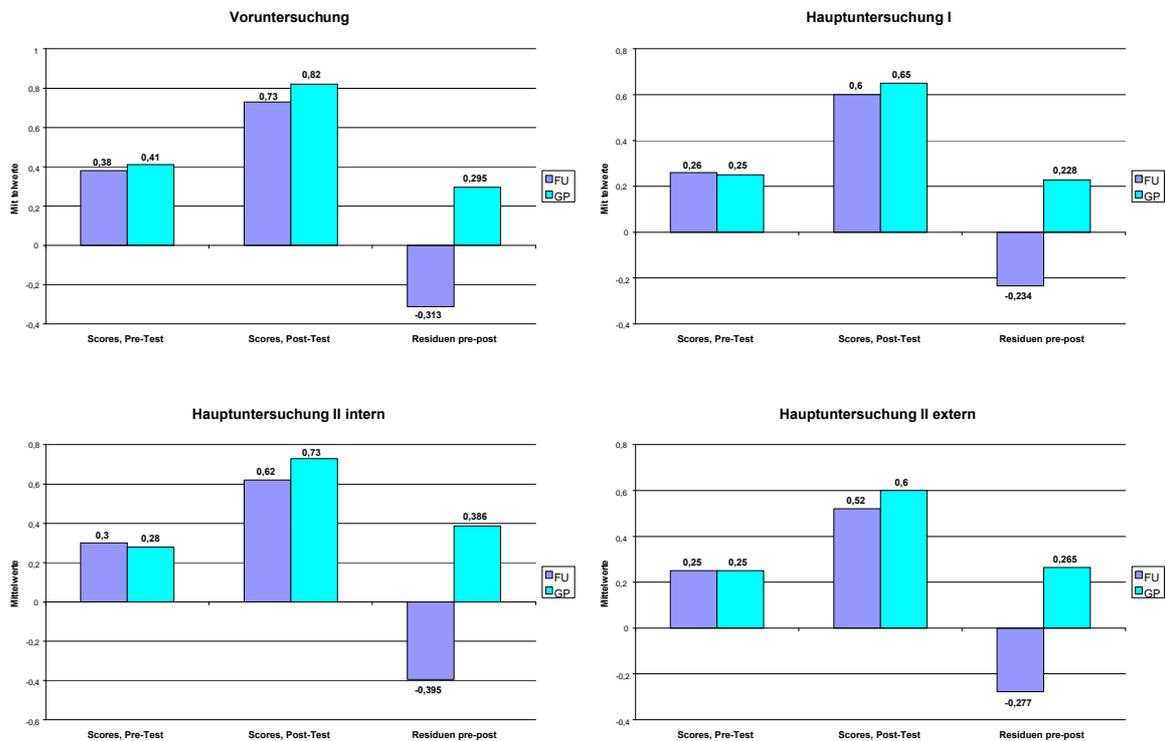


Abbildung 17: Ergebnisse des Wissenstests für Vor- und Hauptuntersuchungen. Dargestellt sind sowohl Residuen als auch Scores zum Vergleich.

Da die vier Untersuchungen tendenziell gleiche Ergebnisse liefern, sollen sie zusammen beschrieben werden: Vor der Intervention lagen beide Gruppen mit ca. 25-30 % korrekter Antworten ungefähr gleich auf. Lediglich in der Voruntersuchung werden höhere Werte erreicht. Nach der Intervention konnten GP-Schüler mit 60-73 % zutreffender Antworten im Schnitt acht Prozentpunkte mehr lösen als die Kontrollgruppe mit 52-60 %. Auch hier liegen die Werte der Voruntersuchung deutlich darüber. Der Vorteil für das Gruppenpuzzle bleibt auch nach Umrechnung in z-standardisierte Residuen für alle vier Untersuchungen bestehen (siehe Tabelle 6). Erwartungsgemäß unterscheiden sich die Residuen der Voruntersuchung nun nicht mehr gravierend von denen der Hauptuntersuchung, da das bessere Abschneiden im Vortest berücksichtigt wurde.

Tabelle 6: Vergleich der Unterrichtsmethoden: Ergebnisse des U-Tests von Vor- und Hauptuntersuchung. Cronbachs α wurde für den Post-Test bestimmt; die Stichprobengröße, Effektstärke und die Signifikanztests beziehen sich jeweils auf die Residuen, welche aus Pre- und Post-Test errechnet wurden. Beim t-Test wurde die Varianzhomogenität berücksichtigt. Soweit nicht anders angegeben beziehen sich die Irrtumswahrscheinlichkeiten auf die einseitige Testdurchführung.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 33 GP: 35	FU: 78 GP: 80	FU: 89 GP: 91	FU: 115 GP: 120
MW Scores, Pre-Test	FU: 0.38 GP: 0.41	FU: 0.26 GP: 0.25	FU: 0.30 GP: 0.28	FU: 0.25 GP: 0.25
MW Scores, Post-Test	FU: 0.73 GP: 0.82	FU: 0.60 GP: 0.65	FU: 0.62 GP: 0.73	FU: 0.52 GP: 0.60
MW Residuen _{pre-post}	FU: -0.313 GP: 0.295	FU: -0.234 GP: 0.228	FU: -0.395 GP: 0.386	FU: -0.277 GP: 0.265
t-Test Residuen, $p_{(z)}$	0.012* (2-seitig)	0.002**	0.000***	0.000***
U-Test Residuen, $p_{(z)}$	0.028* (2-seitig)	0.000***	0.000***	0.000***
Cronbachs α	0.79	0.73	0.80	0.82
Effektstärke d	0.61	0.46	0.78	0.54

Tabelle 6 gibt im Wesentlichen die Befunde der Grafiken wieder. Zusätzlich ist die Effektstärke aufgeführt. Sie liegt zwischen 0.5 (mittlerer Effekt) und 0.8 (starker Effekt). Vergleicht man die Ergebnisse des t-Tests mit denen des U-Tests, so ergeben sich keine gravierenden Unterschiede: Lediglich in der VU weichen sie voneinander ab.

Der Kolmogorov-Smirnov-Test (KS-Test) auf Normalverteilung bestätigt bis auf die Gruppenpuzzlegruppe der HU I eine Normalverteilung der Stichproben (siehe Tabelle 7).

Tabelle 7: Überprüfung der Stichproben auf Normalverteilung: Ergebnisse des Kolmogorov-Smirnov-Tests für den Wissenstest auf Basis der Residuen. Normalverteilung kann angenommen werden ab einem $p(z) \geq 0.2$, da es sich hierbei um die Wunschhypothese handelt. Fettgedruckt sind möglicherweise nicht normalverteilte Stichproben.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 33 GP: 35	FU: 78 GP: 80	FU: 89 GP: 91	FU: 115 GP: 120
Kolmogorov-Smirnov-Z	FU: 0.436 GP: 0.566	FU: 0.594 GP: 1.306	FU: 0.598 GP: 0.892	FU: 0.493 GP: 0.698
KS-Test $p_{(z)}$, (2-seitig)	FU: 0.991 GP: 0.905	FU: 0.872 GP: 0.066	FU: 0.867 GP: 0.404	FU: 0.968 GP: 0.714

Diskussion

Schüler der Interventionsgruppe schneiden in allen Untersuchungen (VU – HU II) signifikant besser beim Wissenstest ab als ihre Mitschüler aus der Frontalunterricht-Gruppe (vgl. auch TEPNER/MELLE/ROEDER 2006, S. 275 ff). Damit konnten die Ergebnisse der HU I repliziert werden. Die Resultate der Voruntersuchung lagen in allen Tests über denen der Hauptuntersuchung, da sich der Test noch in der Erprobungsphase befand und für die Hauptuntersuchung erschwert wurde. Interessant ist, dass die Anwesenheit der Beobachter keinen Einfluss auf den Lernzuwachs der Gruppenpuzzlegruppe zu haben scheint, da in der zweiten Hauptuntersuchung (bis auf 2 Lehrer intern) auf Beobachtungen verzichtet wurde.

Vergleich mit internationalen Studien

Die Ergebnisse decken sich mit internationalen Studien (z. B. DE BAZ 2001, DORI et al. 1985 oder LAZAROWITZ & KARSENTY 1990). Eine Vergleichbarkeit erscheint jedoch problematisch. DE BAZ 2001 beispielsweise untersuchte lediglich zwei 7. Klassen mit je 40 Schülern in Jordanien. Die Vergleichbarkeit der Experimental- und Kontrollklasse wurde über einen Pre-Wissenstest sichergestellt, sowohl Gruppenpuzzleklasse als auch „traditionelle“ Klasse wurden vom selben Lehrer unterrichtet. Die Varianzanalyse der Mittelwerte ergab einen signifikanten Vorteil für die Gruppenpuzzlegruppe. Eine Übertragung der Ergebnisse auf deutschen Unterricht erscheint einerseits wegen der größeren Lerngruppen schwierig, andererseits wurden Mittelwerte für die Signifikanztests verwendet. Die geringe Stichprobengröße lässt zudem keine belastbaren Zahlen erwarten. Ferner werden zum Ma-

terial (Thema: „Lebender Organismus“) und zur Durchführung (z. B. Dauer) kaum Angaben gemacht. Etwas aussagekräftiger erscheint die Untersuchung von DORI et al. (1985). In dieser behandelten vier 8. Klassen mit insgesamt 112 Schülern in Israel den Zellaufbau. Kritisch anzumerken ist bei der Auswertung auch hier der Verzicht auf Residuen, es wurden lediglich Pre-Testwerte von Post-Testwerten subtrahiert, wobei die Kontrollgruppe im Vortest besser abschnitt und somit per se weniger dazu lernen konnte.

Vergleich mit nationalen Studien

Da eine empirische Vergleichsgruppenuntersuchung im Chemieunterricht bislang nicht existiert (Stand: April 2007), muss auf andere Fächer bzw. Schulformen zurückgegriffen werden. BORSCH et al. (2002) verglichen das Gruppenpuzzle in acht 3. und 4. Klassen im Sachunterricht der Grundschule mit herkömmlich unterrichteten Kontrollklassen. Dabei konnte ein deutlich größerer Zuwachs deklarativen Wissens der nach dem Gruppenpuzzle unterrichteten Kinder festgestellt werden, unabhängig von ihrem jeweiligen Vorwissensniveau (vgl. ebd., S. 180). Allerdings verwendeten sie statt Residuen lediglich die Differenzen zwischen Vor- und Nachtest für ihre Signifikanzanalyse mittels t-Test (ebd., S. 178), was die Vergleichbarkeit mit dieser Studie erschwert.

Zu abweichenden Ergebnissen kamen BERGER & HÄNZE 2004 im Fach Physik (siehe auch Kapitel 2.6). Im Fokus ihrer Untersuchung waren Unterrichtseinheiten zum Rasterelektronenmikroskop und Mikrowellenofen in der Jahrgangsstufe 12. Hierbei erhielten alle Schüler das Gruppenpuzzlematerial und nicht nur die jeweiligen Experten (mündliche Mitteilung M. HÄNZE 2004). Trotzdem konnte die Interventionsgruppe keinen Leistungsvorteil gegenüber der Frontalunterrichtsgruppe erzielen, wenngleich sie auch nicht signifikant schlechter abschnitt (BERGER & HÄNZE ebd., S. 216). Die Ergebnisse sind besonders aussageschwach, da nicht nur auf eine Auswertung über Residuen verzichtet wurde, sondern in Pre- und Post-Wissenstest unterschiedliche Aufgaben gestellt wurden (vgl. BERGER & HÄNZE ebd., S. 215).

Auch JÜRGEN-LOHMANN et al. (2001) bestätigen den Vorteil für das Gruppenpuzzle nicht. Allerdings wurde bei dieser Untersuchung nicht nur die Methode variiert. Zusätzliche Faktoren wie z. B. die intensivere Betreuung der Kontrollgruppe beeinflussten die Ergebnisse. Insofern scheint ein Vergleich wieder problematisch.

Zur Statistik der vorliegenden Studie

Insgesamt weisen die Untersuchungen gute Reliabilitäten auf (siehe Tabelle 6). Das relativ „schlechte“ Cronbachs Alpha und die geringe Effektstärke der HU I kommen durch die Verschärfung der Bedingungen zustande: Wurde ein Lerninhalt im Frontalunterricht vom Lehrer nicht oder nur unvollständig dargeboten, erhielten alle Schüler der FU-Gruppe für die entsprechende Aufgabe im Test einen Punkt – auch wenn sie die Aufgabe vielleicht nicht hätten lösen können. Dieses war bei zwei Klassen in der HU I der Fall, in denen die Lehrkraft es nicht schaffte, in der vorgegebenen Zeit alle Lerninhalte darzubieten (bei der HU II war dies nur in einem Fall mit je einem Item erforderlich). Nimmt man die entsprechenden Items aus der Analyse heraus, so ergibt sich ein Cronbachs Alpha von 0.764. Diese Maßnahme macht die Untersuchung zwar konservativer, führt jedoch auch zu einer schlechteren inneren Konsistenz des Wissenstests. Dass diese aber eigentlich sehr gut war, zeigen die α -Werte der anderen Hauptuntersuchungen, schließlich handelte es sich durchgängig um denselben Test. Zusätzlich tragen die Transferfragen (wie bei jeder Stichprobe) erheblich zur schlechten Reliabilität bei.

t-Test vs. U-Test

Grundsätzlich bestehen keine gravierenden Unterschiede zwischen den beiden Testarten. Bei relativ kleinen Stichproben wie in der Voruntersuchung scheint der t-Test jedoch eher zugunsten der H_1 zu entscheiden; der U-Test misst dementsprechend konservativer. Ein Einsatz des t-Tests verbietet sich aber aufgrund der Nicht-Normalverteilung der HU I (GP-Gruppe) (siehe Tabelle 7, siehe Kapitel 7.1.1). Dadurch wird ein Vergleich mit anderen Untersuchungen, die ihn dennoch verwenden, schwierig. Dem Argument der Vergleichbarkeit darf insofern nicht viel Relevanz beigemessen werden, da viele Untersuchungen per se nicht vergleichbar sind. So macht es wenig Sinn, nur Irrtumswahrscheinlichkeiten miteinander zu vergleichen, wenn hingegen Vergleiche der Konzeption, der Durchführung sowie der Dateneingabe selten bis gar nicht stattfinden und durch fehlende Information auch nicht möglich sind. Insofern ist der U-Test als Testinstrument für zwei unabhängige Stichproben besser geeignet.

Determinanten erfolgreicher Gruppenpuzzles

Mehrere Erklärungen für das gute Abschneiden des Gruppenpuzzles scheinen plausibel. Das Gruppenpuzzle vereint und unterstützt mindestens acht von zehn Merkmalen „guten Unterrichts“ (vgl. MEYER, 2003, S. 36 ff). Zwar bezieht er sich bei seinen aus zahlreichen

Studien zusammengefassten Determinanten, die erfolgreichen Unterricht ausmachen, auf einen eher frontal organisierten Unterricht (vgl. ebd., S. 37), jedoch scheinen gerade diese Merkmale besonders gut mit dem Gruppenpuzzle umgesetzt werden zu können.

Zum einen weist das Gruppenpuzzle konstruktionsbedingt eine klare Strukturiertheit auf. Diese bezieht sich nicht nur auf das Layout, sondern vielmehr auf den Unterrichtsablauf: Eine starke Gliederung in die drei Phasen sowie eine enge zeitliche Vorgabe mögen einigen Schülern geholfen haben, sich über den Prozess, die Ziele und Inhalte klar zu werden. Jeder Schüler wusste zu jedem Zeitpunkt, was er zu tun hatte (Rollenklarheit). Hinzu kommt eine starke inhaltliche Trennung der drei Expertenthemen.

Die scheinbar gute und mittels einer Voruntersuchung sichergestellte Verständlichkeit der Aufgabenstellung, Monitoring des Lernverlaufs durch den Lehrer, Plausibilität des thematischen Gangs sowie eine verbindliche, klar formulierte Ergebnissicherung (Lernkontrolle) sorgten zum anderen für eine inhaltliche Klarheit bei den Schülern.

Das Gruppenpuzzle per se, aber besonders das verwendete Gruppenpuzzle Seifen, zeichnet sich durch Methodenvielfalt aus: Es werden nicht nur Informationen in Textform dargeboten, die selbstständig in Einzelarbeit erarbeitet werden müssen. Der zeitlich überwiegende Teil findet in Gruppenarbeit statt, ergänzt durch den Einsatz von einfachen Schülerexperimenten. Bei der Vermittlung werden mehrere Lernkanäle – z. B. Grafiken für visuelle Lerntypen, verbale Informationen für auditive Lerntypen oder Experimente für die praktisch Lernenden – verwendet. Das klassische Unterrichtsgespräch findet gar nicht mehr statt, davon übrig geblieben ist lediglich ein kleiner Lehrervortrag zum Ablauf des Gruppenpuzzles. So kann mit dem Gruppenpuzzle eine Vielfalt an Handlungsmustern umgesetzt werden.

Zum hohen Anteil echter Lernzeit trugen die gute Vorbereitung insbesondere der Versuche sowie die klare Aufgabenstellung in einem abgesteckten zeitlichen Rahmen bei. Diese Faktoren mögen auch den Verlauf des Frontalunterrichts positiv beeinflusst haben. Allerdings verweisen JOHNSON & JOHNSON (1999, S. 205) auf 30 Studien, in denen Kooperative Lerner signifikant mehr „time on task“ aufbringen als individuell bzw. wettbewerblich Lernende.

MEYER führt weiterhin Sinnstiftendes Kommunizieren an. Aus lerntheoretischer Sicht unterstützt *„die Auseinandersetzung mit und die Kommunikation über einen Sachverhalt aus verschiedenen Perspektiven sowie das Verknüpfen mit ähnlichen Sachverhalten in der*

Kleingruppe“ (ENTWISTLE 1995, zitiert in GREEN & GREEN 2005, S. 20) das Lernen, wobei so genannte *Deep-processing-strategies* zum Tragen kommen. Darunter verstehen (WILD, KRAPP & WINTELER 1992, S. 283 ff, zitiert in GREEN & GREEN 2005 ebd.) „*das Bemühen von Lernenden, die tiefere Bedeutung eines Sachverhaltes zu verstehen und Beziehungen zu anderen Wissens- oder Erfahrungsbereichen herzustellen, anstatt sich mit Surface-level-strategies und damit im Wesentlichen mit dem Einprägen von Sachwissen o. Ä. zu begnügen*“. Der im gegenseitigen Austausch von Informationen, Gedanken, Fragen stattfindende Lehr-Lernprozess und die damit verbundenen persönliche Identifikation mit dem Lerngegenstand verleiht ihm eine für die Schüler persönliche Bedeutung (vgl. MEYER 2004, S. 67).

Besonders das Gruppenpuzzle schafft ein für guten Unterricht notwendiges lernförderliches Unterrichtsklima, da es zwangsläufig viel von seinen Teilnehmern fordert: gegenseitigen Respekt der Gruppenmitglieder, verlässlich eingehaltene Regeln, Verantwortungsübernahme für jede Rolle (vgl. MEYER 2004, S. 47). Variieren kann man bei wiederholter Durchführung des Gruppenpuzzles die Zusammenstellung der Gruppen, so dass nicht immer die gleichen Schüler in einer Gruppe zusammen arbeiten, wenngleich sich nicht alle Schüler gleichermaßen sympathisch sind. Aber gerade hier liegen die Stärken des Gruppenpuzzles, da es die Schüler praktisch „zu einer Zweckgemeinschaft zusammenschweißt“ (*positive Abhängigkeit*, siehe Kapitel 2.2).

Individuelles Fördern erfolgt beim Gruppenpuzzle auf verschiedene Art und Weise: Zum einen kann jeder Schüler sein Lerntempo, seine Lernwege und auch inhaltliche Schwerpunkte im Rahmen der vorgegeben Zeit selbst bestimmen (vgl. BOSSE 2003, S. 27). Diese innere Differenzierung bezieht sich vor allem auf die Erarbeitungsphase, aber auch in der Vermittlungsphase konnten die Schüler die Methode frei wählen. Gleichzeitig können hier eher schlechte Schüler in Gruppen mit guten Schülern integriert werden. Trotz zeitlicher Vorgaben herrschte kein Zeitdruck, da die Zeiten ausreichend dimensioniert waren.

Auch dem Lehrer steht mehr Zeit für die individuelle Förderung zur Verfügung, da er gezielt in die Kleingruppen hineingehen kann. Um die hier vorliegenden Untersuchungsergebnisse nicht zu verfälschen, wurden die Lehrer dieser Studie jedoch gebeten, keine individuellen Förderungen vorzunehmen. Diese erfolgte, wenn überhaupt, durch Mitschüler.

Für das erfolgreiche Abschneiden im Test waren transparente Leistungserwartungen (vgl. MEYER 2004, S. 113 ff) mitverantwortlich. Mit Hilfe der auf den Arbeitsblättern aufge-

fürten Lernziele sowie der wichtigen Begriffe wussten die Schüler nicht nur, was sie ihren Mitschülern vermitteln sollten, sondern ebenfalls, was von ihnen im Test gefordert wurde (vgl. FREY-EILING & FREY 1999). Aus neurobiologischer Sicht macht ein solcher Überblick Sinn, da die Schüler so auf die Aufnahme der Inhalte vorbereitet werden (vgl. BRAND & MARKOWITSCH 2004, S. 6). Sie begründen dieses Vorgehen mit zwei Funktionen: *„erstens verarbeiten Schüler unbewusst [...] bereits die wesentlichen Aspekte, was eine Bahnung für eine spätere tiefe Verarbeitung bedeutet. Zweitens dienen Überblick und Zieldefinition auch als Strukturierung bei der Einspeicherung, was eine tiefere Verarbeitung im Sinne der Bildung geeigneter Assoziationen wahrscheinlicher macht“* (ebd., S. 6).

Die Transparenz der Lernziele galt eingeschränkt ebenso für die Frontalunterrichtsgruppe, da die Lehrer ebenfalls die Lernziele hätten nennen können, was allerdings nur in einigen Fällen erfolgte. So präsentierte z. B. eine Lehrerin vor der Unterrichtseinheit einen *Advance Organizer* an der Tafel. Diese von AUSUBEL (1980) entwickelte „Lernlandkarte“ wird den Schülern vor der eigentlichen Stoffarbeit dargeboten und stellt eine Visualisierung der Lerninhalte und ihrer Zusammenhänge im Sinne einer nichtlinearen Didaktik dar²³. Im Prinzip sollte diese Vorausbeschreibung so gut sein, *„dass die Lernenden das Wesentliche sofort begreifen“* (vgl. FREY-EILING & FREY 1999, S. 55). Dabei erleichtert die Vorstrukturierung die Verknüpfung von neuem Stoff mit bereits vorhandenem Wissen (vgl. AUSUBEL 1980); der *Advance Organizer* basiert dementsprechend auf dem lerntheoretischen Konstruktivismus. Diese Lerntheorie geht davon aus, dass *„die Lehrperson ihr Wissen [keineswegs] durch schlichte Mitteilung aus ihrem Kopf in die Köpfe der Lernenden übertragen“* kann (STORK 1995, S. 17). Grund ist, dass *„wichtige Teile der Naturwissenschaften“*; [also Begriffe und Regeln,] *„in einem Zusammenhang stehen, ein Netz bilden. Und aktives, das heißt anwendbares, vielfach verknüpftes Wissen dieser Art läßt sich nur erwerben, indem jedes lernende Individuum es für sich konstruiert“* (STORK, ebd.).

Eine fertige Übersicht im Sinne eines *Advance Organizers (AO)*, wie sie FREY-EILING & FREY (1999) empfehlen, wurde den Schülern zwar nicht gegeben, um Effekte des Gruppenpuzzles nicht durch den AO zu verfälschen. Dennoch sind einige Elemente des AOs im Gruppenpuzzle Seifen wiederzufinden. So geben die am Ende des Expertenmaterials aufgeführten Lernziele zusammen mit den wichtigen Begriffen einen Überblick über den zu

²³ vgl. http://lehrerfortbildung-bw.de/unterricht/sol/07_advance_organizer/

lernenden Stoff. Zur Effektivität des *Advance Organizers* im Chemieunterricht: vgl. ZIMMER & MELLE 2007.

Ferner findet im Gruppenpuzzle ein Selbstorganisieren des Wissens zumindest ansatzweise während der Expertenphase statt, in der die Schüler sich selbstständig einen Vortrag erarbeiten, mit dem sie ihr Expertenwissen den anderen Schülern vermitteln. Dabei fließen zwangsläufig auch eigene Gedanken und Vorstellungen mit ein.

Ebenfalls zum guten Ergebnis beigetragen haben mögen die Arbeitsblätter – speziell die Selbstkontrollen der Schüler, etwa durch Protokollbögen während der Expertenphase. Durch die Neuigkeit der Methode hervorgerufene motivationale Aspekte dürften eine eher untergeordnete Rolle gespielt haben. Schließlich war einigen Schülern das Gruppenpuzzle bekannt. Bedeutsamer ist vermutlich die Perspektive der sozialen Kohäsion (vgl. GRÄSEL & GRUBER 2000, S. 164). In diesem motivationalen Ansatz wird davon ausgegangen, dass Schüler durch den sozialen Zusammenhalt motiviert werden, gemeinsam zu lernen, „*weil ihnen aneinander und an der Gruppe etwas liegt*“ (ebd.). Insofern spielt die Motivation durch die Gruppe an sich eine entscheidende Rolle.

Den Schlüssel für den Lernerfolg beim Gruppenpuzzle sieht MEYER (2007, S. 3) jedoch in der wechselseitigen Abhängigkeit (siehe Abschnitt 2.2 und vgl. COHEN 1993), „*da alle Schüler auf ihre Gruppenkameraden angewiesen sind; diese müssen die notwendigen Informationen in die Gruppe einbringen, aufgrund ihres Expertenwissens Fragen beantworten können usw., damit alle Gruppenmitglieder die Gesamtinformation, die zum Test nötig ist, erhalten*“²⁴. Jedes Mitglied muss in die Rolle des Lehrenden wie auch des Lernenden schlüpfen.

Dabei scheint besonders die Lehrerrolle für die Lernerfolge verantwortlich zu sein: Nach RENKL (1997) führt die so genannte „Lehrerwartung“ zu einer höheren Motivation bei den Schülern (siehe hierzu auch Abschnitt 8.1.2). DECI & RYAN (1993, S. 229) postulieren in ihrer Selbstbestimmungstheorie drei angeborene psychologische Bedürfnisse, die einen motivationalen Einfluss haben: Das Bedürfnis nach Kompetenz, nach Autonomie der Selbstbestimmung und nach sozialer Eingebundenheit. Alle drei Bedürfnisse sowie die intrinsische Motivation allgemein können durch das Gruppenpuzzle im Vergleich zum

²⁴ <http://www.member.uni-oldenburg.de/hilbert.meyer/download/Gruppenpuzzle-einzeln.pdf> (Letzter Zugriff am 3.4.2007)

Frontalunterricht deutlich erhöht werden (vgl. BERGER & HÄNZE 2003, S. 247 ff). Sie vermuten, dass besonders die Förderung des Kompetenzerlebens „*ein Schlüssel für eine gleichzeitige Steigerung der Leistungen*“ ist (vgl. BERGER & HÄNZE 2004b, S. 33).

Schlussendlich mag auch der zu erwartende Abschlusstest für eine erhöhte Motivation bei den Schülern gesorgt haben, was allerdings auf die Kontrollgruppe gleichermaßen zutrifft.

Generell gilt, dass kooperative Unterrichtsformen stark auf die Selbstständigkeit und die Verantwortungsübernahme der Schüler für ihr Lernergebnis (und das der Gruppe) setzen (siehe Abschnitt 2.2). Hinzu kommt die Arbeit in Kleingruppen: Evtl. trauen sich Schüler eher, ihren Mitschülern Fragen zu stellen als dem Lehrer. Eine Art „Bloßstellen“ vor der Klasse und die damit verbundene Zurückhaltung wird somit vermieden. FROTSCHER (2004, S. 9) konstatiert: „*Lernschwache Kinder sind mit der Kleingruppe im Hintergrund oft mutiger bei sprachlichen Äußerungen und anderen Lernaktivitäten und trauen sich eher, noch vorhandene Schwierigkeiten anzusprechen*“.

Zudem herrscht in Kleingruppen eine höhere Aktivität als im Frontalunterricht, die eventuelle schlechtere Erklärungsleistungen der Mitschüler kompensiert. Zusätzlich ist die Möglichkeit, Anstöße und Hilfestellungen von den Gruppenmitgliedern zu erhalten, größer (vgl. HUBER 1999). Sie erklärt den Erfolg des Gruppenlernens auch durch das größere (kumulative) Wissenspotenzial, das durch die größere Anzahl der „Mitdenker“ entsteht.

Diese im Vergleich zur Standardsituation in der Klasse sehr informelle Situation der Kleingruppe lässt Schüler untereinander stärker umgangssprachlich miteinander kommunizieren. So kommen Schüler leichter zu Wort, die Probleme mit der im Plenum favorisierten Hochsprache haben (vgl. HAAG 2005, S. 28).

Klassenweise Analyse

Nach wissenschaftlichen (statistischen) Kriterien ist eine klassenweise Analyse nicht zulässig: Die Stichprobengröße wäre für belastbare Aussagen erstens zu klein, und zweitens hätte für einen Vergleich der Klassen untereinander im Vorfeld sichergestellt werden müssen, dass diese tatsächlich vergleichbar sind, was jedoch aus oben genannten Gründen kaum möglich ist (siehe Abschnitt 5.4.1). Zur Vervollständigung soll dennoch eine kurze Analyse der Gruppen innerhalb einer Klasse erfolgen.

Grundsätzlich spiegelt sich das Bild der vier Hauptuntersuchungen in fast allen Klassen wider: Gruppenpuzzle-Schüler schneiden besser beim Wissenstest ab als frontal unterricht-

tete Schüler. Die Unterschiede in den Klassen sind trotz der geringen Schülerzahl in sieben Fällen signifikant. Lediglich in Klasse 14 und 21 (siehe Tabelle 8) kehrt sich das Bild um, wobei die Unterschiede nicht signifikant sind.

Tabelle 8: Klassenweiser Vergleich der Unterrichtsmethoden: Ergebnisse des U-Tests nach Mann und Whitney (zweiseitig). Aufgeführt sind sowohl Klassen (mp-Prinzip) als auch Gruppen (parallel Design). Die Residuen basieren auf Pre- und Post-Test und wurden für jede Lerngruppe separat berechnet. Fett gedruckt sind Klassen, in denen die GP-Gruppe schlechter abgeschnitten hat. Signifikante Unterschiede sind mit entsprechenden Sternchen markiert. mp bedeutet Matched-Pairs-Design, para = parallel. Die Lerngruppen 4 und 12 sind etwas größer, da hier zwei kleinere Klassen zusammengelegt werden konnten. Cronbachs Alpha wurde nicht berechnet, da die Auswertung nach Klassen aufgrund ihrer kleinen Stichprobengröße keinerlei statistische Aussagekraft besitzt. Die einzelnen Untersuchungen (VU, HU I, HU II_{intern}, HU II_{extern}) sind durch unterschiedliche Graustufen gekennzeichnet.

Klasse/Gruppe (Design)	N	MW _{Scores, pre}	MW _{Scores, post}	MW _{Residuen}	U-Test _{Res, p(z)}
1 (mp)	FU: 11 GP: 12	FU: 0.34 GP: 0.39	FU: 0.68 GP: 0.77	FU: -0.186 GP: 0.171	0.786
2 (mp)	FU: 11 GP: 12	FU: 0.38 GP: 0.42	FU: 0.71 GP: 0.84	FU: -0.413 GP: 0.378	0.079
3 (mp)	FU: 11 GP: 11	FU: 0.42 GP: 0.42	FU: 0.79 GP: 0.85	FU: -0.370 GP: 0.370	0.076
4 (mp)	FU: 17 GP: 18	FU: 0.21 GP: 0.23	FU: 0.57 GP: 0.63	FU: -0.187 GP: 0.176	0.258
5 (mp)	FU: 13 GP: 13	FU: 0.24 GP: 0.25	FU: 0.53 GP: 0.55	FU: -0.062 GP: 0.062	0.687
6 (mp)	FU: 13 GP: 13	FU: 0.31 GP: 0.25	FU: 0.66 GP: 0.68	FU: -0.253 GP: 0.253	0.113
7 (mp)	FU: 12 GP: 12	FU: 0.28 GP: 0.26	FU: 0.61 GP: 0.64	FU: -0.216 GP: 0.216	0.291
8 (mp)	FU: 12 GP: 13	FU: 0.26 GP: 0.27	FU: 0.59 GP: 0.74	FU: -0.719 GP: 0.664	0.000 ^{***}
9 (mp)	FU: 11 GP: 11	FU: 0.26 GP: 0.28	FU: 0.61 GP: 0.68	FU: -0.220 GP: 0.220	0.217
10 (mp)	FU: 14 GP: 13	FU: 0.29 GP: 0.32	FU: 0.64 GP: 0.73	FU: -0.206 GP: 0.222	0.550
11 (mp)	FU: 14 GP: 15	FU: 0.32 GP: 0.31	FU: 0.66 GP: 0.83	FU: -0.642 GP: 0.599	0.001 ^{**}

Klasse/Gruppe (Design)	N	MW _{Scores, pre}	MW _{Scores, post}	MW _{Residuen}	U-Test _{Res, p(z)}
12 (mp)	FU: 17 GP: 18	FU: 0.33 GP: 0.27	FU: 0.61 GP: 0.71	FU: -0.386 GP: 0.365	0.004**
13 (mp)	FU: 14 GP: 15	FU: 0.25 GP: 0.23	FU: 0.45 GP: 0.68	FU: -0.686 GP: 0.640	0.000***
14 (mp)	FU: 15 GP: 15	FU: 0.28 GP: 0.26	FU: 0.70 GP: 0.64	FU: 0.170 GP: -0.170	0.436
15 (mp)	FU: 15 GP: 15	FU: 0.31 GP: 0.32	FU: 0.67 GP: 0.82	FU: -0.566 GP: 0.566	0.002
16 (mp)	FU: 13 GP: 14	FU: 0.25 GP: 0.28	FU: 0.62 GP: 0.81	FU: -0.660 GP: 0.613	0.000***
17 (mp)	FU: 10 GP: 10	FU: 0.31 GP: 0.36	FU: 0.70 GP: 0.82	FU: -0.334 GP: 0.334	0.165
18 (para)	FU: 14 GP: 15	FU: 0.19 GP: 0.23	FU: 0.33 GP: 0.41	FU: -0.235 GP: 0.219	0.252
19 (mp)	FU: 14 GP: 14	FU: 0.23 GP: 0.23	FU: 0.51 GP: 0.55	FU: -0.183 GP: 0.183	0.306
20 (para)	FU: 23 GP: 19	FU: 0.27 GP: 0.24	FU: 0.50 GP: 0.54	FU: -0.186 GP: 0.225	0.129
21 (mp)	FU: 9 GP: 16	FU: 0.28 GP: 0.21	FU: 0.60 GP: 0.47	FU: 0.261 GP: -0.147	0.388
22 (para)	FU: 20 GP: 18	FU: 0.21 GP: 0.23	FU: 0.43 GP: 0.62	FU: -0.626 GP: 0.695	0.000***
23 (mp)	FU: 12 GP: 14	FU: 0.27 GP: 0.27	FU: 0.59 GP: 0.69	FU: -0.492 GP: 0.421	0.013*

Diskussion

Obwohl es sich bei Nichtsignifikanz keineswegs um ein „schlechteres Abschneiden“ handelt, da die Ergebnisse zufallsbedingt sein können, sollen die Klassen 14 und 21 hier dennoch diskutiert werden: Die Lehrerin der Klasse 14 hat das Thema Seifen bereits 2004 unterrichtet, verfügte somit über mehr Erfahrung. Ein Vergleich der Videos von 2004 und 2005 ergab, dass sie zweimal auf exakt die gleiche Art unterrichtet hat. Das bessere Abschneiden der Kontrollgruppe im Jahr 2005 muss folglich andere Gründe haben:

In Betracht kommen erstens eine sehr unvollständige Randomisierung nach dem Matched-Pairs-Prinzip, die in dieser Klasse durch äußere Umstände notwendig war: Der Franzö-

sisch-Kurs der Klasse 10 war zu Beginn der Untersuchung im Rahmen eines Schüleraustauschs in Frankreich, so dass diese Schüler aufgrund des zeitlichen Ablaufs der Untersuchung automatisch der GP-Gruppe zugeteilt wurden. Von einer zufälligen Auswahl kann somit nicht mehr die Rede sein, da davon acht Schüler betroffen waren.

Des Weiteren war die Motivation in der GP-Gruppe nicht so gut. Das mag zum einen an der Kursfahrt direkt vor den Osterferien gelegen haben, so dass die m. E. wichtige Motivationsphase, in der die Untersuchung und ihr Sinn vorgestellt wurden, nicht erfolgen konnte. Erschwerend kam hinzu, dass der Abschlusstest am letzten Schultag der Abiturienten geschrieben wurde, die wie gewöhnlich die Klassenräume „stürmten“ und dadurch für Unruhe sorgten. Die Erwartung der Feier direkt im Anschluss an den Test mag die Konzentration der Schüler zusätzlich geschwächt haben (siehe auch Abschnitt 8.5).

Auch bei der zweiten „Abweichler-Gruppe“ (Berufskolleg) spielte das Matched-Pairs-Prinzip eine entscheidende Rolle: Aufgrund von Krankheiten und Organisationsproblemen hat ein Lehrer die Gruppenzusammensetzung ändern müssen. Zudem konnte er nicht die gesamte Zeit während des Gruppenpuzzles anwesend sein, so dass die Schüler ohne Aufsicht waren. Außerdem fiel die Expertenphase komplett weg.

Fazit: Das Matched-Pairs-Prinzip scheint eine entscheidende Rolle für die saubere Durchführung einer Vergleichsuntersuchung zu sein. Weiterhin scheint die Motivationsphase in der ersten Stunde vor Beginn der Untersuchung eine entscheidende Rolle zu spielen. Dabei wurden der zu diesem Zeitpunkt noch kompletten Klasse der Sinn und die Ernsthaftigkeit der Untersuchung dargelegt und die Schüler aufgefordert, sich anzustrengen.

Nachgegangen werden könnte der Frage, welche Rolle die Zusammenfassung am Ende der Stunde in Form von Begriffen, zu denen jeder Schüler blitzlichtartig einen Satz sagen musste, spielt. Die Lehrerin der Lerngruppe 14 führte diese Form der Sicherung im FU durch. Allerdings konnte sie mit dieser Methode im Jahr zuvor keine besseren Ergebnisse erzielen, so dass hierzu weiterer Forschungsbedarf besteht.

8.1.2 Abhängigkeit der Lernleistung von der Art der Wissensvermittlung

Unabhängig davon, welches Themengebiet man betrachtet, erreichen die Experten für das jeweilige Thema deutlich höhere Wissenszuwächse als Schüler, denen ihre Mitschüler

(Experten) oder der Lehrer das Thema vermittelt haben (siehe Abbildung 18). Im Fall der ersten Hauptuntersuchung beträgt der Unterschied fast eine ganze Standardabweichung. Alle Unterschiede sind bis auf Thema 3 in der Voruntersuchung mindestens hoch signifikant (siehe Tabelle 9). Damit konnten die Ergebnisse der ersten HU bestätigt werden.

Selbst-Beigebrachtes und Fremdvermitteltes im Vergleich

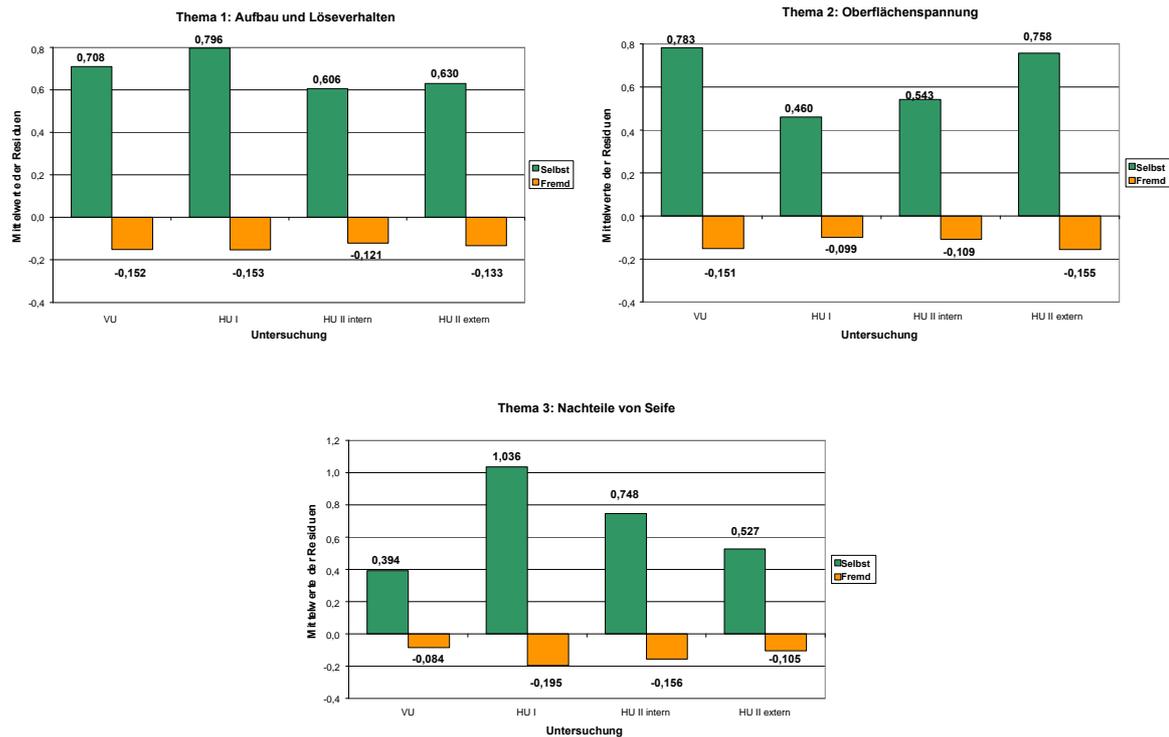


Abbildung 18: Selbst- und fremd vermitteltes Wissen im Vergleich: Ergebnisse der Vor- und Hauptuntersuchungen. Die Residuen wurden für jedes Thema und jede Untersuchung separat berechnet. Die Unterschiede in den Beträgen der Residuen sind auf starke Stichprobeunterschiede zurückzuführen.

Tabelle 9: Selbst-Beigebrachtes vs. Fremdvermitteltes: Ergebnisse des U-Tests nach Mann und Whitney. Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Signifikanzen sind mit entsprechenden Sternchen markiert.

	Thema	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N		68	158	180	235
U-Test $p_{(z)}$ (2-seitig)	1: Aufbau und Löseverhalten	0.003**	0.000***	0.000***	0.000***
	2: Oberflächenspannung	0.003**	0.006**	0.000***	0.000***
	3: Nachteile von Seife	0.120	0.000***	0.000***	0.000***

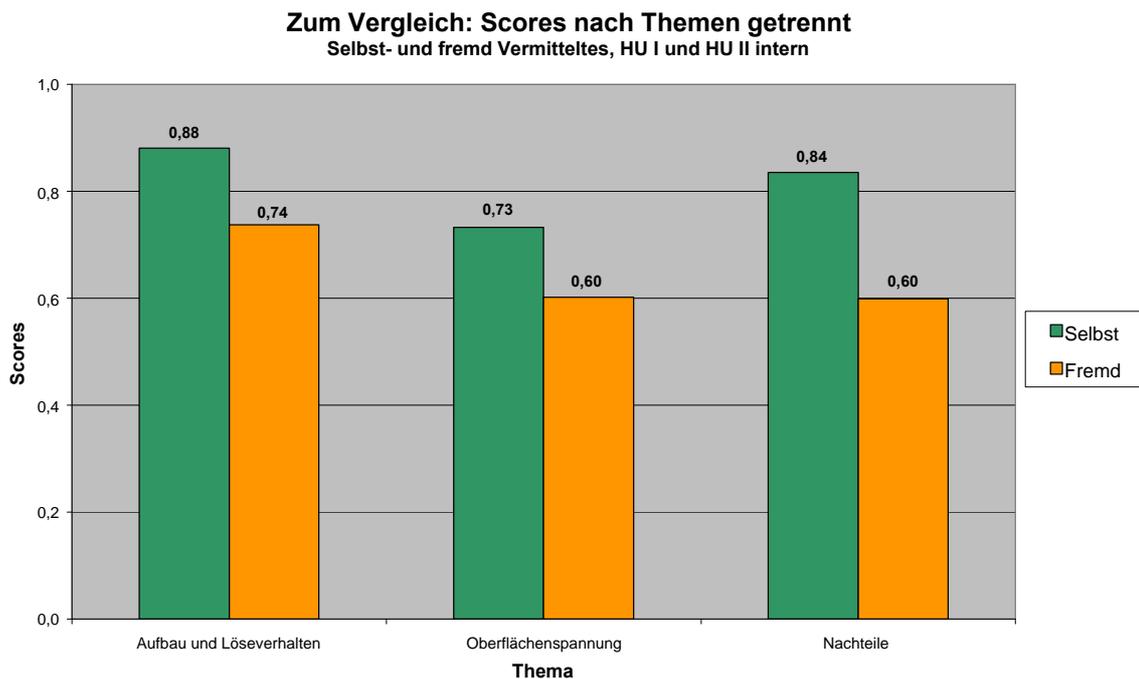


Abbildung 19: Selbst- und fremd vermitteltes Wissen im Vergleich – Scores: Ergebnisse des Post-Tests für die Hauptuntersuchungen (intern).

Unabhängig vom Thema waren Schüler, die Experten für das jeweilige Thema darstellten, erfolgreicher als Schüler, welche sich die Themen nicht selbst angeeignet haben (siehe Abbildung 19). Mit durchschnittlich knapp 82 % korrekter Antworten im Post-Test liegen sie ca. 17 %-Punkte über den „Zuhörern“ mit knapp 65 % richtiger Antworten. Dabei variiert der Vorsprung zwischen 13 und 24 Prozentpunkten.

Im Folgenden ist die Fremdvermittlung nochmals aufgeteilt in Lehrer (FU) bzw. Schüler (GP) dargestellt (siehe Abbildung 20). Dabei entsprechen die Ergebnisse im Wesentlichen denen von Abbildung 18: Die Experten sind auf ihrem Gebiet deutlich erfolgreicher als auf anderen Gebieten. Allerdings fällt der Vergleich zwischen Mitschülern und Lehrern weniger eindeutig aus.

Vergleich der drei Vermittlungsarten

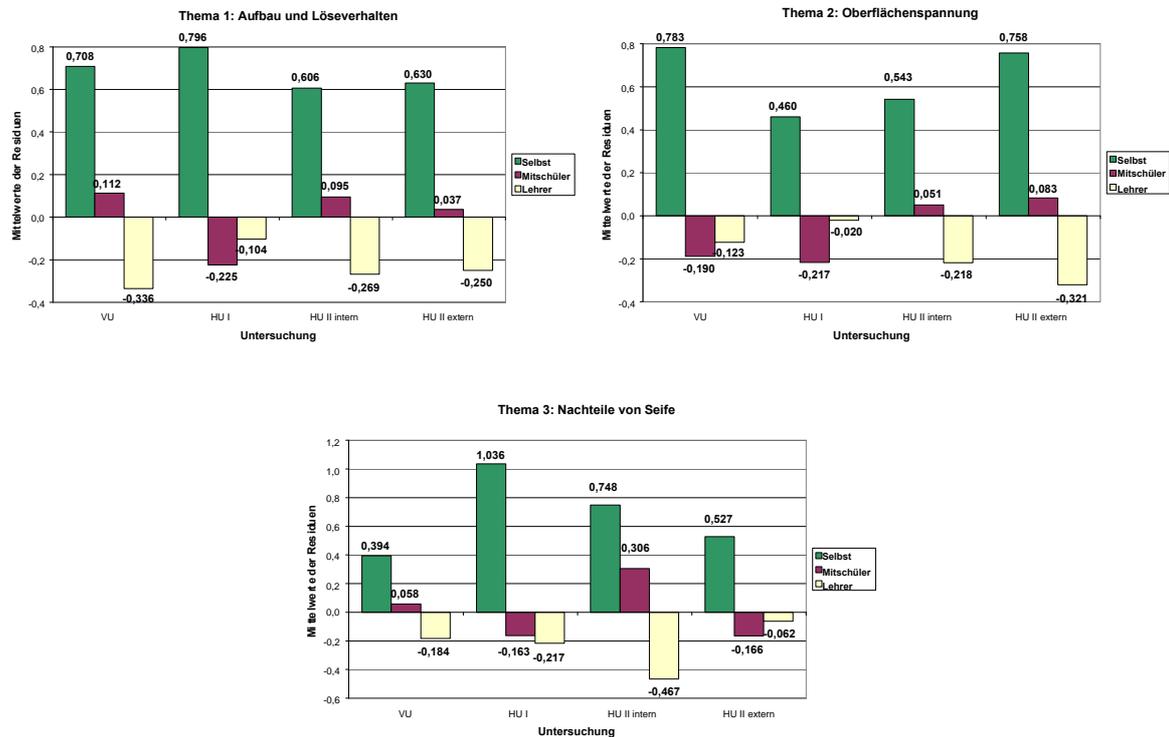


Abbildung 20: Vergleich der drei Arten der Wissensvermittlung – Residuen: Ergebnisse der Vor- und Hauptuntersuchungen. Die Residuen wurden für jedes Thema und jede Untersuchung separat berechnet.

Thema 1

In drei von vier Fällen ergeben sich Vorteile für das in den Kleingruppen vermittelte Wissen. Lediglich in der ersten Hauptuntersuchung schneidet der Frontalunterricht – wenn auch nicht signifikant – besser ab.

Thema 2

Für die VU und die HU I ergeben sich leichte Vorteile für das vom Lehrer vermittelte Wissen. Die Unterschiede in der HU II_{extern}, wo die Vermittlung durch Schüler erfolgreicher ist, sind signifikant (siehe Tabelle 10).

Thema 3

Die frontal unterrichteten Schüler schneiden nur in der externen HU II besser ab als ihre Gruppenpuzzle-Kollegen. Signifikante Unterschiede treten allerdings in der internen HU II auf, zugunsten der Schüler-Vermittlung.

Tabelle 10: Fremdvermittlung – Schüler vs. Lehrer: Ergebnisse des U-Tests nach Mann und Whitney. Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Signifikanzen sind mit entsprechenden Sternchen markiert.

	Thema	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N		56	130	150	194
U-Test $p_{(z)}$ (2-seitig)	1: Aufbau und Löseverhalten	0.220	0.756	0.018	0.041
	2: Oberflächenspannung	0.559	0.525	0.118	0.002*
	3: Nachteile von Seife	0.275	0.929	0.000***	0.450

Tabelle 10 belegt in lediglich zwei von 12 Fällen einen Unterschied in der Fremdvermittlung. Treten signifikante Unterschiede auf, so scheinen die Mitschüler einen größeren Wissenszuwachs bei den Probanden hervorzurufen als die Lehrer (siehe Abbildung 20).

Tabelle 11: Art der Wissensvermittlung: Ergebnisse des Jonckheere-Terpstra-Trendtests. Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Signifikanzen sind mit entsprechenden Sternchen markiert.

	Thema	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N		68	158	180	233
$p_{(z)}$, (2-seitig)	1: Aufbau und Löseverhalten	0.003**	0.001***	0.000***	0.000***
	2: Oberflächenspannung	0.093	0.125	0.000***	0.000***
	3: Nachteile	0.060	0.000***	0.000***	0.030*

Der Trendtest bestätigt die Uneinheitlichkeit der Unterschiede, liefert aber keine zusätzlichen Informationen.

Zum besseren Verständnis erfolgt ergänzend in Abbildung 21 eine Aufstellung der Scores:

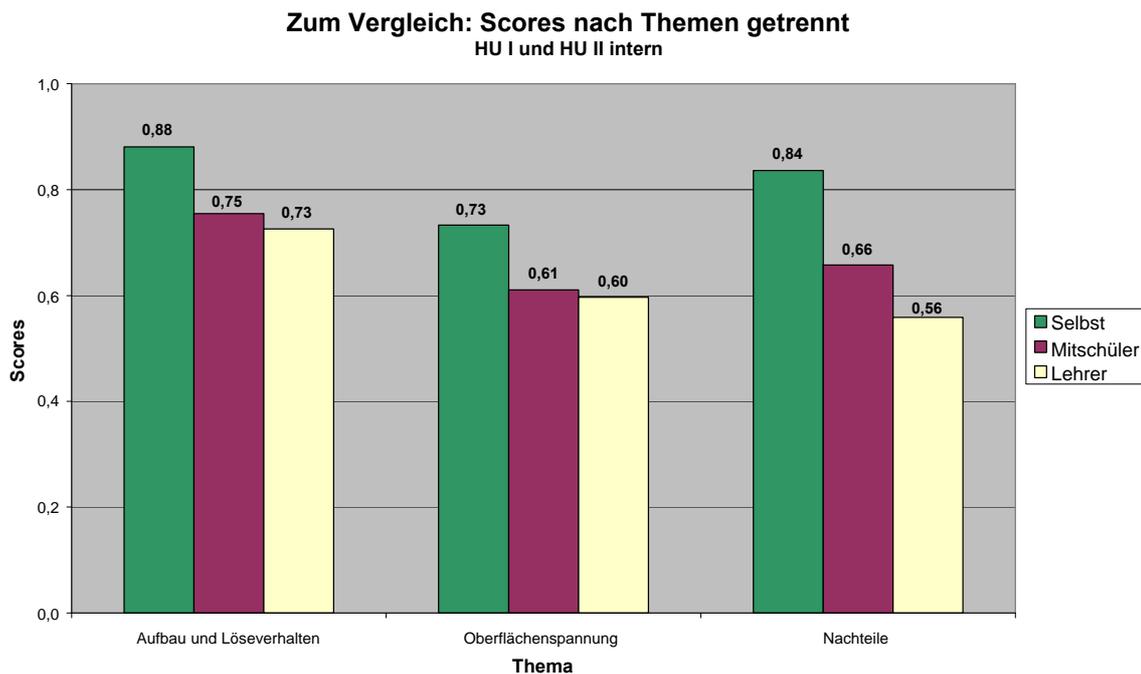


Abbildung 21: Vergleich der drei Arten der Wissensvermittlung – Scores: Ergebnisse der Vor- und Hauptuntersuchungen.

Werden die HU I und ihre Replikation (HU II_{intern}) im Jahre 2005 als Ganzes betrachtet, so ergibt sich folgendes Bild für die erreichten Scores im Nachtest: In allen drei GP-Themen liegen die Experten mit 73-88 % zutreffender Antworten vorne. Im Mittelfeld befinden sich die GP-Nicht-Experten (61-75 %). Am wenigsten scheinen die Frontalunterrichts-Schüler zu profitieren, da dort nur 56-73 % des Tests richtig gelöst werden konnte.

Diskussion

Die Analyse der Vermittlungsarten zeigt die hohe Überlegenheit von selbst beigebrachtem Wissen gegenüber rezeptiv erworbenem Wissen. Diese Erkenntnis deckt sich mit Untersuchungen von BERGER & HÄNZE (2004, 2005). Zwar sind diese nicht unbedingt mit der vorliegenden Studie vergleichbar, z. B. war die Stichprobe (2004) sehr viel kleiner, vor dem Gruppenpuzzle fand eine Frontalunterrichtssequenz statt, eine Auswertung über Residuen wurde nicht durchgeführt und beide Gruppen wichen in ihren Vorleistungen stark voneinander ab (BERGER & HÄNZE 2004, S. 216), jedoch zeigen die Ergebnisse die gleiche Tendenz. Auch BORSCH et al. (2002, S. 172) bestätigen den Experten-Vorteil für den Sachunterricht. Die Zuhörer der Kontrollgruppe hingegen schnitten in ihrer Studie besser ab als die Zuhörer bzw. „Novizen“ des Gruppenpuzzles, die Ergebnisse der vorliegenden Untersu-

chung sind diesbezüglich zumindest nicht einheitlich, tendieren aber zu einem Vorteil für die Schüler.

Beim Thema 2 (Oberflächenspannung) schnitten alle Schüler geringfügig schlechter ab. Zwar sollten die Themen alle gleich anspruchsvoll sein, jedoch mag der Anteil an Wiederholung (Wasserstoff-Brücken) dazu geführt haben, dass im Vortest einige Schüler durch ihr Vorwissen den Effekt überlagerten.

Fazit

Schüler, die sich ihr Thema selbst angeeignet haben, schnitten beim Wissenstest signifikant besser ab als Schüler, denen das Wissen lediglich durch ihre Mitschüler bzw. den Lehrer vermittelt wurde.

Die Unterschiede zwischen den beiden Fremdvermittlungsarten fallen hingegen deutlich geringer aus und sind nur in 2 von 12 Fällen signifikant, wie der U-Test nach Mann und Whitney ergab (siehe Tabelle 10). Da die Ergebnisse sehr uneinheitlich sind und nicht repliziert werden konnten, muss das Fazit diesbezüglich sehr vorsichtig formuliert werden: In einigen Fällen ist das Erklären durch Mitschüler effektiver als durch den Lehrer, zumindest aber gleich effektiv. Somit konnte die Annahme, dass die Mitschüler effektiver vermitteln, nicht eindeutig bewiesen werden.

Erklärungsversuche /-ansätze

Zwei Faktoren dürften für das bessere Abschneiden der Experten und damit des gesamten Gruppenpuzzles verantwortlich sein: Punkt 1 ist der höhere Anteil an Selbsttätigkeit durch das eigenständige Aneignen von Wissen. Dieses der Arbeitsschulbewegung entstammende Prinzip geht auf seine wohl wichtigsten deutschen Reformpädagogen, den Leibziger Schulleiter HUGO GAUDIG und den Münchener Stadtschulrat und Professor GEORG KERSCHENSTEINER zurück (vgl. GUDJONS 1995, S. 103). Obwohl Kontrahenten, sahen beide die selbstständige geistige Arbeit der Schüler als wichtigste Voraussetzung für das Lernen. Allerdings mit unterschiedlichen Zielen: Während KERSCHENSTEINER die staatsbürgerliche Erziehung im Vordergrund sah, betonte GAUDIG die Bildung zur „Persönlichkeit“ (GUDJONS, ebd.). Er entwickelte eine Reihe praktischer Methoden, die das eigenständige Erarbeiten von Themen durch die Schüler fördern sollten, gemäß DEWEYS „learning by doing“. Diese aktive Auseinandersetzung mit dem Lerngegenstand ist Teil der konstruktivistischen Lerntheorie (siehe Kapitel 2.4).

Allerdings darf die Einzelleistung nicht für sich betrachtet werden: Erst wenn neben der Einzelleistung auch die Möglichkeit gegeben ist, die Ergebnisse abzusichern, wie es beim Gruppenpuzzle in der Expertenrunde der Fall ist, kann der Vorteil entstehen. Darauf deuten die Ergebnisse einer Metastudie von JOHNSON & JOHNSON (1994) hin. In dieser wurden 59 Einzelstudien auf die Effektivität von kooperativem Lernen und individuellem Lernen untersucht. SLAVIN (1993, S. 170) nennt weiterhin die Bedeutung von Gruppenzielen und die individuelle Verantwortlichkeit.

Trotz dieser eindeutigen Ergebnisse und Erkenntnisse aus der Lehr-Lernforschung sind Methoden des selbstständigen Lernens auch zu Beginn des 21. Jahrhunderts besonders in den Sekundarstufen der allgemein bildenden Schulen nur wenig verbreitet (vgl. UNRUH/PETERSEN 2006, S. 55).

GREEN & GREEN (2005, S. 12) sehen die Vermittlungsleistung der Experten (Punkt 2) für das bessere Abschneiden verantwortlich: „[Schüler] *begreifen 95 % von dem, was sie gelernt haben, erst dann, wenn sie es anderen vermitteln*“. Das Gruppenpuzzle beinhaltet diese als Lernen durch Lehren bekannte Variante kooperativen Lernens. Sowohl RENKL (2000, zitiert in MEYER 2004, S. 83) als auch HUBER, KONRAD & WAHL 2001 diskutieren und bestätigen die Vorteile dieser Methode, die bereits auf COMENIUS (1592-1670) zurückgeht:

„Sehr wahr ist nämlich das Wort: `Wer andere lehrt, der bildet sich selbst`; und zwar nicht bloß, weil er durch Wiederholung das aufgenommene Wissen in sich befestigt, sondern weil er Gelegenheit findet, tiefer in die Dinge einzudringen. Der geistreiche Joachim Fortius bezeugt, was er jemals bloß gehört oder gelesen habe, das sei ihm schon innerhalb eines Monats wieder entfallen, was er aber andere gelehrt habe, das sei ihm so vertraut wie die Zahl seiner Finger, und er glaube, es könne ihm nur durch den Tod entrissen werden.“²⁵

Zu begründen ist dies mit neueren Forschungen aus der Psychologie (vgl. BENNETT et al. 2002, S. 7): Dem konstruktivistischen Ansatz folgend werden beim Lernen neue Informationen mit alten verknüpft. Dazu ist eine Art kognitive Umstrukturierung – auch Elaboration genannt – des Materials erforderlich (vgl. WITTRÖCK 1986). Eine der effektivsten Ela-

²⁵ COMENIUS, J. A. (1954): Große Didaktik. In neuer Übers. hrsg. von ANDREAS FLITNER. Düsseldorf, Kupper, S. 117.

borationsmöglichkeiten ist es, jemand anderem das Material zu erklären (vgl. CHI, DE LEEUW, CHIU & LAVANCHER 1994; DANSEREAU 1998, zitiert in BENNETT et al. 2002, S. 7) sowie die Interaktion mit Gleichaltrigen über Kontexte, die per se eine Erklärung von beiden Seiten notwendig machen.

RENKL (1997) erklärt Effekte des Lernen durch Lehren mit Hilfe dreier Faktoren: Die Lehrerwartung führe zu besserem Verständnis des Lernstoffs sowie einer besseren Lernleistung. Durch das Geben von Erklärungen können Wissenslücken oder Verständnisprobleme auf Seiten der Experten aufgedeckt werden. Zuletzt hilft das Reagieren auf Rückfragen dabei, Sachverhalte noch einmal zu durchdenken und Zusammenhänge herzustellen (vgl. hierzu auch FÜRSTENAU 2002, S. 42).

8.1.3 Wechselwirkungen zwischen Unterrichtsmethode und Ausgangsleistung

Die Residuen von GP- und FU-Gruppe fallen mit der Zunahme der Chemienote nahezu linear ab (siehe Abbildung 22). Diese Korrelation ist signifikant (siehe Tabelle 13). Auffallend ist, dass sich der Abstand der Residuen von FU und GP-Schülern kaum verändert: Erreichen mit der Zensur eins vorbenotete Schüler durch das Gruppenpuzzle 0.61 Standardabweichungen mehr als vergleichbare FU-Schüler, so wird dieser Abstand mit 0.53 (Note 4) bzw. 0.44 (Note 5) fast aufrecht erhalten (relativer Lernzuwachs). Auch mit 5 vorbenotete Schüler schneiden im Vergleich immer noch besser ab als frontal unterrichtete Schüler. Interessant ist, dass im Vorfeld als ausreichend beurteilte Schüler durch das Gruppenpuzzle ihre Leistungen fast auf Einser-Niveau der FU-Schüler anheben konnten.

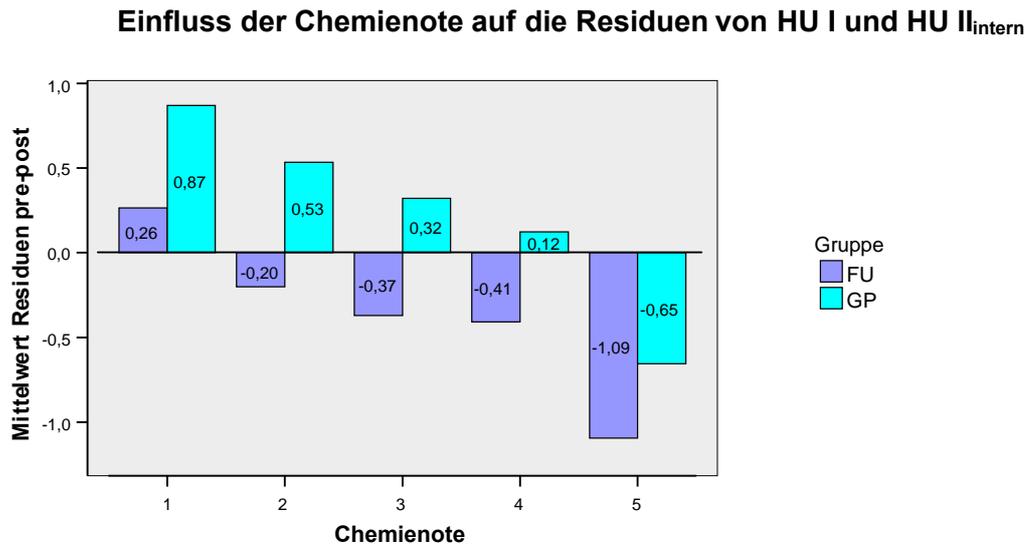


Abbildung 22: Einfluss der Chemie-Zeugnisnote auf den Lernzuwachs. Die Residuen (pre-post) wurden für beide Gruppen zusammen berechnet. Dargestellt sind HU I und HU II_{intern}.

Tabelle 12: Einfluss der Chemienote: Ergebnisse der zweifaktoriellen Varianzanalyse (2-Weg-Wechselwirkungen). Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Werte für Haupteffekte unter 0.01 gelten als signifikant (Sternchen), Wechselwirkungen können ab 0.2 ausgeschlossen werden. Für die Wechselwirkungen sind die Effektstärken η^2 mit angegeben. Bei der Analyse der Haupteffekte wurden die Residuen für beide Gruppen zusammen auf Unterschiedlichkeit getestet.

		VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N		68	158	180	233
Haupteffekte	Methode	0.010	0.003*	0.000**	0.000**
	Chemienote	0.560	0.032	0.000**	0.025
2-Weg-Wechselwirkungen	Methode x Chemienote	0.209	0.752	0.399	0.776
Effektstärke η^2		0.095	0.013	0.023	0.008

Mit Ausnahme der HU II_{intern} unterscheiden sich die Residuen der unterschiedlich vorbereiteten Schüler in keiner Stichprobe signifikant (siehe Haupteffekte, FU und GP gepoolt). Betrachtet man die 2-Weg-Wechselwirkungen von Gruppe und Chemienote, so sind keinerlei Wechselwirkungen nachweisbar.

Diskussion

Der höhere Lernzuwachs der GP-Schüler im Vergleich zu den FU-Schülern scheint unabhängig von ihrer Zeugnisnote zu sein. Gute wie schlechte Schüler profitieren also gleichermaßen vom Gruppenpuzzle.

Zwar sprechen die Irrtumswahrscheinlichkeiten dafür, dass sich die Residuen der unterschiedlich vorbenoteten Schüler nicht signifikant unterscheiden (siehe Haupteffekte Tabelle 12). Die Werte nahe der Signifikanzgrenze jedoch lassen nicht den Umkehrschluss zu, dass sie völlig gleich sind (dann müsste $p \geq 0.2$ sein). Im Prinzip erreichen „gute“ Schüler auch höhere Residuen, wie die Korrelationen in Tabelle 13 zeigen.

Die Vermutung, das Gruppenpuzzle helfe vermehrt guten Schülern (vgl. z. B. DE BAZ 2001), konnte widerlegt werden. Zwar nimmt der Methodenvorteil für schlechtere Schüler etwas ab, jedoch nicht signifikant, wie die Varianzanalyse ergab. Gestützt wird dieser Umstand durch die Tatsache, dass die Residuen-Differenzen (FU – GP) zwischen guten und schlechten Schülern fast gleich bleiben (siehe Abbildung 22). Die Methode wirkt auf alle Leistungstypen gleichermaßen. Damit konnte eine frühere Beobachtung von LAZAROWITZ (1991, S. 20) bestätigt werden: Als eher schwach beurteilte Schüler erreichen durch das kooperative Lernen ähnlich hohe Lernerfolge wie gute, allerdings herkömmlich unterrichtete Schüler.

Der dennoch leicht vorhandene Abfall der mit fünf vorbenoteten Schüler kann m. E. mit Motivations- und Vorwissensproblemen begründet werden. So mögen sich einerseits nicht alle Schüler gleichermaßen engagiert mit dem Text beschäftigt haben, andererseits mögen auch grundlegende chemische Kenntnisse, die nicht per Wissenstest erfasst wurden, bei schlechten Schülern gefehlt haben. Auch Probleme, den Text zu verstehen, erscheinen plausibel. Wie wenig relevant diese Suche nach Begründungen allerdings ist, zeigt die geringe Zahl der „5er“ Schüler: Insgesamt wurden für die HU I und HU II_{intern} lediglich zehn Schüler ($\approx 3\%$) mit 5 vorbenotet. Die Ergebnisse konnten repliziert werden.

Tabelle 13: Korrelation von Chemienote und Residuen für beide Gruppen zusammen. Ein Sternchen: Die Korrelation ist auf dem 0.05 Niveau signifikant. Zwei Sternchen: Die Korrelation ist auf dem 0.01 Niveau signifikant.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	68	158	180	233
Korrelationskoeffizient (Spearman-Rho)	-0.151	-0.222**	-0.246**	-0.168*
Signifikanz (2-seitig)	0.219	0.005	0.001	0.010

8.1.4 Abhängigkeit des Lernerfolgs vom Geschlecht

Für das Gruppenpuzzle (siehe Abbildung 23, jeweils rechts dargestellt) ergibt sich ein uneinheitliches Bild: Einmal sind die Jungen besser (HU I und HU II extern), ein anderes Mal die Mädchen (VU, HU II_{intern}). Grundsätzlich scheinen im Frontalunterricht Jungen etwas besser zu sein als Mädchen (siehe Abbildung 23).

Die Varianzanalyse zeigt jedoch, dass eine Wechselwirkung zwischen Geschlecht und Methode nicht besteht (siehe Tabelle 14). Wechselwirkungen würden dazu führen, dass sich die Geraden stärker überschneiden. Nebenbei bestätigt die Varianzanalyse die Ergebnisse des U-Tests, was die Wirksamkeit der Unterrichtsmethode betrifft.

Wechselwirkungen von Geschlecht und Unterrichtsmethode

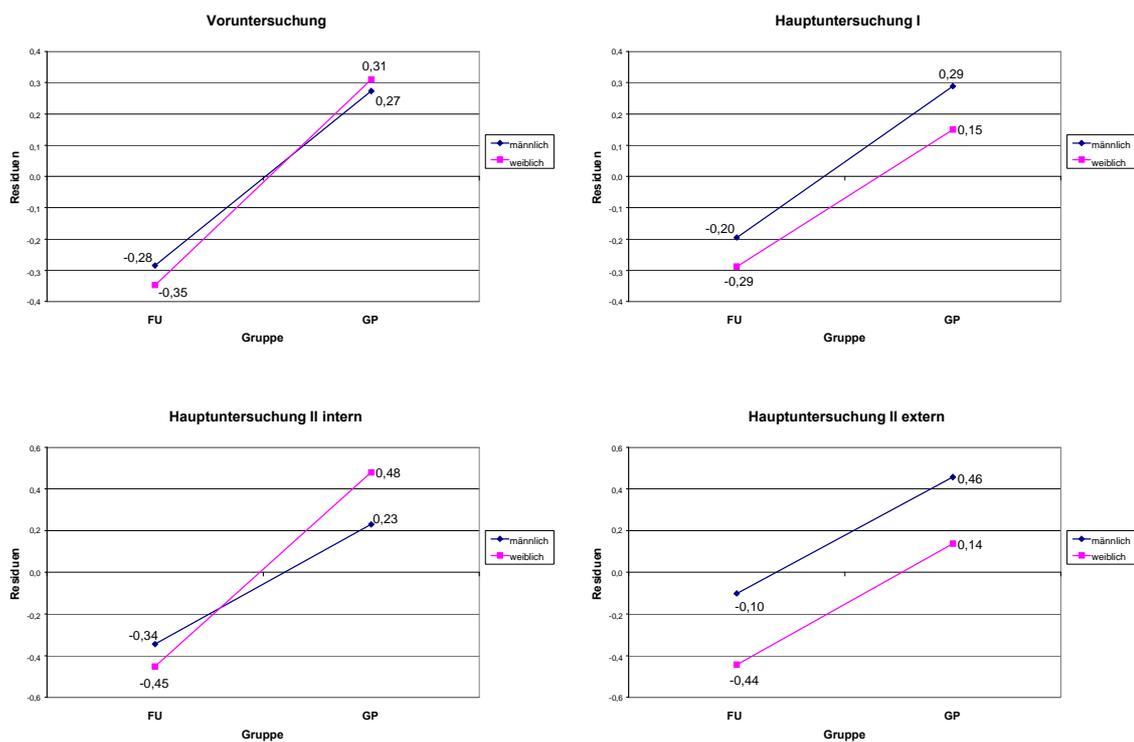


Abbildung 23: Wechselwirkungen von Geschlecht und Unterrichtsmethode. Die Residuen (pre-post) wurden für beide Gruppen zusammen berechnet.

Tabelle 14: Einfluss des Geschlechts: Ergebnisse der zweifaktoriellen Varianzanalyse. Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Werte für Haupteffekte unter 0.01 gelten als signifikant (Sternchen), Wechselwirkungen können ab 0.2 ausgeschlossen werden. Für letztere sind die Effektstärken η^2 mit angegeben.

		VU	HU I	HU II intern	HU II extern	HU II extern Gym.	HU II ext. andere Schulform.
Stichprobenum- fang N		68	158	180	235	47	188
Haupteffekte	Methode	0.012	0.003	0.000	0.000	0.000	0.000
	Geschlecht	0.958	0.486	0.629	0.009*	0.387	0.031
2-Weg- Wechselwirkungen	Methode x Geschlecht	0.834	0.847	0.203	0.927	0.302	0.836
Effektstärke η^2		0.001	0.000	0.009	0.000	0.025	0.000

Geschlechtsunterschiede (Haupteffekte) sind lediglich für die Nicht-Gymnasialklassen (HU II_{extern}, andere Schulformen) zu vermuten. Zwar treten signifikante Haupteffekte in der HU II_{extern} auf, diese verschwinden jedoch, sobald man sie in Gymnasial- und Nicht-Gymnasialklassen splittet.

Diskussion

Tatsächlich sind weder in der Kontroll- noch in der Interventionsgruppe geschlechtsspezifische Unterschiede bezüglich des Lernzuwachses nachzuweisen (siehe Tabelle 14, 2-Weg-Wechselwirkungen). Obwohl in zwei Fällen das Ergebnis durch das Gruppenpuzzle umgekehrt werden konnte, sind diese Effekte nicht auf die Unterrichtsmethode zurückzuführen, wie die zweifaktorielle Varianzanalyse ergab: Die Methode wirkt auf beide Geschlechter gleich gut. LAZAROWITZ (1991, S. 20) konnte zwar leichte Vorteile für weibliche Schüler diagnostizieren, jedoch waren die Unterschiede nicht signifikant. Eine umfangreiche Untersuchung im Fach Biologie zum Thema Fotosynthese in neunzehn 10. Klassen ($n = 708$), bestätigte die Unabhängigkeit des Gruppenpuzzles vom Geschlecht (vgl. LAZAROWITZ & KARSENTY 1990, S. 142).

Die nur aufgrund des abgesenkten Signifikanzniveaus von 0.01 nicht mehr signifikanten Unterschiede bei den Geschlechtern ($p = 0.031$) treten lediglich bei Nicht-Gymnasialklassen auf (siehe Haupteffekte Tabelle 14). Die Gründe dafür liegen wahrscheinlich bei der Klassenzusammensetzung: In den beteiligten Berufsschulklassen waren Jungen eher unterrepräsentiert. Eventuell konnten sie durch das Gruppenpuzzle aus dem

Schatten der Mädchen heraustreten, was anhand dieser Daten allerdings nicht belegt werden kann. Ferner wurde in der externen Hauptuntersuchung nicht immer das Matched-Pairs-Prinzip eingehalten, so dass sich Aussagen – wenn überhaupt – nur mit eingeschränkter Gültigkeit machen lassen. Der nächste Schritt wäre die klassenweise Analyse, welche sich aber aus Gründen der Vergleichbarkeit der zugrunde liegenden Fragestellung und der statistischen Aussagekraft verbietet. Für eine Aussage bezüglich der Geschlechtervorteile an anderen Schulformen müsste eine eigens konstruierte Untersuchung mit eigenen Hypothesen durchgeführt werden.

8.1.5 Abhängigkeit des Lernerfolgs von der Schulform

Unabhängig von der Schulform führt das Gruppenpuzzle zu einem höheren Lernzuwachs als der Frontalunterricht (siehe Abbildung 24). Mit 1.23 Standardabweichungen liegen die Resultate der Gymnasiasten über denen der Realschüler (0.55), Berufsschüler (0.07) sowie der Hauptschüler (-0.38) – jeweils die GP-Gruppe betrachtend. Die Unterschiede zwischen den Schulformen sind – zusammen mit der FU-Gruppe – signifikant (siehe Tabelle 15). Beim Vergleich der beiden Unterrichtsmethoden innerhalb einer Schulform fallen die Differenzen der Lernzuwächse (GP – FU) stark unterschiedlich aus: GP-Realschüler erreichen 1.14 Standardabweichungen mehr als die vergleichbare FU-Gruppe, gefolgt von den Gymnasiasten mit 0.79, Berufskolleg-Schülern mit 0.33 und Hauptschülern mit 0.28.

Wechselwirkungen von Schulform und Unterrichtsmethode

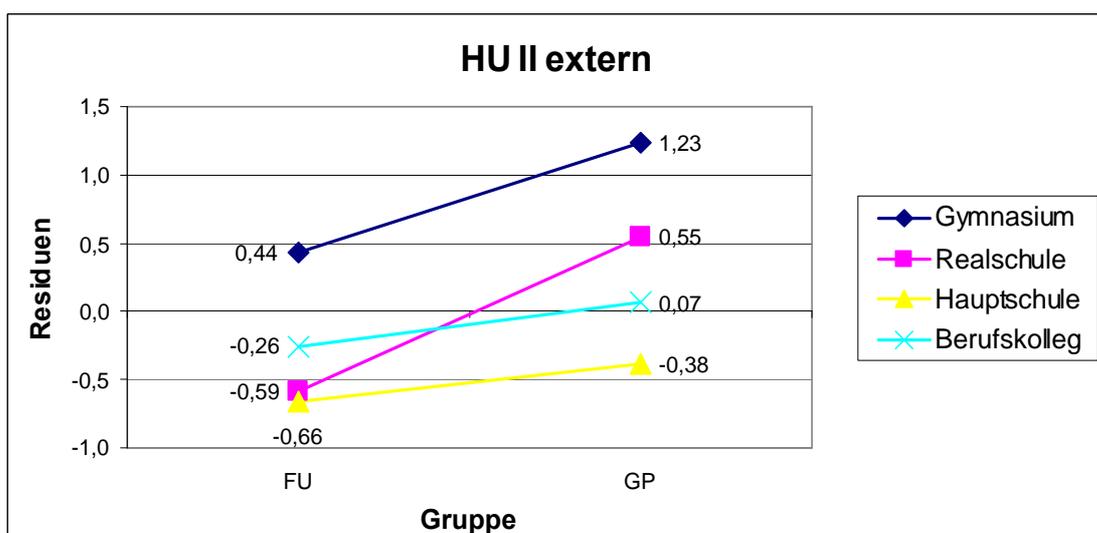


Abbildung 24: Wechselwirkungen von Schulform und Unterrichtsmethode. Die mittleren Residuen (pre-post) wurden für alle Schulformen zusammen berechnet.

Tabelle 15: Wechselwirkungen zwischen Methode, Geschlecht und Schulform für $HU II_{\text{extern}}$: Ergebnisse der dreifaktoriellen Varianzanalyse. Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Werte für Haupteffekte unter 0.01 gelten als signifikant (Sternchen), Wechselwirkungen können ab 0.2 ausgeschlossen werden. Zusätzlich sind die Effektstärken η^2 mit angegeben.

$HU II_{\text{extern}}$	Faktor	p	η^2
Stichprobenumfang $N = 235$			
Haupteffekte	Schulform	0.000***	0.238
2-Weg-Wechselwirkungen	Methode x Schulform	0.050	0.035
	Geschlecht x Schulform	0.695	0.007
3-Weg-Wechselwirkungen	Methode x Geschlecht x Schulform	0.843	0.004

Die Tabelle zeigt signifikante Unterschiede in den Schulformen (Haupteffekte). Bei den 2-Weg-Wechselwirkungen konnte keine signifikante Abhängigkeit von Schulform und Methode nachgewiesen werden. Eine zusätzliche Beeinflussung durch das Geschlecht (2- und 3-Weg-Wechselwirkung) existiert nicht.

Diskussion

Das Gruppenpuzzle scheint auf verschiedene Schulformen ähnlich stark zu wirken. Allerdings existieren Vorteile des Gruppenpuzzles vor allem beim Gymnasium und der Realschule. Dieser Verdacht wird durch die mehrfaktorielle ANOVA untermauert (siehe Tabelle 15). Zwar liegt die Irrtumswahrscheinlichkeit p mit 0.050 relativ knapp über der Signifikanzgrenze von 0.01, da jedoch H_0 Wunschhypothese ist, kann eine Wechselwirkung von Methode und Schulform nicht mehr ausgeschlossen werden. Genauere Rückschlüsse lassen sich aus dieser dafür nicht angelegten Untersuchung nicht ziehen. So konnte z. B. nicht in allen Fällen das Matched-Pairs-Prinzip angewendet werden, eine Vergleichbarkeit zwischen den verschiedenen Stichproben besteht demnach nicht unbedingt. Zudem ist zu beachten, dass das GP für das Gymnasium entwickelt wurde. Daher muss es bei einem Hinweis bleiben. Eine detailliertere (z. B. klassenweise) Analyse macht demzufolge ebenfalls keinen Sinn. Für Aussagen bezüglich der Schulform bedarf es einer speziellen Untersuchung.

8.1.6 Einfluss der Übungsdauer auf den Lernerfolg

Die grafische Auswertung (siehe Abbildung 25) gibt keinerlei Hinweise auf Zusammenhänge zwischen der Zeit, die für den Test geübt wurde, und dem erreichten Lernzuwachs.

Deutlich zu sehen ist allerdings das bessere Abschneiden der Interventionsgruppe – unabhängig von der aufgewendeten Übungszeit.

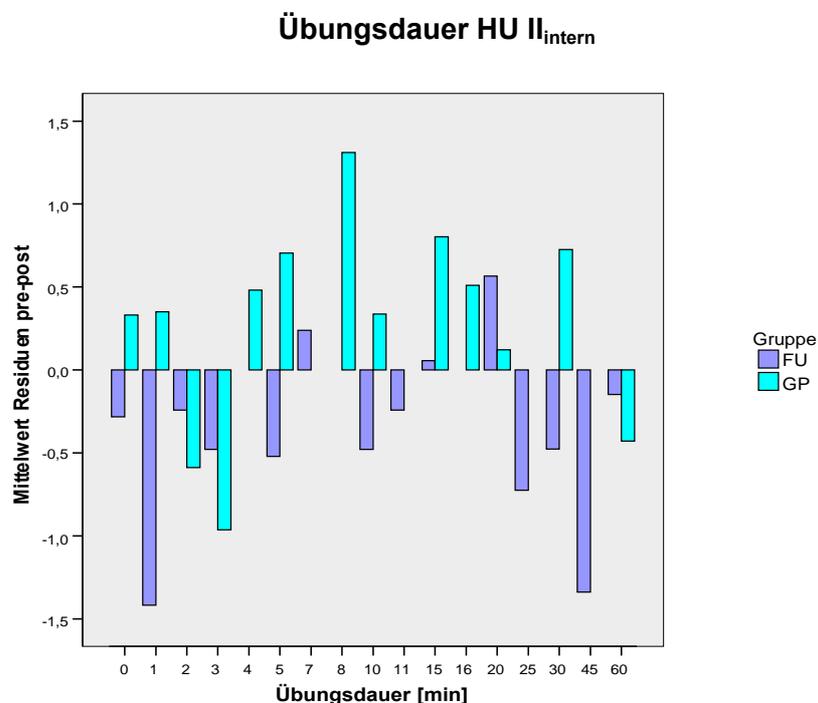


Abbildung 25: Abhängigkeit des Lernzuwachses von der Übungsdauer für die HU II_{intern} (ohne die ersten beiden Klassen, siehe 5.4.2). Die Residuen (pre-post) wurden für beide Gruppen zusammen berechnet.

Tabelle 16: Korrelation Übungsdauer-Lernzuwachs nach Gruppen getrennt. In die Korrelationsanalyse gingen die Residuen aus Vor- und Nachtest der beiden Gruppen sowie deren angegebene Übungsdauer. Sternchen kennzeichnen jeweils signifikante Korrelationen auf dem 5 %-Niveau. Die Fallzahlen weichen von den bisherigen ab, da die Übungsdauer erst ab der HU II_{intern} ohne die ersten beiden Klassen in Hagen erfasst wurde.

	HU II _{int.}	HU II _{ext.} Gymnasien	HU II _{ext.} andere Schulformen
Stichprobe N	FU: 60 GP: 62	FU: 23 GP: 24	FU: 84 GP: 94
Übungsdauer nach Gruppe [min]	FU: 10.6 GP: 10.2	FU: 11.1 GP: 10.8	FU: 20.5 GP: 37.5
Korrelationskoeffizient nach Spearman	FU: 0.001 GP: 0.039	FU: -0.102 GP: 0.096	FU: -0.219* GP: -0.174
Signifikanz (2-seitig)	FU: 0.992 GP: 0.764	FU: 0.644 GP: 0.655	FU: 0.046* GP: 0.093

Bei Gymnasiasten existieren kaum nennenswerte Unterschiede in den Übungszeiten der beiden Stichproben. Mit 10-11 Minuten liegen sie alle gleich auf. Anders sieht es bei den Haupt-, Real-, und Berufsschülern aus: Sie üben nicht nur deutlich mehr im Schnitt, die reinen Mittelwertvergleiche (FU: 20.5 min, GP: 37.5 min) lassen zudem auf starke Unterschiede in den beiden Gruppen schließen, was durch die Varianzanalyse untermauert wird (siehe Tabelle 17). Der Korrelationskoeffizient nach Spearman gibt an, wie Übungsdauer und Lernerfolg miteinander korrelieren: Nur bei frontal unterrichteten Nicht-Gymnasiasten findet sich eine signifikant negative Korrelation (zur Begründung: siehe Diskussion).

Tabelle 17: Abhängigkeit der Übungsdauern von der Methode: Ergebnisse der zweifaktoriellen Varianzanalyse. Die Residuen wurden aus Pre-Test und Post-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Werte für Haupteffekte unter 0.01 gelten als signifikant, Wechselwirkungen können ab 0.2 ausgeschlossen werden. Für letztere sind die Effektstärken η^2 mit angegeben. Fett dargestellt sind Auffälligkeiten. Die Fallzahlen weichen von den bisherigen ab, da die Übungsdauer erst ab der HU II_{intern} ohne die ersten beiden Klassen in Hagen erfasst wurde.

		HU II _{intern}	HU II _{extern} Gym.	HU II _{extern} andere Schulf.
Stichprobenumfang N		122	47	178
Haupteffekte	Methode	0.000	0.000	0.001
	Übungsdauer	0.943	0.951	0.027
2-Weg- Wechselwirkungen	Methode x Übungsdauer	0.842	0.268	0.316
Effektstärke η^2		0.049	0.179	0.081

Während bei den Gymnasiasten weder signifikante Unterschiede in den Übungsdauern noch in den Wechselwirkungen existieren, können bei Nicht-Gymnasiasten Übungszeitunterschiede (beide Gruppen zusammen) zumindest nicht mehr ausgeschlossen werden ($p = 0.027$). Diese sind jedoch in allen Fällen von der Methode unabhängig.

Diskussion

Da nur die anderen Schulformen auffällig geworden sind, sollen die Gymnasien nicht diskutiert werden.

Einerseits scheinen die angesprochenen GP-Schüler signifikant mehr als FU-Schüler zu üben (siehe Tabelle 16), für eine konkrete Aussage wären allerdings weitere Untersuchungen notwendig. Der zweite Befund – also die hohe negative Korrelation von Übungszeit und Lernerfolg – vermag zu dem Schluss leiten: Je mehr man übt, desto schlechter wird der Lernzuwachs. M. E. ist dieser scheinbare Widerspruch mit der Ausgangsleistung eini-

ger Schüler zu erklären: Schlechte Schüler üben per se mehr für Tests, auch wenn das nicht unbedingt zu einem höheren Lernerfolg führt, zumindest wenn man diesen in Relation zur Klasse sieht. Im Fall des Parallelklassendesigns wurde immer die etwas schlechtere Klasse zur Interventionsgruppe, was die höhere Übungsdauer der GP-Gruppe erklären mag.

Die generell höhere Lernzeit der Nicht-Gymnasiasten erscheint insofern plausibel, als dass sich einige Zehntklässler für das Abschlusszeugnis noch einmal richtig angestrengt haben mögen. In keinem Fall jedoch besteht eine Wechselwirkung zwischen Methode und Übungsdauer für den Test, zu sehen an den hohen Irrtumswahrscheinlichkeiten der 2-Weg-Wechselwirkungen (siehe Tabelle 17).

8.1.7 Nachhaltigkeit des erworbenen Wissens

Wie schon beim Post-Test schneidet die Interventionsgruppe durchweg besser beim Follow-up-Test ab als die Kontrollgruppe (siehe Abbildung 26). Dabei schwanken die Scores der GP-Gruppe von 0.5 bis 0.64, die der FU-Gruppe von 0.42-0.54. Die einstigen Unterschiede von 8 %-Punkten im Post-Test schrumpfen damit auf durchschnittlich 6.3 %-Punkte zusammen. Wiederum liegen die Werte der Voruntersuchung etwas höher, da die erste (leichtere) Testversion verwendet wurde.

Follow-up-Tests

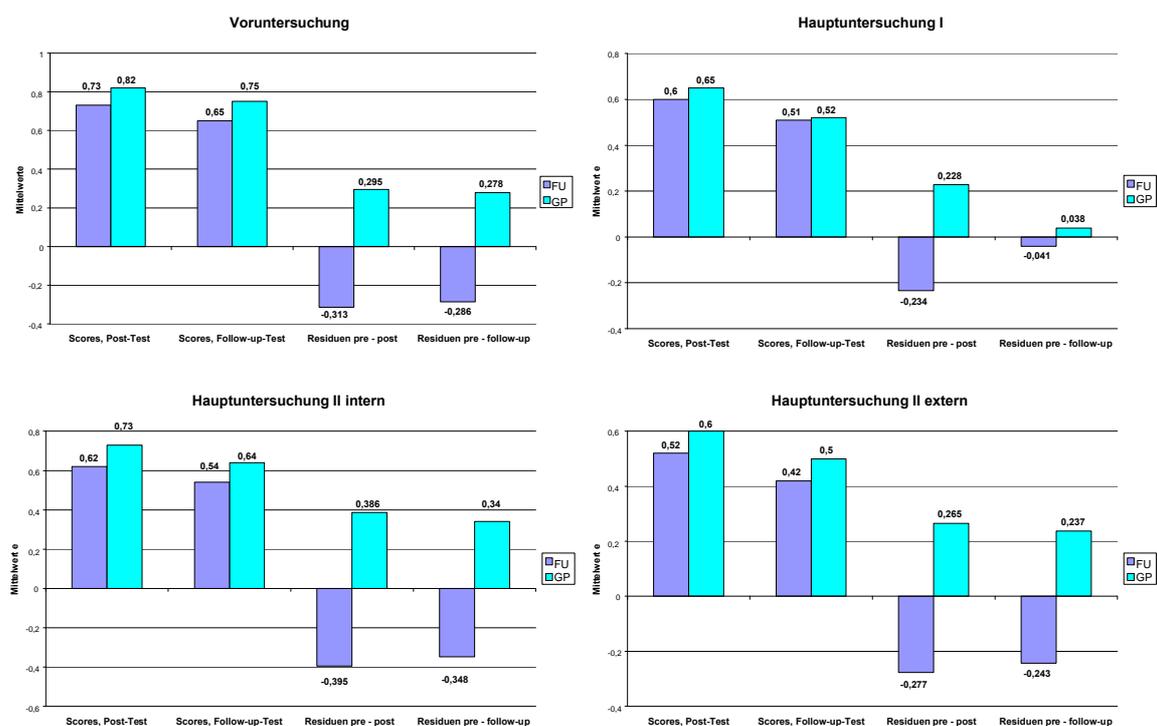


Abbildung 26: Ergebnisse der Vor- und Hauptuntersuchungen. Dargestellt sind Residuen sowie Scores.

Betrachtet man die Residuen des Follow-up-Tests, so ergibt sich das tendenziell gleiche Bild: Die kooperative Lernform liegt mit 0.239 Standardabweichungen (alle HU gemittelt) vor der FU-Gruppe mit -0.211 Standardabweichungen. Die Unterschiede sind bis auf die HU I signifikant (siehe Tabelle 18). Im Vergleich zu den Residuen des Post-Tests sind sie generell weniger stark ausgeprägt: Die Frontalunterrichtsgruppe erreicht etwas bessere Werte, die GP-Gruppe etwas schlechtere Werte. Es findet demnach eine Angleichung statt. Die Unterschiede zwischen Post- und Follow-up-Test sind in beiden Gruppen nicht signifikant, wie der Wilcoxon-Test ergeben hat (siehe Tabelle 18).

Tabelle 18: Nachhaltigkeit des erworbenen Wissens: Ergebnisse von Vor- (VU) und Hauptuntersuchung (HU). Cronbachs α wurde für den Follow-up-Test bestimmt, Stichprobengröße und Effektstärke beziehen sich jeweils auf die Residuen, welche aus Pre- und Follow-up-Test errechnet wurden.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 33 GP: 34	FU: 72 GP: 77	FU: 89 GP: 91	FU: 111 GP: 114
MW Scores, Pre-Test	FU: 0.38 GP: 0.41	FU: 0.26 GP: 0.25	FU: 0.30 GP: 0.28	FU: 0.25 GP: 0.25
MW Scores, Post-Test	FU: 0.73 GP: 0.82	FU: 0.60 GP: 0.65	FU: 0.62 GP: 0.73	FU: 0.52 GP: 0.60
MW Scores, Follow-up-Test	FU: 0.65 GP: 0.75	FU: 0.51 GP: 0.52	FU: 0.54 GP: 0.64	FU: 0.42 GP: 0.50
MW Residuen _{pre-post}	FU: -0.313 GP: 0.295	FU: -0.234 GP: 0.228	FU: -0.395 GP: 0.386	FU: -0.277 GP: 0.265
MW Residuen _{pre-follow-up}	FU: -0.286 GP: 0.278	FU: -0.041 GP: 0.038	FU: -0.348 GP: 0.340	FU: -0.243 GP: 0.237
U-Test Residuen _{pre-follow-up} , $p_{(z)}$, zweiseitig	0.022*	0.968	0.000***	0.000***
Wilcoxon-Test Residuen _{pre-post / pre-follow-up} , $p_{(z)}$, zweiseitig	FU: 0.728 GP: 0.726	FU: 0.563 GP: 0.228	FU: 0.723 GP: 0.282	FU: 0.493 GP: 0.509
Cronbachs α	0.79	0.82	0.81	0.80
Effektstärke d	0.57	0.08	0.69	0.48
Periode zwischen Post- und Follow-up-Test in Tagen (MW)	FU: 56 GP: 54	FU: 48 GP: 42	FU: 40 GP: 39	FU: 44 GP: 46

Folgende Tendenz ist erkennbar: Je später der Follow-up-Test geschrieben wird, desto geringer ist der noch messbare Lernzuwachs nach dieser Zeit – unabhängig von der Methode (siehe Abbildung 27). Die Abstände zwischen Post- und Follow-up-Test liegen mit Ausnahme der HU I für beide Gruppen in etwa auf gleicher Höhe (6-7 Wochen). Für die Voruntersuchung wurden größere Abstände gewählt.

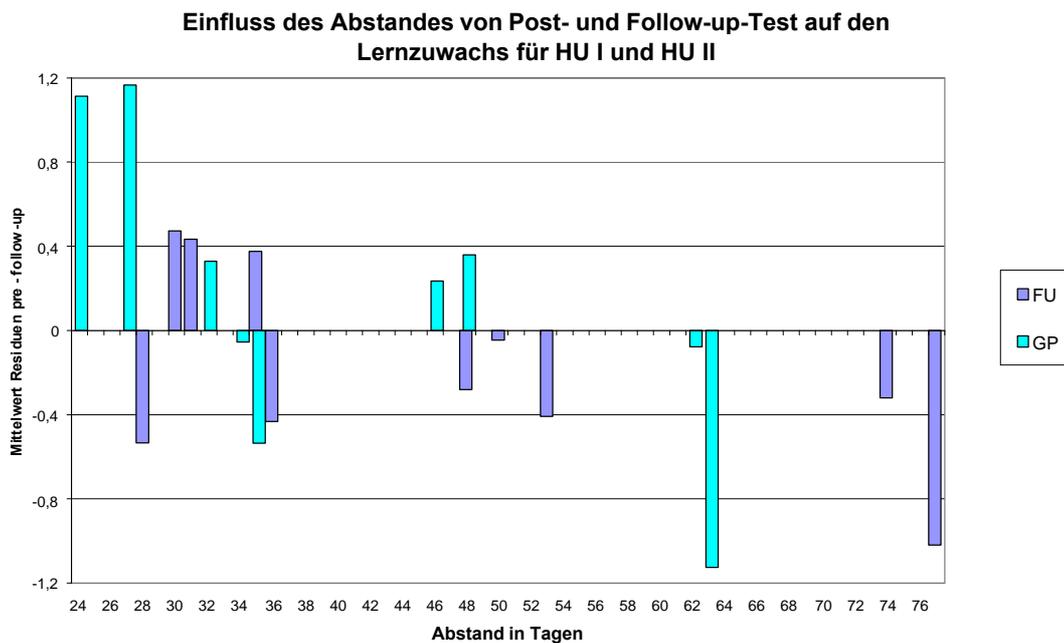


Abbildung 27: Abhängigkeit der mittleren Residuen (Pre-Test – Follow-up-Test) vom Abstand zum Post-Test – Follow-up-Test. Aufgeführt sind alle Gruppen der HU I und HU II_{intern}.

Tabelle 19: Einfluss der Unterrichtsmethode auf die Behaltensleistung: Ergebnisse der zweifaktoriellen Varianzanalyse. Die Residuen wurden aus Pre-Test und Follow-up-Test berechnet. Angegeben ist die Irrtumswahrscheinlichkeit p . Werte für Haupteffekte unter 0.01 gelten als signifikant (Sternchen), Wechselwirkungen können ab 0.2 ausgeschlossen werden. Für letztere sind die Effektstärken η^2 mit angegeben. Die Abstände wurden mediandichotomisiert, (siehe 7.1.7).

		VU	HU I	HU II intern	HU II extern
Stichprobenumfang N		67	149	180	225
Haupteffekte	Methode	0.020	0.605	0.000***	0.000***
	Abstand	0.680	0.000***	0.209	0.710
2-Weg-Wechselwirkungen	Methode x Abstand	0.342	0.347	0.399	0.852
Effektstärke η^2		0.014	0.006	0.004	0.000

Es sind keine Wechselwirkungen von Methode und Abständen nachweisbar.

Diskussion

Sowohl bei den Scores als auch bei den Residuen erreichten die Schüler der Interventionsgruppe durchweg bessere Werte, der Vorsprung blieb also erhalten. Damit konnten Ergebnisse von BORSCH et al. (2002) bestätigt werden. Erstaunlich ist, dass selbst nach einer langen Zeit (6-8 Wochen) keine gravierenden Wissensseinbußen zu verzeichnen sind, wie der Wilcoxon-Test mit Residuen (siehe Tabelle 18) zeigt. Diese Tatsache trifft auf beide Gruppen gleichermaßen zu. Rechnet man den Wilcoxon-Test mit Scores und vergleicht den Post-Test mit dem Follow-up-Test, wird die Wissensabnahme signifikant (siehe Anhang), und zwar für beide Gruppen. Anzumerken ist jedoch, dass die Verwendung von Scores für die Auswertung aus oben genannte Gründen (siehe Kapitel 7.1.1) zugunsten der Residuen nicht erfolgen soll. Insofern ist ein Vergleich mit den Ergebnissen von SUMFLETH et al. (2004, S. 290) schwierig, die in ihrer Untersuchung einen signifikanten Rückgang der mittleren Punktzahl (Scores Post-Test – Follow-up-Test) festgestellt haben, allerdings erst nach sechs Monaten.

Für die Erscheinung, dass es in der ersten Hauptuntersuchung keine Unterschiede mehr zwischen den beiden Stichproben gibt, sind verschiedene Erklärungen denkbar:

Einerseits mag es daran liegen, dass ein Lehrer den bei der Frontalunterrichtsgruppe nicht durchbekommenen Stoff nachgeholt hat und somit praktisch für den Test „geübt“ hat, was zum Verwischen von evtl. einmal vorhandenen Unterschieden geführt haben könnte.

Ferner konnte der Follow-up-Test in zwei Klassen erst nach den Sommerferien geschrieben werden, so dass zum einen die Motivation nicht mehr besonders groß war, zum anderen ebenso das Wissen, so dass auch hier Unterschiede verschwimmen. Auch fehlende Kontrolle durch den Untersuchungsleiter nach dem Post-Test mag zu Unwägbarkeiten geführt haben: Einige Lehrer haben den Follow-up-Test angekündigt, mit der Folge, dass deren Schüler mit evtl. eigenen Methoden dafür geübt und somit das Ergebnis verfälscht haben.

Zur Vergleichbarkeit der Abstände

Die Abstände innerhalb der Untersuchungen variierten von Klasse zu Klasse (nicht gruppenweise) stark – je nach Jahreszeit: Lag die Untersuchung noch weit vor den Sommerferien, so konnte auch der Follow-up-Test bis zu acht Wochen später stattfinden (z. B. VU), was bei Untersuchungen kurz vor den Sommerferien nicht mehr möglich war. Ebenfalls beeinflussend wirkten Elternsprechtage, Klassenfahrten etc. So war nicht gewährleistet,

dass wirklich in beiden Gruppen nach exakt der gleichen Periode die Follow-up-Tests geschrieben wurden. Für die erste Hauptuntersuchung betrug die mittlere Differenz sogar sechs Tage. Generell ist die Planung von Follow-up-Tests relativ schwierig, da im Schulalltag immer etwas dazwischen kommen kann. Da diese (Tests) jedoch nur einen kleinen Teil der Untersuchung ausmachen, soll ihre Aussagekraft nicht überbewertet werden.

Die Ergebnisse für die Hauptuntersuchungen sind nicht eindeutig und konnten nicht bestätigt werden. Daher soll das Fazit vorsichtiger formuliert werden: Auch nach 4-8 Wochen scheinen Gruppenpuzzle-Schüler einen Wissensvorteil gegenüber frontal unterrichteten Schülern zu haben. Es existieren keinerlei Wechselwirkungen von Methode und Abständen, d. h. die Unterrichtsmethode beeinflusst die Behaltensleistung der Schüler nicht. Um genauere Aussagen bezüglich der Langzeitwirkung der Methode treffen zu können wären weitere Studien erforderlich.

8.2 Meinungstest

8.2.1 Einstellung zum Fach Chemie

Die Ergebnisse des Meinungstests zum Fach Chemie fallen uneinheitlich aus: Einmal beurteilt die Kontrollgruppe das Fach etwas besser (HU I + HU II_{intern}), ein anderes Mal die Interventionsgruppe (VU, HU II_{extern}, siehe Abbildung 28). In keinem Fall sind die Unterschiede signifikant (siehe Tabelle 20 sowie folgende Diskussion).

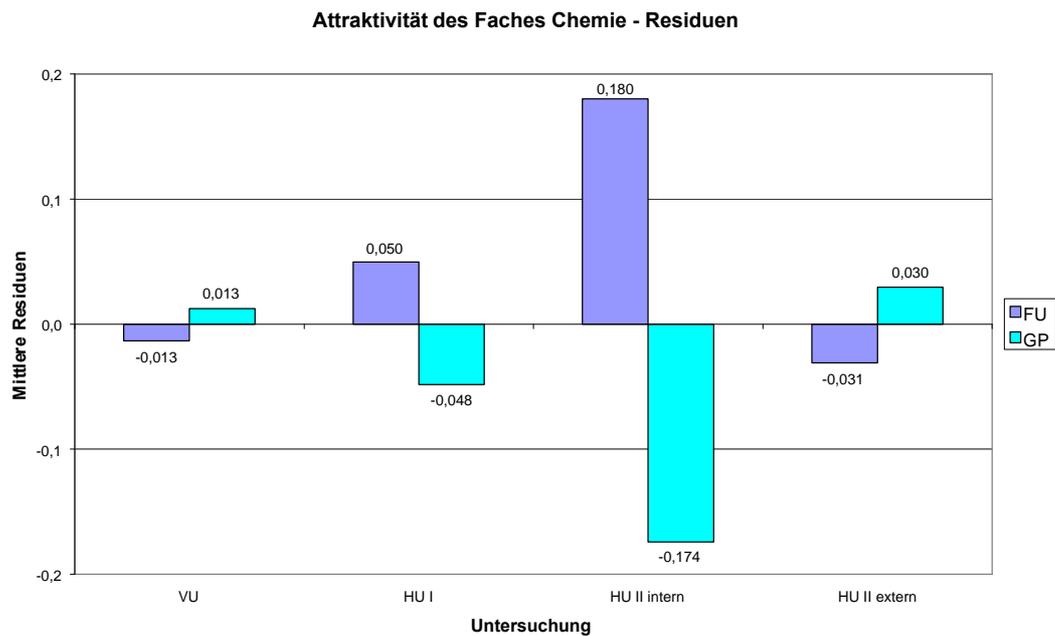


Abbildung 28: Attraktivität des Faches Chemie – Residuen: Ergebnisse des Meinungstests für Vor- und Hauptuntersuchung. Die Residuen wurden aus den Scores von Pre- und Post-Test berechnet.

Werden die Scores des Post-Tests betrachtet, so scheint die Gruppenpuzzlegruppe das Fach etwas attraktiver zu finden, lediglich die HU II_{intern} zeigt den gegenteiligen Befund (siehe Abbildung 29). Auch hier sind keine Signifikanzen auszumachen (siehe Tabelle 20).

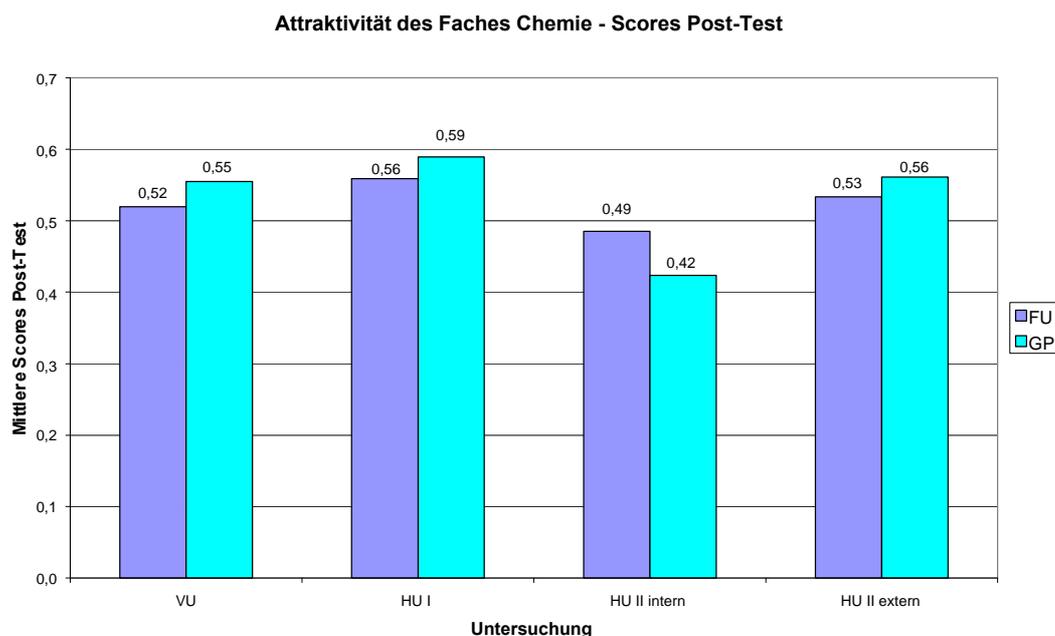


Abbildung 29: Attraktivität des Faches Chemie – Scores: Ergebnisse des Meinungstests für Vor- und Hauptuntersuchung.

Tabelle 20: Attraktivität des Faches Chemie: Ergebnisse des U-Tests von Vor- und Hauptuntersuchung. Cronbachs α wurde für den Post-Test auf Basis der umkodierte, noch nicht mediandichotomisierten Rohdaten bestimmt; die Stichprobengröße, die Signifikanztests und die Effektstärke beziehen sich jeweils auf die Residuen, welche aus Pre- und Post-Test-Scores errechnet wurden. Beim t-Test wurde die Varianzinhomogenität berücksichtigt. Soweit nicht anders angegeben beziehen sich die Irrtumswahrscheinlichkeiten auf die zweiseitige Testdurchführung. Grundsätzlich gilt: je höher der Score und je niedriger der Likert-Wert, desto positiver die Einschätzung.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 33 GP: 35	FU: 78 GP: 80	FU: 89 GP: 92	FU: 114 GP: 119
MW Scores, Pre-Test	FU: 0.46 GP: 0.50	FU: 0.50 GP: 0.57	FU: 0.46 GP: 0.44	FU: 0.50 GP: 0.53
MW Scores, Post-Test	FU: 0.52 GP: 0.55	FU: 0.56 GP: 0.59	FU: 0.49 GP: 0.42	FU: 0.53 GP: 0.56
MW Residuen _{pre-post}	FU: -0.013 GP: 0.013	FU: 0.050 GP: -0.048	FU: 0.180 GP: -0.174	FU: -0.031 GP: 0.030
MW Likert	FU: 2.95 GP: 2.85	FU: 3.01 GP: 2.96	FU: 2.60 GP: 2.69	FU: 2.95 GP: 2.90
t-Test Residuen, $p_{(z)}$	0.914	0.539	0.016	0.645
U-Test Residuen, $p_{(z)}$	0.941	0.680	0.013	0.632
U-Test Scores, $p_{(z)}$	0.497	0.402	0.101	0.494
Cronbachs α	0.913	0.934	0.924	0.938
Effektstärke d	0.03	0.10	0.36	0.06

Die Tabelle gibt im Wesentlichen die Befunde der Grafiken wieder.

Diskussion

Auf den ersten Blick sind die Residuen der HU II_{intern} signifikant unterschiedlich. Bedenkt man jedoch, dass für die Replikation ein Signifikanzniveau von $\alpha = 0.01$ gilt, so müssen die Unterschiede als zufällig entstanden bezeichnet werden. Die Scores der HU II_{intern} sind ebenfalls nicht signifikant different.

Zusammenfassend lässt sich festhalten, dass etwaige Unterschiede zufälliger Natur sind und die Methode keinen Einfluss auf die Beurteilung des Faches Chemie hat. Dieses Ergebnis verwundert nicht, da die Interventionsmaßnahme lediglich über drei Stunden erfolg-

te, ein Effekt allerdings erst nach längerer Dauer nachzuweisen wäre. Offensichtlich scheinen bereits vorhandene und evtl. manifestierte Schülereinstellungen durch kurze Interventionsmaßnahmen nicht beeinflusst zu werden.

8.2.2 Attraktivität der Unterrichtseinheit

Im Anschluss an die Untersuchung wurde die Unterrichtseinheit nach dem Gruppenpuzzle mit durchweg höheren Scores bewertet als der Frontalunterricht (siehe Abbildung 30). Die Unterschiede sind bis auf die HU II_{intern} signifikant (siehe Tabelle 21). Analog zu den Scores wurde auch auf der Likert-Skala dem Gruppenpuzzle eine positivere Beurteilung zugebracht (siehe Tabelle 21).

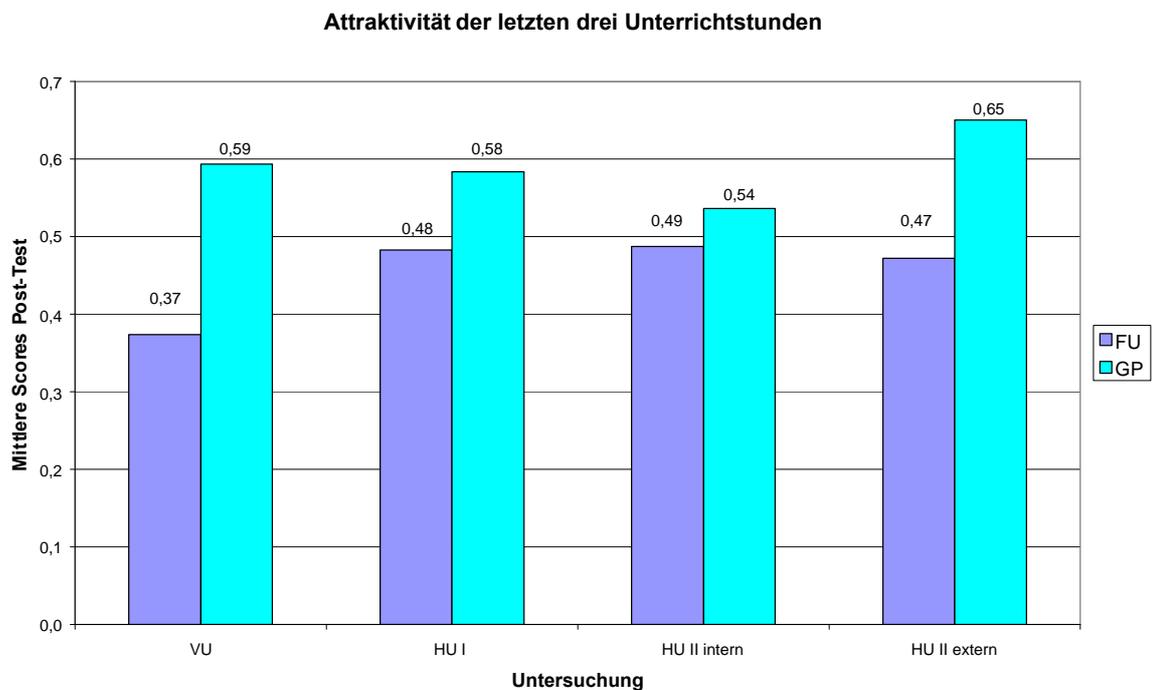


Abbildung 30: Attraktivität der Unterrichtseinheit – Scores: Ergebnisse des Meinungstests für Vor- und Hauptuntersuchung.

Tabelle 21: Attraktivität der letzten drei Unterrichtsstunden (nur Post-Test): Ergebnisse des U-Tests von Vor- und Hauptuntersuchung. Cronbachs α wurde auf Basis der umkodierten, noch nicht mediandichotomisierten Rohdaten bestimmt; die Stichprobengröße, die Signifikanztests und die Effektstärke beziehen sich jeweils auf Scores. Beim t-Test wurde die Varianzhomogenität berücksichtigt. Soweit nicht anders angegeben beziehen sich die Irrtumswahrscheinlichkeiten auf die einseitige Testdurchführung.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 35 GP: 36	FU: 78 GP: 81	FU: 89 GP: 92	FU: 113 GP: 118
MW Scores, Post-Test	FU: 0.37 GP: 0.59	FU: 0.48 GP: 0.58	FU: 0.49 GP: 0.54	FU: 0.47 GP: 0.65
MW Likert, Post-Test	FU: 2.45 GP: 2.05	FU: 2.36 GP: 2.13	FU: 2.25 GP: 2.15	FU: 2.55 GP: 2.21
t-Test Scores, $p_{(z)}$	0.000***	0.003**	0.070	0.000***
U-Test Scores, $p_{(z)}$	0.000***	0.004**	0.073	0.000***
Cronbachs α	0.914	0.914	0.900	0.923
Effektstärke d	0.98	0.43	0.22	0.69

Tabelle 22: Attraktivität der letzten drei Unterrichtsstunden nach Herausnahme der drei Klassen mit besonderen Rahmenbedingungen (siehe Text): Ergebnisse des U-Tests für Hauptuntersuchung II_{intern}. Die Irrtumswahrscheinlichkeiten beziehen sich auf die einseitige Testdurchführung.

	Stichprobenumfang N	MW Scores, Post-Test	U-Test Scores, $p_{(z)}$
HU II _{intern} ohne Dortmund	FU: 45 GP: 47	FU: 0.47 GP: 0.57	0.014

Diskussion

Im Prinzip sprechen die Ergebnisse dafür, dass den Schülern das GP besser gefallen hat als der Frontalunterricht. Dies konnte in drei unabhängigen Teil-Untersuchungen bestätigt werden. Auch BERGER & HÄNZE (2004b, S. 33) konnten eine Verbesserung der intrinsischen Motivation für das Gruppenpuzzle im Physikunterricht nachweisen. Zwei kooperativ durchgeführte Seminare einer deutschen Hochschule schnitten im Vergleich zur jeweiligen Kontrollgruppe ebenfalls deutlich positiver ab (vgl. JÜRGEN-LOHMANN et al. 2001, S. 74).

Die interne Hauptuntersuchung II (Replikation) jedoch weist keine signifikanten Unterschiede mehr auf. Daher wurde eine Detailanalyse durchgeführt (siehe Tabelle 22), in der die drei Klassen von der statistischen Analyse ausgeschlossen wurden, bei denen besondere Rahmenbedingungen galten (siehe nächster Absatz). Nach Herausnehmen dieser Klassen werden die Unterschiede zumindest auf dem 5 % Niveau signifikant, was allerdings für die Replikation zu wenig ist. Somit wird die Signifikanzgrenze knapp verfehlt.

Gründe für die weniger positive Beurteilung der Gruppenpuzzle-Einheit könnten die besonderen Umstände gewesen sein: In zwei Klassen des ersten Dortmunder Gymnasiums wurde der Test am letzten Schultag der Jahrgangsstufe 13 geschrieben, die Feier auf dem Schulhof und in der Schule war schon im Gange (siehe Abschnitt 8.1.1). Möglicherweise spielt auch die unzureichende Randomisierung (eine dieser beiden Dortmunder Klassen) eine Rolle. Hypothese: Französisch-Schüler, aus denen die GP-Gruppe zum großen Teil bestand, mögen das Fach Chemie generell nicht so gerne und dies speziell nach einer Woche Schüleraustausch sowie zwei Wochen Osterferien (siehe Abschnitt 8.1.1).

Die GP-Gruppe am zweiten Dortmunder Gymnasium musste ihren Test in der letzten Stunde vor den Pfingstferien schreiben. Dazu wurde er extra vorverlegt und noch dazu in die 6. Stunde. Das hatte zwar keinen Einfluss auf die Lernleistung, evtl. jedoch auf die Motivation.

Die schlechteren Scores der Gruppenpuzzlegruppe könnten aber auch schlichtweg durch positiveres Beurteilen der Frontalunterrichtseinheit verursacht werden: Vergleicht man nämlich die Mittelwerte der Likert-Skala (die nur Hinweise geben darf!) (siehe Tabelle 21), so fällt auf, dass das Gruppenpuzzle nicht schlechter, sondern der Frontalunterricht deutlich besser beurteilt wurde. Im Zuge der Mediandichotomisierung äußerte sich das in geringeren Scores für die GP-Einheit. Diese Erklärungsversuche bleiben allerdings spekulativ.

Warum letztendlich das Gruppenpuzzle in der HU II_{intern} im Vergleich nicht mehr ganz so positiv beurteilt wurde wie die vergleichbaren Gruppenpuzzle-Einheiten, kann mit dieser Untersuchung nicht geklärt werden.

Die dennoch deutlich bessere Beurteilung der Gruppenpuzzle-Methode ist insofern wichtig, als dass Zufriedenheit und Wohlbefinden im Unterricht eine entscheidende Rolle für den Lernerfolg spielen (vgl. BOLTE 2004, S. 5). Nur eine von den Schülern akzeptierte und gemochte Methode wird auch erfolgreich eingesetzt werden können. Auch Detailauswer-

tungen der PISA-Studie bestätigen, dass „*effektives Lernen nicht einfach als spezielle Fertigkeit gelehrt werden kann, sondern in erheblichem Maße von der Entwicklung einer positiven Einstellung abhängt*“ (ARTELT et al. 2004, S. 3).

8.2.3 Attraktivität des Gruppenpuzzles

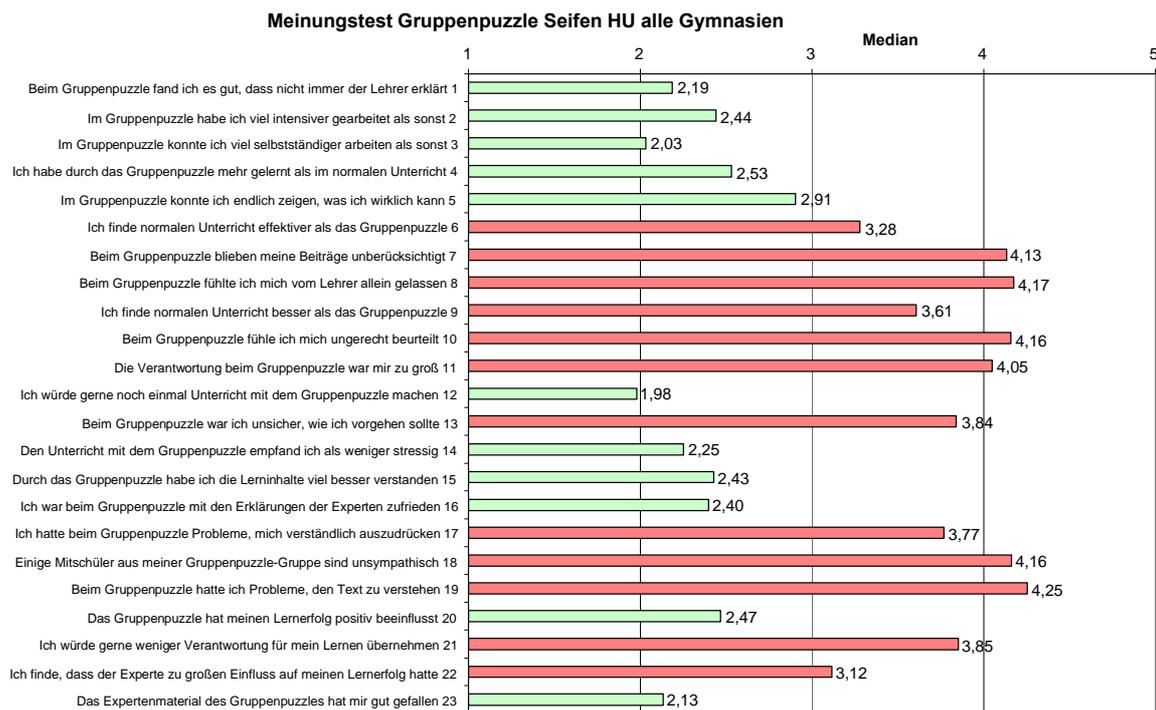


Abbildung 31: Ergebnisse des Meinungstests zur Attraktivität des Gruppenpuzzles für alle Gymnasien. Grüne Balken: positiv formulierte Items, rote Balken: negativ formulierte Items. Die Abbildung zeigt nicht umkodierte Rohdaten (Likert-Skala). Cronbachs α liegt zwischen 0.83 (VU) und 0.91 (HU II_{extern}).

Beschreibung und Diskussion

Das Balkendiagramm zeigt ausschließlich positive Wertungen für das Gruppenpuzzle zum Thema Seifen: Sämtliche grüne Balken liegen unter 3 („mittel“ auf der Likert-Skala), alle roten Balken über 3 (siehe Abbildung 31). Als besonders zutreffend wurde das selbstständige Arbeiten (Item 3) im Gruppenpuzzle beurteilt – ein Zeichen für den bisher hohen Frontalunterrichtanteil in den teilnehmenden Klassen.

Nicht so gut fanden die Schüler, dass das Abschneiden im Test ihrer Meinung nach relativ stark durch den jeweiligen Experten bestimmt wurde (siehe Item 22).

Dass die Schüler selbst von sich und ihren Vermittlungsleistungen nicht ganz überzeugt sind, zeigt Item 6: Zumindest ein kleiner Teil hält den herkömmlichen Unterricht für effektiver. Das mag daran liegen, dass sie ihre Ergebnisse des Wissenstests zum Zeitpunkt des Meinungstest noch nicht erfahren hatten.

Fast alle Schüler haben die Verantwortung des Vermittlers als nicht zu groß empfunden (siehe Item 11).

Aus Schülerperspektive zeichnet sich das GP-Seifen durch ein hohes Potenzial für selbstständiges Arbeiten und gute Textverständlichkeit (siehe Item 3, 19) aus. Dies führte zusammen mit der Möglichkeit sich einzubringen (Item 7) dazu, dass die Mehrheit das GP gerne noch einmal durchführen würde (siehe Item 12) und auch bereit wäre, künftig mehr Verantwortung für das eigene Lernen zu übernehmen (siehe Item 21).

8.3 Individuelle Schülerbeurteilungen

Die Aufteilung der Gruppen nach dem Matched-Pairs-Prinzip hat funktioniert: Beide Gruppen wurden von den Lehrern gleich gut eingeschätzt, wie der Mann-Whitney-U-Test ergab (siehe Tabelle 23). Dementsprechend niedrig sind die Effektstärken. Die Reliabilität hingegen ist mit einem Cronbachs Alpha-Werten über 0.9 sehr gut.

Tabelle 23: Vergleichbarkeit der Stichproben: Ergebnisse von Vor- (VU) und Hauptuntersuchung (HU) für das Polaritätenprofil „Mitarbeit“. Cronbachs α , Stichprobengröße und Effektstärke beziehen sich jeweils auf den Pre-Test.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 37 GP: 37	FU: 78 GP: 83	FU: 89 GP: 92	FU: 123 GP: 121
MW Scores, Pre-Test	FU: 0.51 GP: 0.53	FU: 0.56 GP: 0.58	FU: 0.53 GP: 0.54	FU: 0.57 GP: 0.53
U-Test, Scores $p_{(z)}$, zweiseitig	0.892	0.749	0.776	0.429
Cronbachs α	0.945	0.969	0.957	0.957
Effektstärke d	0.05	0.05	0.04	0.06

Tabelle 24: Korrelation Scores Schülerbeurteilung – Chemienote. Ergebnisse von Vor- (VU) und Hauptuntersuchung (HU) für das Polaritätenprofil „Mitarbeit“. Cronbachs α , Stichprobengröße und Effektstärke beziehen sich jeweils auf den Pre-Test.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 36 GP: 36	FU: 77 GP: 81	FU: 88 GP: 92	FU: 114 GP: 119
Korrelationskoeffizient nach Spearman	FU: -0.689 GP: -0.637	FU: -0.730 GP: -0.728	FU: -0.725 GP: -0.799	FU: -0.748 GP: -0.660
Signifikanz (2-seitig)	FU: 0.000 ^{***} GP: 0.000 ^{***}			

Tabelle 24 zeigt in allen Stichproben negative, hoch signifikante Korrelationen zwischen den Chemienoten und den auf Basis der individuellen Schülerbeurteilungen errechneten Scores.

Tabelle 25: Normalverteilung der individuellen Schülerbeurteilungen: Ergebnisse des Kolmogorov-Smirnov-Tests der beiden Stichproben (Mitarbeit, Pre-Test). Normalverteilung darf angenommen ab einem $p(z) \geq 0.2$, da es sich hierbei um die Wunschhypothese handelt.

	VU	HU I	HU II _{intern}	HU II _{extern}
Stichprobenumfang N	FU: 37 GP: 37	FU: 78 GP: 83	FU: 89 GP: 92	FU: 123 GP: 121
Kolmogorov-Smirnov-Z	FU: 1.175 GP: 0.962	FU: 1.424 GP: 1.449	FU: 1.128 GP: 1.096	FU: 1.414 GP: 1.550
Asymptotische Signifikanz (2-seitig)	FU: 0.126 GP: 0.313	FU: 0.035 GP: 0.030	FU: 0.157 GP: 0.181	FU: 0.037 GP: 0.016

Der Kolmogorov-Smirnov-Test auf Normalverteilung zeigt, dass bis auf die Gruppenpuzzlegruppe der Voruntersuchung keine Stichprobe normalverteilt ist. Das Ergebnis erzwingt die Berechnung des Korrelationskoeffizienten nach Spearman.

Ein erster Vergleich von Pre- und Post-Test ließ kaum Veränderungen erkennen, auch der U-Test nach Mann und Whitney konnte keine Unterschiede in den Beurteilungen und Residuen der beiden Gruppen aufzeigen (siehe Anhang).

Diskussion

Die hohe negative Korrelation von Chemienoten und den errechneten Scores lässt den Schluss zu, dass Chemienoten in ähnlicher Weise geeignet sind, die Leistung eines Schülers so einzuschätzen, dass damit ein Schülerranking zur Erstellung der Matched-Pairs erzeugt werden kann. Zwar sind Noten niemals so genau wie ein 16 Items umfassender Fragebogen, jedoch rechtfertigt dieser nur minimale Vorteil nicht den hohen zeitlichen Aufwand für die betroffenen Lehrer. In Zukunft können die Stichproben durchaus auf Basis der letzten Zeugnisnoten im Fach Chemie gebildet werden. Das Ergebnis ist nicht verwunderlich, wenn man bedenkt, dass sowohl Zeugnisnote als auch Beurteilung von ein und derselben Lehrperson stammen.

Das Schülerverhalten wurde im Nachhinein kaum anders beurteilt, woraus sich zwei Schlüsse ableiten lassen: Einerseits konnte die Beurteilbarkeit von sog. „Soft Skills“ durch die zusätzlichen Freiräume während des Gruppenpuzzles nicht verbessert werden. Andererseits scheinen die Lehrer auch im „normalen“ Unterricht das Schülerverhalten gut einschätzen zu können. Evtl. aber haben die Lehrer die Pre-Beurteilungen zu einem großen Teil für den Post-Fragebogen nur übernommen, da sie eine Änderung bei drei Stunden Intervention nicht für realistisch hielten.

8.4 Audioaufzeichnungen

Im Folgenden werden die Rede- bzw. Schweigedauern im Frontal- bzw. Gruppenpuzzleunterricht miteinander verglichen²⁶.

²⁶ Die gesamten Berechnungen wurden von B. ROEDER mit dem *BRS (Behavior Recording System)* und *Analyse* erstellt.

Konfigurationsfrequenzanalyse – Ergebnisse

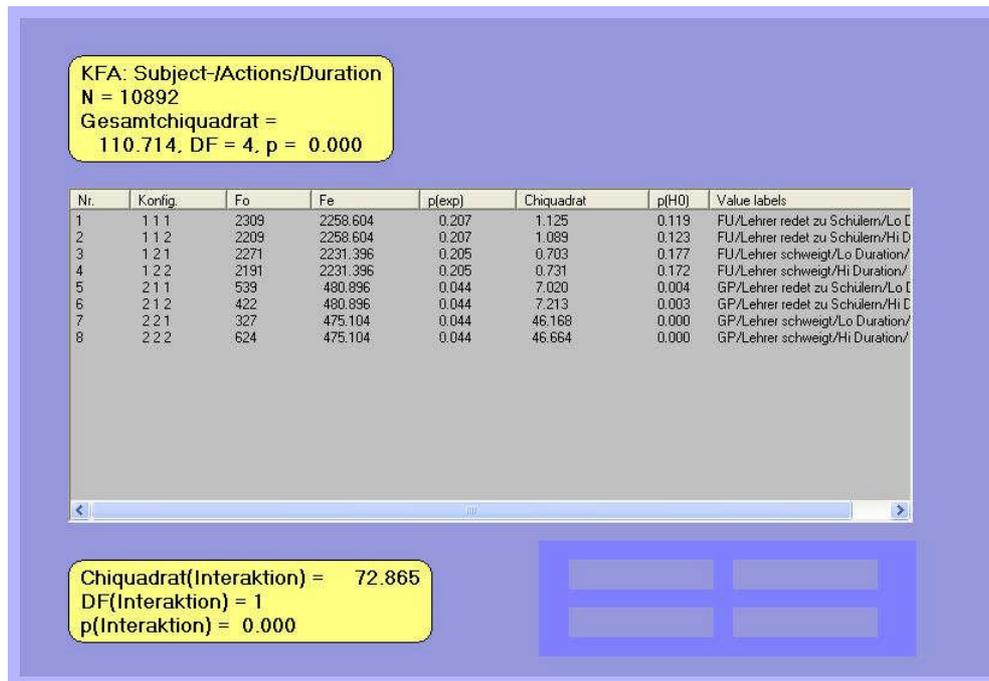


Abbildung 32: Ergebnisse der KFA mit Analyse.

Aus den drei Merkmalen Subject (FU/GP), Action (Lehrer redet zu Schülern/Lehrer schweigt) und Duration (high/low) resultieren insgesamt acht Zweier-Kombinationen. Für jede wurde ein χ^2 -Test durchgeführt. Dieser vergleicht die beobachteten Häufigkeiten F_o (Frequency observed), Spalte 3, mit den auf Basis der H_0 errechneten erwarteten Häufigkeiten F_e (F. expected), Spalte 4. In Spalte 6 sind die einzelnen Chiquadratkomponenten aufgeführt. Die exakte Wahrscheinlichkeit p (Binomialtest, Spalte 7) gibt an, ob die Abweichung der beobachteten Häufigkeiten F_o von den erwarteten Häufigkeiten F_e zufällig ist. Ein $p < 0.01$ sagt aus, dass dies zutrifft.

Das Gesamt- χ^2 liegt mit 110.714 deutlich über dem für 4 Freiheitsgrade²⁷ kritischen χ^2 -Wert von 13.28, die Unterschiede sind hoch signifikant ($p = 0.000$). Auffällig ist, dass die beobachteten Häufigkeiten F_o mit 327-624 ca. ein Fünftel so hoch sind im Vergleich zu den Häufigkeiten im Frontalunterricht (2191-2309). Besonders interessant ist, ob dieser beim Gruppenpuzzle seltenere Wechsel zwischen Reden und Schweigen zusätzlich noch mit kürzeren Redezeiten gekoppelt ist. Als statistische Prüfgröße hierzu dient das Chiquadrat der Wechselwirkung aus der hierarchischen KFA: Chiquadrat(Interaktion) =

²⁷ Die Freiheitsgrade für das Gesamt Chi-Quadrat mit m Merkmalen berechnen sich nach BORTZ 2005, S. 176 wie folgt: $df = 2^m - m - 1$. Für 3 Merkmale (Subject, Action, Duration) also: $2^3 - 3 - 1 = 4$.

72.865, $DF = 1$, $p < 0.000$ bei $\alpha = 0.01$. Dies bedeutet, dass die kurzen Redezeiten im Gruppenpuzzle häufiger als die langen vorkommen und zwar im Vergleich zum Frontalunterricht, bei dem eine Gleichverteilung vorliegt (siehe Abbildung 33). Bei den „Schweigezeiten“ im GP ist es umgekehrt: Die längeren sind häufiger.

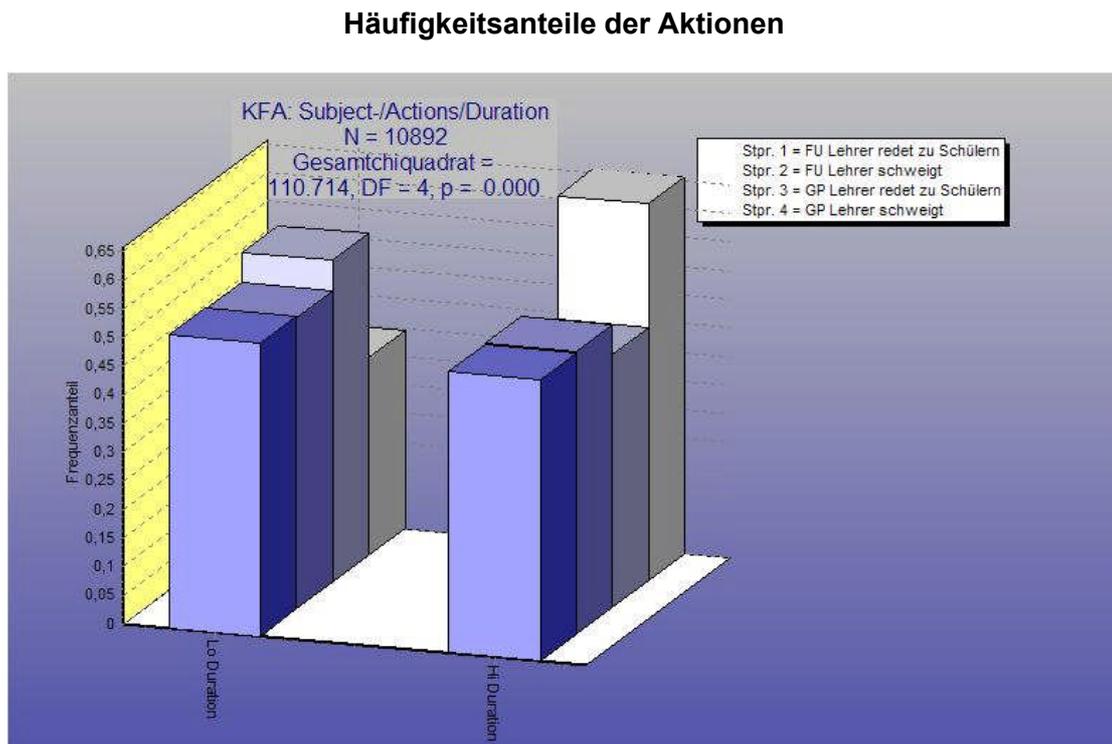


Abbildung 33: Häufigkeitsanteile der Actions (Reden-Schweigen) nach Dauern getrennt. Berücksichtigt wurden fünf bzw. sechs Lehrer/innen, eine Lehrkraft nahm mit zwei Klassen teil. Insgesamt wurden knapp 11.000 Eingaben gemacht (N=10892). Die vorderen beiden Säulen gehören zur FU-Gruppe, die hinteren beiden zur GP-Gruppe. Die Säulen sind von vorne nach hinten durchnummeriert. Innerhalb einer Stichprobe ergänzen sich „low“ und „high“ zu 1.

Deutlich zu erkennen sind die Unterschiede zwischen den beiden Gruppen: Sowohl für Reden als auch Schweigen liegen die Häufigkeitsanteile der FU-Gruppe (vordere zwei Säulen) auf einer Höhe. Das gilt für kurze Rededauern („low“, ca. 0.51) genauso wie für lange Rededauern („high“, ca. 0.49).

Für den Unterricht in der GP-Gruppe (Säule 3 + 4) ergibt sich auf den ersten Blick ein gemischtes Bild: Kurze Redeanteile treten etwas häufiger auf als im Frontalunterricht, gleichzeitig wird weniger lang gesprochen. Dementsprechend seltener sind die kurzen Schweigeperioden im Vergleich zu den längeren. Anders ausgedrückt: Im Gruppenpuzzleunterricht wird eher länger geschwiegen und kürzer geredet. Diese Unterschiede zwischen den beiden Gruppen sind hoch signifikant.

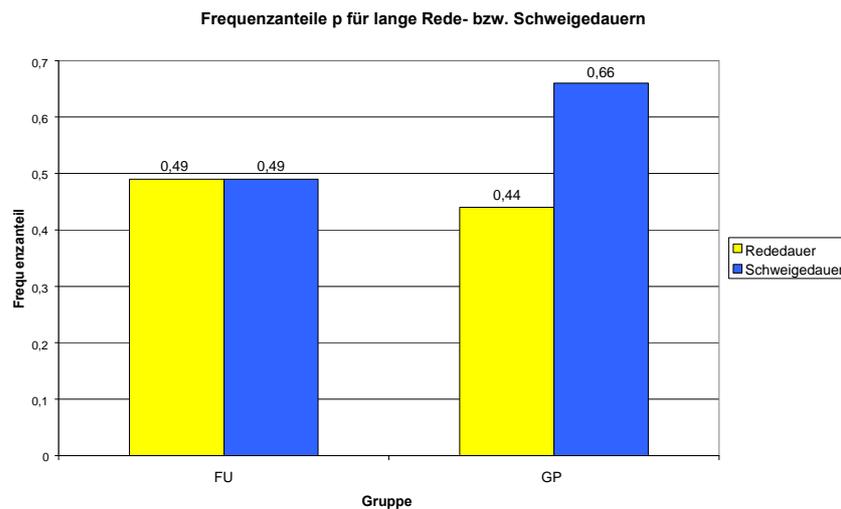


Abbildung 34: Vergleich der Häufigkeitsanteile (= geschätzte Wahrscheinlichkeiten) für lange Rede- oder Schweigedauer (> Median) nach Gruppen getrennt.

Mit 0.49 Anteilen ist die Wahrscheinlichkeit für eine lange Rededauer im Frontalunterricht genau so groß wie die Wahrscheinlichkeit einer langen Schweigedauer. Im kooperativen Unterricht redet die Lehrkraft weniger und schweigt dafür mehr, was in Zahlen heißt:

mittlere Rededauer im FU: 14 min
mittlere Rededauer im GP: 5 min²⁸.

Diskussion

Die Gründe für die stark unterschiedlichen Rededauern und Häufigkeiten liegen in der Unterrichtsmethode: Im klassischen fragend-entwickelnden Frontalunterricht findet konstruktionsbedingt ein häufigerer Wechsel von Reden und Schweigen statt, da der Unterricht auf einem „Frage-Antwort-Spiel“ basiert. Im Gruppenpuzzle übernimmt der Lehrer vorwiegend die Aufgabe eines Moderators bzw. Organisations, weshalb er deutlich weniger mit der Klasse interagiert. Der Redeanteil der Schüler dürfte im Gegenzug deutlich höher ausfallen, was an der hohen Aktivität in den Kleingruppen des Gruppenpuzzles liegt.

Fehlerbetrachtung

Die Datenkodierung erfolgte durch eine an der Untersuchung unbeteiligte Person. Insofern kann nicht unbedingt sichergestellt werden, dass nicht auch Gespräche mit dem Untersu-

²⁸ Anmerkung: Die absoluten Redezeiten wurden für lediglich fünf Lehrer, d. h. 30 Unterrichtsstunden, ermittelt. Je nach Engagement der Lehrkraft und benötigter Individualförderung können die Redezeiten während des Gruppenpuzzles stark variieren.

chungsleiter als U-Gespräch kodiert worden sind. Die Kodiererin wurde im Vorfeld zwar auf diese Gefahr hingewiesen, derartige Fehler sind mit letzter Sicherheit jedoch nicht auszuschließen. Als einzige Lösungsmöglichkeit bestünde die Kodierung durch den Untersuchungsleiter, was aus Zeitgründen nicht möglich gewesen ist. Daher bleibt eine Restunsicherheit.

8.5 Videoaufzeichnungen

Wie bereits in Abschnitt 6.5 erwähnt, sollten Videoaufzeichnungen nur zur Protokollierung eingesetzt werden, so dass im Nachhinein überprüft werden konnte, welche Umstände und Faktoren für das Abschneiden einer bestimmten Gruppe verantwortlich sein könnten. Tatsächlich war es notwendig, sich drei Gruppen einer Lehrkraft noch einmal anzusehen: Klasse Nr. 14 (FU und GP) im Jahr 2005 und zum Vergleich Klasse Nr. 8 (nur FU) in 2004 (siehe Tabelle 8).

Auslöser für das erneute Sichten der Bänder waren Abweichungen in den Ergebnissen des Wissenstests: Klasse 14 war in der HU II_{intern} die einzige Klasse, in der die Gruppenpuzzlegruppe einen geringeren Lernzuwachs als die Kontrollgruppe erzielt hat. Obwohl die Unterschiede nicht signifikant waren (siehe Abschnitt 8.1.1) wurden Erklärungen gesucht – und auch gefunden. Die Sichtung der Bänder führte zu folgendem Resultat:

Im Vergleich zu 2004 hat die Lehrkraft keine Unterschiede bei der Erteilung des Frontalunterrichts gemacht, insofern dürften die Gründe in der Testdurchführung oder bei der Durchführung des Gruppenpuzzles liegen. Bei der Durchführung des Tests ging es bei der GP-Gruppe relativ laut zu, da die Jahrgangsstufe 13 ihren letzten Schultag hatte und feiernd durch die Schule lief. Dieser Umstand sowie die Tatsache, dass direkt nach dem Test die Feier auf dem Schulhof stattfand, könnte zum schlechteren Abschneiden beigetragen haben.

Hinzu kam, dass acht Schüler bei der Erklärung des Gruppenpuzzles wegen eines Schüleraustauschs in Frankreich waren. So musste diese nachgeholt werden, was ca. 10 min Zeit in Anspruch nahm. Am Ende führte dies dazu, dass einige Versuche nur unter Zeitdruck und nach dem Klingeln durchgeführt wurden.

8.6 Stundenprotokolle

Die Durchsicht ergab, dass in den meisten Fällen die Lerninhalte dargeboten wurden. Im abweichenden Fall wurde dies bei der Auswertung berücksichtigt, indem die betreffenden Fragen des Wissenstests als korrekt gewertet wurden.

8.7 Beobachtungsbögen

Die Sichtung der Bögen ergab, dass alle Lerninhalte von den Schülern zumindest dargeboten wurden.

8.8 Klassen- und Lehrerfeedback

Nach dem jeweiligen Untersuchungsabschnitt wurde von den Schülern meist ohne Beisein des Lehrers eine kurze Rückmeldung bezüglich ihrer Unterrichtseinheit eingeholt. Empirisch fundierte Aussagen lassen sich daraus nicht ableiten, jedoch soll eine Auswahl zentraler Statements aufgeführt werden, die das durch den geschlossenen Fragebogen erhaltene Bild von der Untersuchung ergänzen soll:

FU-Gruppe

- Die Lehrer haben „normalen Unterricht“ gemacht und sich nicht verstellt.
- Einige Lehrer waren etwas aufgeregt wegen der Kamera/Beobachtung durch den Untersuchungsleiter.
- Einzelne Lehrer waren freundlicher als sonst (evtl. wegen der Kamera).
- Manche Lehrer haben etwas schneller unterrichtet.
- Die meist kleinere Gruppengröße war positiv.
- Es wurden mehr Experimente als sonst gemacht.
- Das Thema Seifen wurde als durchweg interessant empfunden.

GP-Gruppe

- Die Schüler fanden das GP gut.
- Das Material war gut erklärt und ansprechend.
- Die Schüler fanden gut, dass auch die Experten Schüler waren, da diese eher die „Sprache der Schüler“ sprechen.
- Die Schüler haben sich eher getraut nachzufragen als beim Lehrer.
- Der Test war fair.

- Die Schüler kamen mit dem Material gut alleine zurecht.
- Die Arbeit mit dem Gruppenpuzzle schien anstrengender für die Schüler zu sein.

Die Lehrer ergänzten, dass sie auch eher stillere Schüler als aktiver in der Kleingruppenarbeit wahrgenommen hätten. Diese Beobachtung wird von COHEN (1993, S. 51) bestätigt. Sie spricht diesbezüglich allgemein von „statusniedrigen“ Schülern.

9 Fazit und Ausblick

Bisher galten kooperative Unterrichtsmethoden trotz diverser positiver Erfahrungsberichte bei vielen Lehrern als unproduktiv und eher als Abwechslung denn als eine ernstzunehmende Unterrichtsmethode, die sie aus Überzeugung im Unterricht einsetzen würden. Mit dieser Studie konnte nachgewiesen werden, dass sich mit der Gruppenpuzzle-Methode nicht nur höhere Lernerfolge erzielen lassen, sondern dass diese auch gut von Schülern angenommen wird.

Folgende Hypothesen konnten bestätigt bzw. widerlegt werden:

Hypothese 1: *Schüler, die nach der Gruppenpuzzlemethode unterrichtet werden, erzielen einen höheren Wissenszuwachs als vergleichbar frontal unterrichtete Schüler.*

Diese Hypothese konnte zumindest für das konzipierte Gruppenpuzzle „Seifen“ in vollem Umfang bestätigt werden. Damit lassen sich erstmals Befunde internationaler Studien auf deutschen Chemieunterricht am Gymnasium übertragen.

Hypothese 2: *Die Methode des Gruppenpuzzles ist für Schüler attraktiver als vergleichbarer Frontalunterricht.*

Auch diese Hypothese konnte verifiziert werden, allerdings bleibt eine gewisse Restunsicherheit, da lediglich drei von vier Untersuchungen die nötigen Signifikanzschranken unterschritten haben. Die Replikation verfehlte das Signifikanzniveau knapp. Als Gründe lassen sich eine nicht absolute Randomisierung anführen, welche auf die realen schulischen Rahmenbedingungen zurückzuführen war, sowie ungünstige äußere Umstände, was jedoch bei einer derartigen Untersuchung nicht immer vermeidbar ist.

Nebenhypothese 1: *Das Gruppenpuzzle eignet sich für leistungsstarke bzw. -schwache Schüler gleichermaßen.*

Diese in internationalen Studien angezweifelte Hypothese erwies sich zumindest in der vorliegenden Untersuchung als zutreffend. Es konnten keinerlei Abhängigkeiten von Methode und Chemie-Zeugnisnote festgestellt werden. Interessant ist, dass mit „ausreichend“

vorbenotete Schüler durch das Gruppenpuzzle fast das Leistungsniveau von „Einser“-Frontalschülern erreichten.

Nebenhypothese 2: *Bei Themen, die sich die Schüler selbst angeeignet haben, erzielen sie einen höheren Wissenszuwachs als bei Themen, die ihnen von ihren Lehrern bzw. Mitschülern vermittelt wurden.*

Auch diese Hypothese ist positiv belegt worden: Die Experten für ein Gruppenpuzzle-thema schneiden hochsignifikant besser auf ihrem Themengebiet ab als ihre Mitschüler. Als Ursachen können die so genannte Lehr-Erwartung und eine aktivere Auseinandersetzung mit dem Lerngegenstand (Elaboration) vermutet werden. Analysiert man die „Zuhörer“, so haben die kooperativ „belehrten“ Schüler durchweg einen leichten Vorteil gegenüber den vom Lehrer unterrichteten Schülern. Die Ergebnisse sind jedoch nur in wenigen Fällen signifikant.

Nebenhypothese 3: *Das Gruppenpuzzle beeinflusst die Einstellung der Schüler zum Fach Chemie.*

Diese Hypothese muss negiert werden. Das Ergebnis wird erklärbar, wenn man die äußerst kurze Interventionszeit (3 Stunden) bedenkt. Um derartige Hypothesen zu überprüfen, müssten längerfristige Untersuchungen durchgeführt werden.

Die Untersuchung liefert Hinweise, dass das Matched-Pairs-Prinzip als Instrument der Randomisierung für die Durchführung einer komparativen Studie von großem Vorteil ist. Allerdings ist der zeitliche und organisatorische Aufwand sehr hoch, so dass es sich vorwiegend für kurze Unterrichtseinheiten eignet. Um den Arbeitsaufwand für die Lehrer bei der Aufteilung der Gruppen in Zukunft zu verringern, könnten statt individueller Schülerbeurteilungen auch Zeugnisnoten verwendet werden.

Die Auswertung über Residuen, die Verwendung des U-Tests sowie die Dichotomisierung der Meinungstests mittels Median erwies sich als praktikabel und zuverlässig. Gleichzeitig bietet diese Form der Auswertung mehr Vergleichsmöglichkeiten mit z. B. nicht normal-

verteilten Daten, wie sie entstehen können, wenn die Untersuchung spontan auf andere Schulformen erweitert wird.

Als Nebenergebnis konnte festgestellt werden, dass die Methode des Gruppenpuzzles auf beide Geschlechter sowie unterschiedlich leistungsfähige Schüler (Vornote) gleichermaßen positiv wirkt.

Aufgrund des stichpunktartigen Einsatzes an anderen Schulformen sowie in anderen Bundesländern kann vermutet werden, dass sich die Ergebnisse auch über NRW und das Gymnasium hinweg verallgemeinern lassen. Zur tatsächlichen Wirkung des Gruppenpuzzles auf andere Schulformen bzw. Altersklassen liefert diese Arbeit zwar Hinweise, jedoch sind weitere Studien notwendig. Genauso kritisch sollte das Gruppenpuzzle Seifen betrachtet werden: Wie erfolgreich andere – evtl. schlechtere – Gruppenpuzzles sind, muss in weiteren Studien evaluiert werden. Forschungsbedarf besteht ebenso bezüglich der Frage, wie sich der häufigere Einsatz des Gruppenpuzzles auf die Motivation der Schüler auswirkt.

10 Zusammenfassung

Ziel dieser Arbeit war es zu evaluieren, ob der Einsatz des Gruppenpuzzles im Chemieunterricht der Sekundarstufe I die Lernleistung und die Einstellung der Schüler im direkten Vergleich zu einer herkömmlich unterrichteten Lerngruppe erhöhen kann. Dabei wurde eine reale Unterrichtssituation experimentell untersucht.

Für eine maximale Vergleichbarkeit der Untersuchungsgruppen wurden die beteiligten Klassen nach dem Matched-Pairs-Prinzip in zwei Hälften geteilt. Grundlage hierfür bildeten individuelle Schülerbeurteilungen, die im Vorfeld von den Lehrern erstellt worden waren. Durch diese Randomisierung entstanden zwei weitgehend gleich leistungsstarke Lerngruppen. Sie erhielten nacheinander von der gewohnten Lehrkraft Unterricht zum Thema Seifen, wobei dieses bei der ersten Gruppe „herkömmlich“ (also im weitesten Sinne darlegend-informativ bzw. fragend-entwickelnd) und bei der Experimentalgruppe nach der Gruppenpuzzle-Methode geschah. In beiden Fällen waren die Lerninhalte identisch, da der Frontalunterricht ebenfalls auf Basis der Gruppenpuzzle-Materialien konzipiert wurde. Mit Hilfe von Stundenprotokollen, die durch den Untersuchungsleiter in jeder Stunde erstellt wurden, konnte sichergestellt werden, dass alle Inhalte dargeboten wurden. Eine abschließende Zusammenfassung bzw. Besprechung der Aufgaben fand beim Gruppenpuzzle nicht statt, um Methodeneffekte nicht durch den Eingriff der Lehrperson zu verfälschen. Die jeweils gerade nicht an der Untersuchung teilnehmende Gruppe wurde von studentischen Hilfskräften themenfremd unterrichtet. Unmittelbar vor bzw. nach Durchführung der Unterrichtsreihe wurden Wissens- und anonyme Meinungstests eingesetzt.

Insgesamt nahmen 413 Chemieschüler von 10. Gymnasialklassen an der Studie teil, ergänzt durch 235 Schüler anderer Schulformen bzw. mit einem abweichenden Untersuchungsdesign.

Ein weiterer Schwerpunkt war die Konzeption und Erprobung eines im realen Unterricht einsetzbaren Evaluationsinstrumentes, mit dem verschiedene Unterrichtsmethoden miteinander verglichen werden können. Dies betraf nicht nur die Randomisierung nach dem Matched-Pairs-Prinzip, sondern auch eine Auswertung über Residuen, U-Test sowie die Dichotomisierung der Meinungstests mittels Median.

Ergebnisse:

1. Das Gruppenpuzzle Seifen führt zu einem signifikant höheren Wissenszuwachs als vergleichbarer Frontalunterricht.
2. In drei von vier Untersuchungen wird das Gruppenpuzzle als signifikant attraktiver von den Schülern beurteilt.
3. Bringen sich die Schüler ihr Wissen eigenständig bei, erreichen sie signifikant höhere Werte im Wissenstest als wenn ihnen die Inhalte vom Lehrer oder ihren Mitschülern vermittelt wurden.
4. Das Gruppenpuzzle eignet sich für leistungsstarke bzw. -schwache Schüler sowie für beide Geschlechter gleichermaßen.
5. Eine Einstellungsänderung der Schüler zum Fach Chemie durch das Gruppenpuzzle konnte nicht festgestellt werden.
6. Im Gruppenpuzzle liegen die Rededauern des Lehrers deutlich unter denen im Frontalunterricht.
7. Die Auswertung über Residuen, der Einsatz des U-Tests sowie die Dichotomisierung der Meinungstests mittels Median erwies sich als praktikabel und vorteilhaft.
8. Das verwendete Evaluationsinstrument ist geeignet, um verlässliche Daten aus einer Vergleichsgruppenuntersuchung zu erheben.

Somit konnte festgestellt werden, dass die Methode des Gruppenpuzzles nicht nur beliebter bei Schülern ist, sondern auch zu einem signifikant höheren Lernzuwachs führt. Damit konnte ein wichtiges Argument der Kritiker von kooperativen Unterrichtsformen widerlegt werden. Zwar waren die Lerngruppen aufgrund des Untersuchungsdesigns (Teilung der Klassen) sowohl beim Gruppenpuzzle als auch beim Frontalunterricht kleiner als im realen Unterricht, jedoch dürfte sich die geringe Größe der Klasse bei der Frontalunterrichtsgruppe noch positiver auswirken. Eine abschließende Besprechung der Aufgaben, auf die aber aus o. g. Gründen in der Untersuchung verzichtet wurde, dürfte die Effektivität des Gruppenpuzzles ebenfalls zusätzlich steigern.

Ein Vergleich der Effektivität von selbst vermitteltem vs. fremd vermitteltem Wissen ergab, dass die Art der Wissensvermittlung eine entscheidende Rolle spielt: Die Experten der Gruppenpuzzlegruppe schnitten hoch signifikant besser ab als ihre Mitschüler, denen die Lerninhalte entweder von anderen Experten oder im Frontalunterricht vom Lehrer vermittelt wurden. Betrachtet man die Fremdvermittlungsarten separat, fallen die Ergebnisse sehr uneinheitlich aus und sind nicht replizierbar. Daher muss das Fazit diesbezüglich sehr vorsichtig formuliert werden: In einigen Fällen ist das Erklären durch Mitschüler effektiver als durch den Lehrer, zumindest aber gleich effektiv.

Weder in der Kontroll- noch in der Interventionsgruppe konnten geschlechtsspezifische Unterschiede bezüglich des Lernzuwachses nachgewiesen werden: Die Methode wirkt auf beide Geschlechter nahezu identisch. Gleiches gilt für Wechselwirkungen zwischen Methode und Chemie-Zeugnisnote: Das Gruppenpuzzle eignet sich für leistungsstarke bzw. -schwache Schüler gleichermaßen.

Die Auswertung der Meinungstests lieferte nur teilweise signifikante Unterschiede: Konnten kaum Differenzen bei der Einstellung zum Fach Chemie gefunden werden, so wurde doch die Attraktivität der letzten drei Gruppenpuzzle-Stunden von den betreffenden Schülern deutlich besser beurteilt als die letzten drei Frontalunterrichtsstunden. Allerdings sind die Unterschiede nur in drei von vier Untersuchungen signifikant. Dennoch kann von einer höheren Beliebtheit des Gruppenpuzzles bei den Schülern ausgegangen werden. Wie sich der häufigere Einsatz dieser Methode auswirkt, stellt ein zukünftiges Forschungsdesideratum dar.

Neben Effizienz und Beliebtheit konnte ein deutlicher Rückgang des Redeanteils des Lehrers während des Gruppenpuzzles festgestellt werden, so dass diese Unterrichtsmethode einen Teil zur Entlastung der Lehrer beitragen kann.

Mit dem eingesetzten Evaluationsinstrument, bestehend aus Wissens-/Meinungstest im Pre-Post-Design, der Randomisierung nach dem Matched-Pairs-Prinzip sowie Audio- und Videoaufzeichnungen, konnten replizierbare Daten aus der Interventions-/Kontrollgruppenuntersuchung gewonnen werden. Dabei kamen neben den klassischen Mittelwertvergleichen auch relativ unbekanntere Auswerteverfahren wie das über Residuen oder eine Mediandichotomisierung zum Einsatz. Der entstandene Aufwand war hoch, jedoch gerechtfertigt.

Einschränkend bleibt zu sagen, dass die Effektivität des Gruppenpuzzles zum Thema Seifen lediglich für Chemieschüler der Jahrgangsstufe 10 an nordrhein-westfälischen Gymnasien evaluiert wurde. Die exemplarische Ausweitung auf andere Bundesländer und andere Schulformen, die erfolgreiche Durchführung auch ohne Anwesenheit des Lehrers, eine streng konservative Auslegung der Untersuchung sowie die Kongruenz mit Ergebnissen anderer Studien sprechen jedoch dafür, dass mit dieser kooperativen Lernform ein effektiver Beitrag zur Unterrichtsverbesserung im deutschen Chemieunterricht geleistet werden kann.

Danksagung

Mein Dank gilt allen, die zum Gelingen meiner Arbeit in den Jahren 2003 – 2007 beigetragen haben. Besonders bedanken möchte ich mich bei:

Meiner Chefin Frau Prof. Dr. Insa Melle. Das mir von ihr entgegengebrachte Vertrauen, die ausgezeichneten Arbeitsbedingungen am Lehrstuhl, die Möglichkeit für ein selbstbestimmtes, freies Forschen und ein offenes, ehrliches Verhältnis trugen erheblich zum Gelingen der Arbeit bei.

Meinem inoffiziellen „2. Doktorvater“ Herrn Prof. Dr. em. Burkhard Roeder. Seine Unterstützung in statistischen Fragen sowie seine Begeisterung für diese Arbeit, die sich z. B. in einem offenen Ohr zu jeder Tages- und Nachtzeit äußerte, waren eine unerlässliche Hilfe.

Herrn Prof. Dr. Bernd Ralle für die Erstellung des Korreferats.

Meinem Bruder und Arbeitskollegen Oliver Tepner. Sein kritisches Hinterfragen sowie seine Fähigkeit und sein Wille, sich auch in meine Arbeit hineinzudenken, haben mich an entscheidenden Stellen weitergebracht. Ich denke, dass wir in den letzten fünf Jahren ein ziemlich gutes Team waren, das mit Humor und gegenseitiger Hilfe für das gemeinsame Ziel gestritten hat. Danke!

Den Lehrstühlen DC 1 und DC 2 für ein kollegiales und humorvolles Arbeitsklima. Vor allem Frau Kirsten Krebs und den studentischen Hilfskräften gilt mein Dank, ohne die ein derartig großes Projekt nicht hätte durchgeführt werden können. Weiterhin danken möchte ich Werner Pöpping für seine tatkräftige Hilfe bei der Entwicklung des Gruppenpuzzles sowie Monika Zimmer für ihre Informationen zum Advance Organizer.

Meinen Eltern, die mir den Rücken frei gehalten und mich zu einem neugierigen Menschen gemacht haben.

Henner Frebel und Cornelia Hülsken für ihre Durchsicht in Rekordzeit.

Allen Lehrerinnen und Lehrern für ihren Einsatz und ihre teilweise unorthodoxe Art, mir bei organisatorischen Schwierigkeiten zu helfen.

Den Schülern, die mehr oder weniger freiwillig diese Untersuchung erst ermöglicht haben.

Nicht nur bedanken sondern vor allem entschuldigen möchte ich bei meiner Freundin Cornelia Hülsken, die viele Sonntage ohne mich „Tatort“ gucken musste...

11 Literatur

1. AMERICAN EDUCATIONAL RESEARCH ASSOCIATION (AERA) (2005): Auf die Lehrer kommt es an - Neuere Forschungsergebnisse zu den Bedingungen des Lernfortschritts von Schülern. In: Pädagogik, Bd. 57, Heft 7-8, S. 70-75.
2. ARONSON, E., STEPHEN, C., SIKES, J., BLANEY, N., & SNAPP, M. (1978): The Jigsaw Classroom. Sage Publication, Beverly Hills, CA.
3. ARTELT, C., BAUMERT, J., JULIUS-McELVANY N. & PESCHAR, J. (2004): Das Lernen lernen: Voraussetzungen für lebensbegleitendes Lernen. Ergebnisse von PISA 2000. OECD, Paris.
4. AUSUBEL, D. P. (1980): Psychologie des Unterrichts. Bd. 1, 2. Beltz: Weinheim, Basel.
5. BUNDESARBEITSGEMEINSCHAFT FÜR HALTUNGS- UND BEWEGUNGSFÖRDERUNG E.V. (BAG) (2005): MAGAZIN - Bewegung im Unterricht fördert die Konzentration. In: Pädagogik, Bd. 57, Heft 9, S. 59.
6. BALLSTAEDT, S.-P. (1993): Schriftliche Unterlagen für Schülerhand. In: Pädagogik, 5/93, S. 24-27.
7. BERGER, R. & HÄNZE, M. (2003). Motivation, Lernen und Leistung im kooperativen Physikunterricht. In: A. Pitton (Hrsg.) Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung, S. 247-249. LIT, Münster.
8. BERGER, R. & HÄNZE, M. (2004): Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II – Einfluss auf Motivation, Lernen und Leistung. Zeitschrift für Didaktik der Naturwissenschaften; Jg. 10, S. 205-219.
9. BERGER, R. & HÄNZE, M. (2004b): Förderung intrinsischer Motivation durch das Gruppenpuzzle. In: *Naturwissenschaften im Unterricht - Physik*, 15, Heft 84, S. 33.
10. BERGER, R. & HÄNZE, M. (2005): Gruppenpuzzle im Physikunterricht der Sekundarstufe II – welche Unterschiede gibt es zum Lernzirkel? In: Pitton, Anja (Hrsg.): Rele-

- vanz fachdidaktischer Forschungsergebnisse für die Lehrerbildung, S. 152-154. LIT, Münster.
11. BODNER, G.M. (1986): Constructivism: A theory of knowledge. *Journal of Chemical Education*, Vol. 63, pp. 873-878.
 12. BOLTE, CLAUS (2004): Motivation und Lernerfolg im Chemieunterricht der Sekundarstufe I. *PdN-Ch - Chemie in der Schule* 2/53, 2-5.
 13. BORSCH, F., JÜRGEN-LOHMANN, J. & GIESEN, H. (2002). Kooperatives Lernen in Grundschulen: Leistungssteigerung durch den Einsatz des Gruppenpuzzles im Sachunterricht. *Psychologie in Erziehung und Unterricht*. 49, 172-183.
 14. BORTZ, J. (2005): Statistik für Human- und Sozialwissenschaftler. Springer Medizin Verlag, Heidelberg.
 15. BORTZ J., DÖRING, N. (1995): Forschungsmethoden und Evaluation für Sozialwissenschaftler. Springer, Berlin.
 16. BORTZ J., DÖRING, N. (2003): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Springer, Berlin.
 17. BORTZ, J. LIENERT, G.A. (2003): Kurzgefasste Statistik für die klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben. Springer, Berlin.
 18. BOSSE, D. (2003): Differenziertes Lernen bis zum Abitur. In: *Pädagogik* - 55 9, S. 24-27.
 19. BOWEN, C. W. (2000): A quantitative literature review of cooperative learning effects on high-school and college chemistry achievement. *Journal of Chemical Education*, 77(1), 116-119.
 20. BRAND, M., MARKOWITSCH, H. J. (2004): Lernen und Gedächtnis. In: *Praxis der Naturwissenschaften – Biologie in der Schule*, 53 (2004) 7, S. 1-7.
 21. BROSIUS, F. (2002): SPSS 11. mitp Verlag, Bonn.

22. BÜHNER, M. (2004): Einführung in die Test- und Fragebogenkonstruktion. Pearson Studium, München.
23. BÜHL, A., ZÖFEL, P. (2005): SPSS 12. Einführung in die moderne Datenanalyse unter Windows, 9. Auflage. Pearson Studium, München.
24. CATTEL, R. B. (1966): The scree test for the number of factors. *Multivariate behav. Res.* 1, 245-276.
25. CHI, M. T. H., DE LEEUW, N., CHIU, M. H. & LAVANCHER, C. (1994): Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-477.
26. COHEN, E. G. (1993). Bedingungen für produktive Kleingruppen. In: G.L. Huber. *Neue Perspektiven der Kooperation: ausgewählte Beiträge der Internationalen Konferenz 1992 über Kooperatives Lernen in Utrecht, Holland.* (S. 45-53). Baltmannsweiler: Schneider, Hohengehren.
27. COHEN, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. In: *Review of Educational Research*, 64, S. 1-35.
28. COMENIUS, J. A. (1954). *Große Didaktik.* In neuer Übers. hrsg. von Andreas Flitner, S. 117. Küpper, Düsseldorf.
29. DANSEREAU, D. F. (1998): Cooperative learning strategies. In: C. E. Weinstein, E. T. Goetz & P. A. Alexander (Eds.), *Learning and Study Strategies: Issues in Assessment, Instruction and Evaluation* (pp. 103-120). Academic Press, New York.
30. DE BAZ, T. (2001): The Effectiveness of Jigsaw Cooperative Learning on Students' Achievement and Attitudes toward Science. In: *Science Education International*, Vol. 12, 4, p 6-11.
31. DECI, E. L. & RYAN, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik* 39, 223-238.
32. DER SPIEGEL (2005): Lernen ohne Lehrer. 23/2005, S. 63.
33. DORI, Y. J., YAROSLAVSKY, O. and LAZAROWITZ, R. (1995): The Effect of Teaching the Cell Topic Using the Jigsaw Method on Students' Achievement and Learning Ac-

- tivity, ERIC Clearinghouse for Science, Mathematics and Environmental Education. Columbus, Ohio.
34. DUIT, R.: Research on students' conceptions -- developments and trends. In: Novak, J. (Ed): Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics. Ithaca: Cornell University (Zugang per Computerdiskette oder Internet), 1993.
 35. EILKS, I. (2003a): Kooperatives Lernen im Chemieunterricht (Teil 1). In: Der mathematische und naturwissenschaftliche Unterricht 1/56, S. 51-55.
 36. EILKS, I. (2003b): Kooperatives Lernen im Chemieunterricht (Teil 2). In: Der mathematische und naturwissenschaftliche Unterricht 2/56, S. 111-115.
 37. EILKS, I., RALLE, B. (2002): Partizipative Fachdidaktische Aktionsforschung – Ein Modell für eine begründete und praxisnahe curriculare Entwicklungsforschung in der Chemiedidaktik. In: Chemie konkret, Jahrgang 1, Heft Nr. 9, S. 13-18.
 38. EILKS, I., STÄUDEL, L. (2005): Kooperativ Lernen – Naturwissenschaften im Unterricht Chemie 16, Nr. 88/89.
 39. EILKS, I., STÄUDEL, L. (2005a): Vorwort. In: Naturwissenschaften im Unterricht Chemie 16, Nr. 88/89, S. 3.
 40. EILKS, I., STÄUDEL, L. (2005b): Warum kooperatives Lernen? In: Naturwissenschaften im Unterricht Chemie 16, Nr. 88/89, S. 4-5.
 41. EILKS, I., WITTECK, T., RUMANN, S., SUMFLETH, E. (2005): Kooperatives Lernen. In: Naturwissenschaften im Unterricht Chemie 16, Nr. 88/89, S. 6-11.
 42. ENTWISTLE, N. (1995): The use of research on student learning in quality assessment. In: Gibbs, G.: Improving Student Learning through assessment and Evaluation, Oxford.
 43. EPPLER, R., WINTER, M., HUBER, G. L. (1985): Kooperatives Lernen als "Gruppenpuzzle". Tuebingen.

-
44. FALLER, H. (2004): Signifikanz, Effektstärke und Konfidenzintervall. Serie Methoden in der Rehabilitationsforschung, *Rehabilitation*, 43 (3), S. 174-178.
 45. FISCHER, H. E., KLEMM, K., LEUTNER, D., SUMFLETH, E., TIEMANN, R. & WIRTH, J. (2003): Naturwissenschaftsdidaktische Lehr-Lernforschung: Defizite und Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*; Jg. 9, S. 179-209.
 46. FÖLLING-ALBERS, M. (1992): *Schulkinder heute. Auswirkungen veränderter Kindheit auf Unterricht und Schulleben*. Beltz, Weinheim.
 47. FREY-EILING, A. & FREY, K. (1999). Das Gruppenpuzzle. In J. Wiechmann (Hrsg.), *Zwölf Unterrichtsmethoden. Vielfalt für die Praxis* (S. 50-57). Beltz, Weinheim.
 48. FRIEDE, CH. K. (1994): Sozialkompetenz als Ziel der Berufserziehung – begriffsanalytisch betrachtet. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 90. S. 606-625.
 49. FROTSCHER, J. (2004): Erfolgreiche Lernprozesse durch Ausgewogenheit in der Methodenwahl. In: *Grundschulunterricht* 5, S.7-10.
 50. FÜRSTENAU, B. (2002). Das kooperative Lehr-Lern-Arrangement Gruppenpuzzle. *Berufsbildung* , 56 (76), 41-42.
 51. GERSTENMAIER, J. & MANDL, H. (1995): Wissenserwerb unter konstruktivistischer Perspektive. *Zeitschrift für Pädagogik* 41, 6: S. 867-888.
 52. GLYNN, S. M., YEANY, R. H., & BRITTON, B. K. (1991): A constructive view of learning science. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science*. (pp. 3-19). Erlbaum, Hillsdale, NJ.
 53. GLYNN, S., DUIT, R. (1995): Learning science meaningfully: Constructing conceptual models. In: S. Glynn, R. Duit (Hrsg): *Learning science in the schools: Research reforming practice*, S. 3-33. Lawrence Erlbaum, Mahwah, NJ.
 54. GRADY, K. & STEVENS, M. (2003): *The Impact of School Inputs on Student Performance: An Empirical Study of Private Schools in the United Kingdom*. Economics Series Working Papers 146, University of Oxford, Department of Economics.

-
55. GRÄSEL, C. & GRUBER, H. (2000): Kooperatives Lernen in der Schule. Theoretische Ansätze - Empirische Befunde - Desiderate in der Lehramtsausbildung. In N. Seibert (Hrsg.), Unterrichtsmethoden (S. 161-176). Klinkhardt, Opladen.
 56. GREEN, N. & GREEN, K. (2005): Kooperatives Lernen im Klassenraum und im Kollegium. Das Trainingsbuch. Kallmeyer, Seelze-Velber.
 57. GUDJONS, H. (1995): Pädagogisches Grundwissen. Klinkhardt, Bad Heilbrunn.
 58. HAAG, L. (2005): Gruppenmethoden und Gruppenarbeit, in: Pädagogik, H. 3/2005, S. 26-30.
 59. HELMKE, A. (2006): Was wissen wir über guten Unterricht? Pädagogik 2/06, 42-45.
 60. HUBER, A. (1999): Bedingungen effektiven Lernens in Kleingruppen unter besonderer Berücksichtigung der Rolle von Lernskripten. Huber, Schwangau.
 61. HUBER, A., KONRAD, K., WAHL, D. (2001): Lernen durch wechselseitiges Lehren. In: Pädagogisches Handeln 5/ 2001, S. 33-46.
 62. HUBER, G. L. (1985), Pädagogisch-psychologische Grundlagen für das Lernen in Gruppen, Studienbrief 1: Lernen in Schülergruppen, Tübingen: Deutsches Institut für Fernstudien.
 63. HUBER, G. L. (2000): Lernen in kooperativen Arrangements. In: Duit, R., v. Rhöneck, C. (Hrsg.): Ergebnisse fachdidaktischer und psychologischer Lehr-Lernforschung, Kiel.
 64. HUBER, O. (2005⁴): Das psychologische Experiment: Eine Einführung. Huber, Bern.
 65. JOHNSON, D.W. & JOHNSON, R.T., HOLUBEC, E.J. (1984): Cooperation in the Classroom. Interaction Book Company, Edina, MN.
 66. JOHNSON, D. W., & JOHNSON, R.T. (1989). Cooperation and competition: Theory and research. Interaction Book Company, Edina, MN.
 67. JOHNSON, D.W. & JOHNSON, R.T. (1994): An Overview of Cooperative Learning. In: Thousand, J. S. et al. (Eds.) Creativity and Collaborative Learning: A Practical Guide

- to Empowering Students and Teachers, S. 31–44. Brooks Publishing Co., Baltimore, MD.
68. JOHNSON, D.W. & JOHNSON, R.T. (1999): Learning Together and Alone. Cooperative, Competitive and Individualistic Learning. Allyn & Bacon, Boston.
69. JOHNSON, D.W., JOHNSON, R.T., STANNE, M. B. (2000): Cooperative Learning Methods: A Meta-Analysis. <http://www.co-operation.org/pages/cl-methods.html>
70. JÜRGEN-LOHMANN, J., BORSCH, F. & GIESEN, H. (2001). Kooperatives Lernen an der Hochschule: Evaluation des Gruppenpuzzles in Seminaren der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*, 15, 74-84.
71. JÜRGENS, B. (1996): Verhaltensgestörte Kinder oder sozial inkompetente Pädagoginnen? In: Heckt, D. H. und Jürgens, E.: Anders kommunizieren lernen. Westermann, Braunschweig.
72. LAZAROWITZ, R. (1991) et al.: Learning biology cooperatively. In: Cooperative Learning 11, No. 3, S. 19-21.
73. LAZAROWITZ, R., HERTZ-LAZAROWITZ, R. & BAIRD, H. (1994): Learning science in a cooperative setting: Academic achievement and affective outcomes. In: Journal of Research in Science Teaching, 31(10), 1121-1131.
74. LAZAROWITZ, R. & HERTZ-LAZAROWITZ, R. (1998): Cooperative Learning in the Science Curriculum. In: Fraser, B. J. & Tobin, K. G. (Eds.). International Handbook of Science Education, pp. 449-469. Kluwer Academic Publishers, Netherlands.
75. LAZAROWITZ, R., & KARSENTY, G. (1990). Cooperative learning and students' achievement, process skills, learning environment, and self-esteem in tenth grade biology classrooms. In: S. Sharan (Ed.), Cooperative learning: Theory and research (pp. 123-149). Praeger, New York.
76. LIENERT, G. A., RAATZ, U. (1998): Testaufbau und Testanalyse. Beltz, Weinheim, Basel.

-
77. LIENERT, G. A., VON EYE, A. (1994): Erziehungswissenschaftliche Statistik. Beltz, Weinheim, Basel.
 78. LORD, T. (1999): A comparison between traditional and constructivist teaching in Environmental Science, *The Journal of Environmental Education* 30(3) 22-27.
 79. MARKIC, S., EILKS, I. (2005): Kooperatives und kontextorientiertes Lernen zu Batterien und Akkumulatoren in der SI. In: A. Wellensiek, M. Welzel, T. Nohl (Hrsg.): *Didaktik der Naturwissenschaften - Quo vadis?* Logos, Berlin. S. 115-125.
 80. MASENDORF, F. (1997): *Experimentelle Sonderpädagogik. Ein Lehrbuch zur angewandten Forschung.* Beltz, Weinheim.
 81. MELLE, I. PÖPPING, W. (2002): Unterrichtsverfahren im Chemieunterricht des Gymnasiums. *Zur Didaktik der Physik und Chemie*, Hrsg. H. Behrendt, S. 212–214. Leuchtturm, Alsbach.
 82. MEYER, H. (2003): Zehn Merkmale guten Unterrichts. In: *PÄDAGOGIK* 10/03, S. 36–43.
 83. MEYER, H. (2004): *Was ist guter Unterricht?* Cornelsen, Berlin.
 84. MINISTERIUM FÜR SCHULE UND WEITERBILDUNG, WISSENSCHAFT UND FORSCHUNG DES LANDES NRW (1999): *Richtlinien und Lehrpläne für die Sekundarstufe II – Gymnasium/Gesamtschule in NRW.* Ritterbach-Verlag, Frechen.
 85. MINISTERIUM FÜR SCHULE UND WEITERBILDUNG, WISSENSCHAFT UND FORSCHUNG DES LANDES NRW (1993): *Richtlinien und Lehrpläne für das Gymnasium - Sekundarstufe I –in Nordrhein- Westfalen.* Ritterbach-Verlag, Frechen.
 86. MÖLLER, K. (1999): Konstruktivistisch-orientierte Lehr-Lernprozessforschung im naturwissenschaftlich-technischen Bereich des Sachunterrichts. In: Köhnlein, W./ Marquardt-Mau, B./ Schreier, H. (Hrsg.): *Forschungen zur Didaktik des Sachunterrichts*, Band 3, S. 125-191. Klinkhardt, Bad Heilbrunn.
 87. MÖLLER, K. (2000): Lehr-Lernprozessforschung im naturwissenschaftlich-technischen Bereich des Sachunterrichts. In: Duit, R. & Rhöneck, C. v. (Hrsg.): *Er-*

- gebnisse fachdidaktischer und psychologischer Lehr-Lern-Forschung. S. 131-156. IPN, Kiel.
88. PRENZEL, M., BAUMERT, J., BLUM, W., LEHMANN, R., LEUTNER, D., NEUBRAND, M., PEKRUN, R., ROST, J. & SCHIEFELE, U. (Hrsg.). (2006). PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres – Zusammenfassung.
89. RENKL, A. (1997): Lernen durch Lehren. Zentrale Wirkmechanismen beim kooperativen Lernen. Deutscher Universitätsverlag, Wiesbaden.
90. RENKL, A. (2000). Automatisierung allein reicht nicht aus: Üben aus kognitionspsychologischer Perspektive. In: R. Meier, U. Rampillon, U. Sandfuchs & L. Stäudel (Hrsg.), Üben und Wiederholen (Jahresheft 2000) (S. 16-19). Friedrich Verlag, Seelze.
91. ROEDER, B. (1974a): Zum Problem der Fehlinterpretation bei gehäufte Anwendung von Signifikanztests - Korrekturdiagramm – Biometrische Zeitschrift 16, 287-288.
92. ROEDER, B. (1974b): Die Konfigurationsfrequenzanalyse (KFA) nach Krauth und Lienert - ein handliches Verfahren zur Verarbeitung sozialwissenschaftlicher Daten, demonstriert an einem Beispiel. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 26, 819-844.
93. ROEDER, B. (1980): Die Anwendung der Regressionstransformation in der schulischen Praxis, insbesondere bei der Überprüfung von Fördermaßnahmen. In: Kanter, G. O. & Masendorf, F. (Hrsg.): Fortschritte Sonderpädagogischer Forschung und Praxis. Band 2 (S. 145-179). Marhold, Berlin.
94. ROEDER, B. (1998): Evaluation computerunterstützten Lernens im Mathematikunterricht. In: Greisbach, M., Kullik, U. Souvignier, E. (Hrsg.): Von der Lernbehindertepädagogik zur Praxis schulischer Lernförderung, S. 251-259. Pabst Science Publishers, Lengerich, Berlin, Düsseldorf, Leipzig, Riga, Scottsdale, Wien, Zagreb.
95. ROEDER, B. (2006): Schülerbefragung. In: Rost, D. (Hrsg.): Handwörterbuch Pädagogische Psychologie. 3. Auflage, S. 637-642. Beltz PVU, Weinheim.

-
96. RUMANN, S. (2005a). Kooperatives Experimentieren im Chemieunterricht. Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik. Dissertation Universität Duisburg-Essen. Online-Publikation:
<http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=12048>
97. RUMANN, S. (2005b). Säuren und Basen. Kooperatives Arbeiten beim Schülerexperiment mit Interaktionsboxen. *Naturwissenschaften im Unterricht/ Chemie*, 88/89, S. 57-59.
98. SCHÖLER, H., MÜLLER, I., SCHEIB, K. & ROOS, J. (November 2004). Selbsteinschätzungen der Lern- und Leistungsmotivation von Drittklässlern. Zur Brauchbarkeit der Skalen zur Erfassung der Lern- und Leistungsmotivation (SELLMO) („EVES“-Arbeitsberichte Nr. 5). Heidelberg: Pädagogische Hochschule, Erziehungs- und Sozialwissenschaftliche Fakultät.
99. SHARAN, S. (1980). Cooperative learning in teams: Recent methods and effects on achievement, attitudes and ethnic relations. *Review of Educational Research*, 50, 241-272.
100. SLAVIN, R. E. (1977). Classroom reward structure: An analytical and practical review. *Review of Educational Research*, 47, 633-650.
101. SLAVIN, R.E. (1993): Kooperatives Lernen und Leistung: Eine empirisch fundierte Theorie. In: G.L. Huber (Hrsg.): *Neue Perspektiven der Kooperation. Ausgewählte Beiträge der Internationalen Konferenz 1992 über Kooperatives Lernen (= Grundlagen der Schulpädagogik, Bd.6; S. 151-170)*. Baltmannsweiler: Schneider-Verlag, Hohengehren.
102. SLAVIN, R. E. (1995): *Cooperative learning: Theory, research and practice*. 2nd ed. Allyn and Bacon, Boston.
103. STEINHAUSEN, D., LANGER, K. (1977): *Clusteranalyse*. De Gruyter, Berlin.
104. STORK, H. (1995): Was bedeuten die aktuellen Forderungen „Schülervorstellungen berücksichtigen, `konstruktivistisch` lehren!“ für den Chemieunterricht in der Sekundarstufe I? *Zeitschrift für Didaktik der Naturwissenschaften*, 1 (1995), 15-28.

-
105. SUMFLETH, E., RUMANN, S. & NICOLAI, N. (2004). Schulische und häusliche Kooperation im Chemieanfangsunterricht, in J. Doll u. a. (Hrsg.): Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung, S. 284-302. Waxmann, Münster.
 106. TEPNER, M. & MELLE, I. (2005): Saubere Sache! Ein Gruppenpuzzle zum Thema Seifen. RAABits Chemie Sekundarstufe I. Raabe, Stuttgart.
 107. TEPNER, M., MELLE, I. & ROEDER, B. (2005a): Methodenevaluation: Effektivität des Gruppenpuzzles im Chemieunterricht der Sek. I. In: A. Pitton. Relevanz fachdidaktischer Forschungsergebnisse für die Lehrerbildung, S. 214-216. LIT, Münster.
 108. TEPNER, M., MELLE, I. & ROEDER, B. (2005b): Gruppenpuzzle und Frontalunterricht im Vergleich. In: Naturwissenschaften im Unterricht Chemie 16, Nr. 88/89, S. 82-85.
 109. TEPNER, M., MELLE, I. & ROEDER, B. (2005c): Gruppenpuzzle und Frontalunterricht im Vergleich. In: Wellensiek, A., Welzel, M., Nohl, T. (Hrsg.): Didaktik der Naturwissenschaften – Quo Vadis? S. 150-157. Logos, Berlin.
 110. TEPNER, M., MELLE, I. & ROEDER, B. (2006): Effektivität des Gruppenpuzzles im Chemieunterricht der Sekundarstufe I. In: A. Pitton (Hg.): Lehren und Lernen mit neuen Medien, S. 275-277. LIT, Münster.
 111. TEPNER, O.: Effektivität von Aufgaben im Chemieunterricht. Dissertation Universität Dortmund, in Vorbereitung.
 112. UNRUH, T. & PETERSEN, S. (2006): Guter Unterricht Praxishandbuch. AOL, Lichtenau.
 113. WALPUSKI, M. (2006): Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback. Studien zum Physik- und Chemielernen, Band 49. Logos, Berlin.
 114. WEIDNER, M. (1998): Durch Gruppenunterricht zur Teamfähigkeit. Kooperative Lernformen an einer Schule zur Erziehungshilfe. In: Praxis Schule 5 – 10, Heft 5, S. 16-24.

115. WEIDNER, M. (2003): Kooperatives Lernen im Unterricht. Das Arbeitsbuch. Kallmeyer, Seelze.
116. WEINERT, F. E. (1998): Eine Lernmethode allein wird nicht genügen. Frankfurter Allgemeine Zeitung Nr. 272, 23. November.
117. WEINERT, F. E. (1999): Die fünf Irrtümer der Schulreformer. Psychologie heute 26, 7, S. 28-34.
118. WILD, K.-P., KRAPP, A. & WINTELER, A. (1992): Die Bedeutung von Lernstrategien zur Erklärung von Studieninteresse auf Lernleistungen. In: Krapp, A., Prenzel, M. (Hrsg.): Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung, Münster.
119. WITTRICK, M. C. (1986). Students' thought processes. In WITTRICK M. C. (Ed.), Handbook of Research on Teaching (3rd ed.). Macmillan, New York.
120. WODZINSKI, R. (2004): Kooperatives Lernen: mehr als nur Gruppenarbeit. Gründe für kooperatives Arbeiten im Physikunterricht. UNTERRICHT PHYSIK 15/2004, Nr. 84, S. 4-7.
121. WOLFF, D. (1997): Lernen lernen. Wege zur Autonomie des Schülers. In: Lernmethoden - Lehrmethoden. Wege zu Selbständigkeit. Friedrich Jahresheft XV, S. 106-108. E. Friedrich, Seelze.
122. WOLLWEBER, K. (2000): Scientific Literacy: Kooperatives Lernen als Lösungsversuch. In: Zur Didaktik der Physik und Chemie, GDPC-Tagung München 1999, Hrsg. R. Brechel, S. 185-187. Leuchtturm-Verlag, Alsbach.
123. ZIMBARDO, P. & GERRIG, R. (1999). Psychologie, 7. Auflage. Springer, Berlin.
124. ZIMMER, M., MELLE I. (2007): Advance Organizer – Einsatz im Chemieunterricht der Sekundarstufe I. In: Naturwissenschaften im Unterricht Chemie 18, Nr. 97, S. 46-48.
125. ZÖFEL, P. (2002): Statistik verstehen. Addison-Wesley/Pearson Education, München.

Internetquellen

ARONSON, E.:

History of the Jigsaw – An Account from Professor Aronson.

<http://www.jigsaw.org/history.htm>

(Stand: 6.4.2007)

ARTELT, C., BAUMERT, J., JULIUS-MCELVANY N. & PESCHAR, J. (2004):

Learners for Life: Thematischer Bericht der OECD zu PISA 2000. Das Lernen lernen: Voraussetzungen für lebensbegleitendes Lernen. Ergebnisse von PISA 2000. Paris: OECD

http://www.mpib-berlin.mpg.de/Pisa/ergebnisse_learners.htm

(Stand: 6.4.2007)

BAUMANN, R. & BLÄTTLER, M. (1994):

Gruppenunterricht - Seifen und Waschmittel.

http://www.educeth.ch/lehrpersonen/chemie/unterrichtsmaterialien_che/speziellethemen/seifen/seifen.pdf

(Stand: 5.4.07)

BENNETT, C., HOWE, C., TRUSWELL, E. (2002):

Small Group Teaching and Learning in Psychology. A review of research in small-group teaching and suggestions for good practice. Report and Evaluation Series No 1, Dec 2002.

http://www.psychology.heacademy.ac.uk/docs/pdf/p20030116_smallgroupsreport.pdf

(Stand: 6.4.2007)

Houghton Mifflin's Project-Based Learning Space:

<http://college.hmco.com/education/pbl/tc/coop.html#3>

(Stand: 6.4.07)

IM BRAHM, G. (2006):

Klassengröße: eine wichtige Variable von Schule und Unterricht? In: bildungsforschung, Jahrgang 3, Ausgabe 1

<http://www.bildungsforschung.org/Archiv/2006-01/klassengroesse/>

(Stand: 26.07.2007)

JOHNSON, D.W., JOHNSON, R.T.:

The Cooperative Learning Institute

<http://www.co-operation.org/>

(Stand: 7.10.06)

KEIMES, D. (2003):

Einsatz und Reflexion ausgewählter Elemente der Methode des "Kooperativen Lernens" in den Jahrgangsstufen 8 und 11.

www.learn-line.nrw.de/angebote/greenline/lernen/downloads/keimes.pdf

(Stand: 11.10.06)

Landesakademie für Fortbildung und Personalentwicklung an Schulen:

Advance Organizer - Der Advance Organizer und nichtlineare Didaktik: Baustein 5

http://lehrerfortbildung-bw.de/unterricht/sol/07_advance_organizer/

(Stand: 11.4.2007)

Uni Dortmund:

ROEDER, B.: Vorlesung Methodenlehre.

<http://seminarserver.fb14.uni-dortmund.de/roeder/Methodologie/MethDrck4.pdf>

(Stand: 3.8.06)

Uni Dortmund:

ROEDER, B.: Auswertungsalgorithmus für Meinungstests und Ratingverfahren.

seminarserver.fb14.uni-dortmund.de/roeder/Methodologie/MessAlgo.pdf (Stand: 1.12.06)

Uni Dortmund:

LIENING, A. (2006): WIDAWIKI – Forschungsprojekt Wirtschaftsdidaktik.

<http://widawiki.wiso.uni-dortmund.de/index.php/Gruppenpuzzle> (Stand: 24.10.06)

Uni Münster:

HELLBERG-RODE, G. (2002): HyperSoil - Entwicklung einer hypermedialen Lern- und Arbeitsumgebung zum Themenfeld "Boden" im (Sach-)Unterricht – Didaktische Legitimation.

<http://hypersoil.uni-muenster.de/2/01/07.htm>

(Stand: 1.12.06)

Uni Oldenburg:

MEYER, H. (2007): Erklärung "Gruppenpuzzle". S. 1-9.

<http://www.member.uni-oldenburg.de/hilbert.meyer/download/Gruppenpuzzle-einzeln.pdf>

(Stand: 4.4.07)

Uni Rostock:

PERLETH, C. (2007): Qualitätsaspekte Psychologischer Diagnostik.

[http://www.phf.uni-](http://www.phf.uni-rostock.de/institut/ipp/lehmaterialien/diagnostikvorlesung/vd1_3_qualitaetsaspekte.htm)

[rostock.de/institut/ipp/lehmaterialien/diagnostikvorlesung/vd1_3_qualitaetsaspekte.htm](http://www.phf.uni-rostock.de/institut/ipp/lehmaterialien/diagnostikvorlesung/vd1_3_qualitaetsaspekte.htm)

(Stand 10.04.07)

Statsoft:

Elementary Concepts in Statistics 2003

[http://www.statsoft.com/textbook/esc.html#What%20is%20"statistical%20significance"%](http://www.statsoft.com/textbook/esc.html#What%20is%20%20statistical%20significance%20(p-level)

[20\(p-level](http://www.statsoft.com/textbook/esc.html#What%20is%20%20statistical%20significance%20(p-level)

(Stand: 3.8.06)

Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien (THILLM, 2002):

Entwicklung von Sozial- und Selbstkompetenz durch kooperatives Lernen. Konzeption für die Umsetzung eines Schwerpunkts der Lehrplanimplementation in Thüringen. Bad Berka. www.thillm.de/thillm/pdf/lehrplan/lp_impl_kompetenz.pdf.

(Stand: 22.03.2007)

12 Anhang

12.1 Verzeichnisse

12.1.1 Abbildungsverzeichnis

Abbildung 1:	Allgemeiner Ablauf des Gruppenpuzzles	22
Abbildung 2:	Expertenmaterial	24
Abbildung 3:	Arbeitsblatt Vermittlungsphase	25
Abbildung 4:	„Einsatzkiste“ mit Schülerversuchen.....	26
Abbildung 5:	Karte mit Untersuchungsorten	31
Abbildung 6:	Übersicht über den Projektverlauf	35
Abbildung 7:	Allgemeines Schema des Untersuchungsablaufs	36
Abbildung 8:	Screepplot für den Wissenstest der Hauptuntersuchung	42
Abbildung 9:	Untersuchungsdesign im Detail.....	44
Abbildung 10:	Layout des Wissenstests	48
Abbildung 11:	Layout des Meinungstests	50
Abbildung 12:	Veranschaulichung der Residuen	56
Abbildung 13:	Häufigkeitsverteilung der Residuen für die Voruntersuchung	74
Abbildung 14:	Häufigkeitsverteilung der Residuen für Hauptuntersuchung I.....	75
Abbildung 15:	Häufigkeitsverteilung der Residuen für Hauptuntersuchung II _{intern}	75
Abbildung 16:	Häufigkeitsverteilung der Residuen für Hauptuntersuchung II _{extern}	76
Abbildung 17:	Ergebnisse des Wissenstests – Scores	77
Abbildung 18:	Selbst- und fremd vermitteltes Wissen im Vergleich	90
Abbildung 19:	Selbst- und fremd vermitteltes Wissen im Vergleich – Scores	91
Abbildung 20:	Vergleich der drei Arten der Wissensvermittlung – Residuen	92
Abbildung 21:	Vergleich der drei Arten der Wissensvermittlung – Scores	94
Abbildung 22:	Einfluss der Chemie-Zeugnisnote auf den Lernzuwachs	98
Abbildung 23:	Wechselwirkungen von Geschlecht und Unterrichtsmethode.....	100
Abbildung 24:	Wechselwirkungen von Schulform und Unterrichtsmethode.....	102
Abbildung 25:	Abhängigkeit des Lernzuwachses von der Übungsdauer	104
Abbildung 26:	Follow-up-Tests: Ergebnisse der Vor- und Hauptuntersuchungen	106

Abbildung 27: Einfluss des Abstandes von Post- und Follow-up-Test auf den Lernzuwachs	108
Abbildung 28: Attraktivität des Faches Chemie – Residuen	111
Abbildung 29: Attraktivität des Faches Chemie – Scores	111
Abbildung 30: Attraktivität der Unterrichtseinheit – Scores	113
Abbildung 31: Attraktivität des Gruppenpuzzles	116
Abbildung 32: Audioaufzeichnungen: Ergebnisse der Konfigurationsfrequenzanalyse	120
Abbildung 33: Audioaufzeichnungen: Häufigkeitsanteile der Aktionen	121
Abbildung 34: Audioaufzeichnungen: Häufigkeitsanteile für lange Rede- bzw. Schweigedauern	122

12.1.2 Tabellenverzeichnis

Tabelle 1:	Vergleich: Kooperatives Lernen – traditioneller Gruppenunterricht.....	9
Tabelle 2:	Beteiligte Schulen und Untersuchungsorte.....	31
Tabelle 3:	Umfang der Stichproben.....	32
Tabelle 4:	Signifikanzniveaus.....	59
Tabelle 5:	Aufteilung der Wissenstest-Items für die Hauptuntersuchungen	62
Tabelle 6:	Vergleich der Unterrichtsmethoden.....	78
Tabelle 7:	Überprüfung der Stichproben auf Normalverteilung.....	79
Tabelle 8:	Klassenweiser Vergleich der Unterrichtsmethoden	87
Tabelle 9:	Selbst-Beigebrachtes vs. Fremdvermitteltes.....	90
Tabelle 10:	Fremdvermittlung: Schüler vs. Lehrer.....	93
Tabelle 11:	Art der Wissensvermittlung: Jonckheere-Terpstra-Trendtest.....	93
Tabelle 12:	Einfluss der Chemienote: Ergebnis der Varianzanalyse.....	98
Tabelle 13:	Korrelation von Chemienote und Residuen.....	99
Tabelle 14:	Einfluss des Geschlechts: Ergebnis der Varianzanalyse	101
Tabelle 15:	Wechselwirkungen zwischen Methode, Geschlecht und Schulform.....	103
Tabelle 16:	Korrelation Übungsdauer – Lernzuwachs	104
Tabelle 17:	Einfluss der Übungsdauer: Ergebnis der Varianzanalyse.....	105
Tabelle 18:	Nachhaltigkeit des erworbenen Wissens: U-Test & Wilcoxon-Test.....	107
Tabelle 19:	Nachhaltigkeit des erworbenen Wissens: Ergebnis der Varianzanalyse ...	108
Tabelle 20:	Attraktivität des Faches Chemie: Ergebnis des U-Tests.....	112
Tabelle 21:	Attraktivität der Unterrichtseinheit: Ergebnis des U-Tests	114
Tabelle 22:	Attraktivität der Unterrichtseinheit: Berücksichtigung von besonderen Rahmenbedingungen	114
Tabelle 23:	Individuelle Schülerbeurteilungen: Ergebnis des U-Tests.....	117
Tabelle 24:	Individuelle Schülerbeurteilungen: Korrelation Scores – Chemienote	118
Tabelle 25:	Individuelle Schülerbeurteilungen: Test auf Normalverteilung	118

12.1.3 Abkürzungsverzeichnis

FU = Frontalunterricht

GP = Gruppenpuzzle

VU = Voruntersuchung

HU = Hauptuntersuchung

MP = Matched-Pairs

Gym = Gymnasium

HS = Hauptschule

BK = Berufskolleg

RS = Realschule

IG = Interventionsgruppe

KG = Kontrollgruppe

12.1.4 Inhaltsverzeichnis Anhang

12.2	MATERIALIEN	156
12.2.1	<i>Gruppenpuzzle Seifen</i>	156
12.2.2	<i>Ergänzendes Material für den Frontalunterricht</i>	179
12.2.3	<i>Wissenstest</i>	183
12.2.4	<i>Meinungstest</i>	191
12.2.5	<i>Individuelle Schülerbeurteilungen</i>	198
12.2.6	<i>Schülerbeobachtungen</i>	199
12.2.7	<i>Stundenprotokoll Frontalunterricht</i>	203
12.2.8	<i>Erfahrungen</i>	204
12.3	STATISTISCHE BERECHNUNGEN WISSENSTEST	206
12.3.1	<i>Linearitätsprüfung</i>	206
12.3.2	<i>Test auf Normalverteilung</i>	207
12.3.3	<i>Reliabilität</i>	209
12.3.4	<i>Effektstärken</i>	210
12.3.5	<i>t-Test & U-Test</i>	211
12.3.6	<i>Art der Wissensvermittlung</i>	221
12.3.7	<i>Wechselwirkung Unterrichtsmethode – Ausgangsleistung</i>	230
12.3.8	<i>Abhängigkeit des Lernerfolgs vom Geschlecht</i>	231
12.3.9	<i>Abhängigkeit des Lernerfolgs von der Schulform</i>	231
12.3.10	<i>Einfluss der Übungsdauer auf den Lernerfolg</i>	232
12.3.11	<i>Nachhaltigkeit des erworbenen Wissens</i>	234
12.4	STATISTISCHE BERECHNUNGEN MEINUNGSTEST	236
12.4.1	<i>Faktorenanalyse Hauptuntersuchung post</i>	236
12.4.2	<i>Einstellung zum Fach Chemie – t-Test, U-Test</i>	237
12.4.3	<i>Einstellung zum Fach Chemie – Effektstärken</i>	237
12.4.4	<i>Einstellung zum Fach Chemie – Reliabilität</i>	238
12.4.5	<i>Attraktivität der Unterrichtseinheit – t-Test, U-Test</i>	239
12.4.6	<i>Attraktivität der Unterrichtseinheit – Effektstärken</i>	239
12.4.7	<i>Attraktivität der Unterrichtseinheit – Reliabilität</i>	240
12.4.8	<i>Attraktivität des Gruppenpuzzles – Reliabilität</i>	241

12.5	STATISTISCHE BERECHNUNGEN INDIVIDUELLE SCHÜLERBEURTEILUNGEN	242
12.5.1	<i>Reliabilität</i>	243
12.5.2	<i>Effektstärken</i>	244
12.5.3	<i>Korrelationen Scores – Chemienote</i>	244
12.5.4	<i>Test auf Normalverteilung</i>	245
12.5.5	<i>U-Test auf Unterschiede in der Mitarbeit</i>	246
12.5.6	<i>U-Test auf Unterschiede im Verhalten</i>	247
12.6	AUDIOAUFZEICHNUNGEN	248

12.2 Materialien

12.2.1 Gruppenpuzzle Seifen



Datum:	Thema 1	Vorname:
Klasse:	Expertenmaterial	Erster Buchstabe des Nachnamens:

Aufbau und Löseverhalten von Seife/Waschwirkung

Experiment:

Bedecke eine Wasseroberfläche mit Hilfe eines Pfefferstreuers leicht mit gemahltem Pfeffer und gib anschließend ein kleines Stück Kernseife in die Mitte der Wasseroberfläche. Beobachte!

Im Folgenden wollen wir dieses kleine Schauspiel erklären. Dazu müssen wir jedoch zunächst einiges über den chemischen Aufbau von Seife wissen:

Das Seifenteilchen:

Ein Seifenteilchen besteht aus einem langen, unpolaren Teil und einem kurzen, polaren Anteil. Das Aussehen ähnelt dem eines Streichholzes (siehe Abb. 1). Der ungeladene, unpolare Schwanz löst sich nicht in polarem Wasser, daher ist dieser Bereich „hydrophob“ = wassermeidend. Der negativ geladene, polare Kopf löst sich hingegen gut in Wasser, er ist „hydrophil“ = wasserliebend. Allein er sorgt dafür, dass sich das gesamte Seifenteilchen im Wasser löst. Merke dir folgende Löslichkeits-Regel:

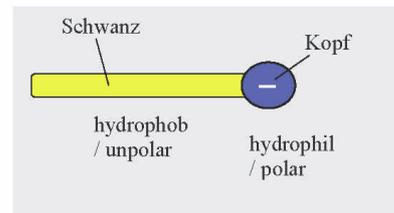


Abb. 1: Streichholzmodell

Polare Stoffe lösen sich in polaren Stoffen, unpolare Stoffe in unpolaren

Verhalten von Seife an der Wasseroberfläche:

Wird Seife in Wasser gegeben, so richtet sich nach dem Streichholzmodell der hydrophile, polare Kopf zum Wasser hin aus, der hydrophobe, unpolare Streichholzstiel wendet sich hingegen vom Wasser ab. Daher stehen die Seifenteilchen senkrecht nebeneinander auf der Wasseroberfläche (siehe Abb. 2). Es bildet sich eine monomolekulare Schicht aus, die viel Platz braucht.

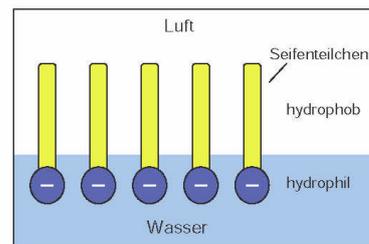


Abb. 2: Seifenteilchen an der Wasseroberfläche

Erklärung der Versuchsbeobachtungen:

Beim Ausbreiten der Seifenteilchen werden die Pfefferkörnchen „beiseite geschoben“, da sich die Kernseife (zunächst einmal) nur an der Wasseroberfläche verteilt.

2

Verhalten von Seife im Wasser:

Erst wenn die Seifenteilchen die gesamte Wasseroberfläche besetzt haben und kein Platz mehr an der Oberfläche ist, bilden sie im Wasser so genannte Mizellen. Das sind Kugeln, bei denen alle unpolaren Schwänze nach innen und alle polaren Streichholzköpfe nach außen zeigen (siehe Abb. 3). Insgesamt ist die Mizelle außen polar und damit in Wasser löslich.

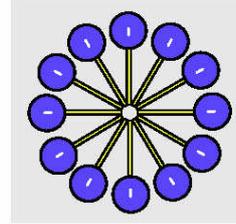


Abb. 3: Vereinfachte Darstellung einer Kugelmizelle

Was macht die Seife mit dem Schmutz?

Um dreckige Wäsche sauber zu waschen, muss der Schmutz im Wasser gelöst werden. Wie ihr sicherlich wisst, ist z. B. Öl nicht in Wasser löslich, sonst könnte man sich ölige Finger auch ganz einfach mit Wasser waschen. Der Grund liegt darin, dass öliger Schmutz unpolar, also hydrophob ist. Und genau deshalb nehmt ihr zum Waschen Seife:

Gemäß der Löslichkeitsregel lagert sich der unpolare Schwanz eines Seifenteilchens an den Schmutz. Gleichzeitig ragt der polare Kopf in das Wasser. Da sich ganz viele Seifenteilchen auf dem Öl bzw. Schmutzteilchen niederlassen, überziehen sie es schließlich mit einer neuen polaren Oberfläche, so dass sich auch unpolare Stoffe wie Schmutz in Wasser lösen (siehe Abb. 4).

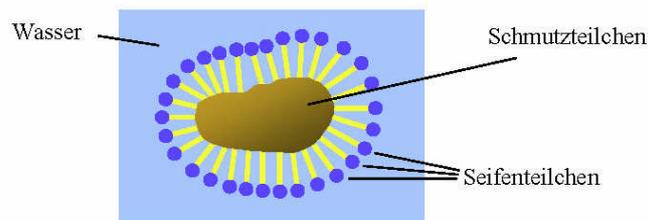


Abb. 4: Gelöster Schmutz in Wasser

Lernziele:

- Du weißt, wie ein Seifenteilchen aufgebaut ist.
- Dir ist klar, was passiert, wenn man ein bisschen Seife in Wasser gibt.
- Du kannst erklären, was Mizellen sind und ab wann sie sich bilden.
- Du weißt, warum sich öliger Schmutz bzw. Öl nicht in Wasser löst.
- Du hast verstanden, warum Schmutz von Seife in Wasser gelöst wird.

Wichtige Begriffe:

polar/unpolar, hydrophob/hydrophil, Streichholzmodell, Kopf, Schwanz, Löslichkeits-Regel, senkrecht, monomolekulare Schicht, viel Platz, Pfefferteilchen werden beiseite geschoben, Mizelle, öliger Schmutz bzw. Öl, neue polare Oberfläche.

Datum:	Thema 2	Vorname:
Klasse:	Expertenmaterial	Erster Buchstabe des Nachnamens:

Herabsetzung der Oberflächenspannung durch Seife

Kleine Mengen Wasser nehmen auf Oberflächen eine nahezu runde Tropfenform an (siehe Abb. 1). Dieses Phänomen kannst du z. B. auf einem Wollpullover oder auf einem frisch polierten Auto beobachten. Der Grund für diese Tatsache ist die Anziehung der Wasserteilchen untereinander, was im Folgenden erläutert wird.



Abb. 1: Wassertropfen auf Wolle

Wiederholung: Theorie zur Wasserstoffbrückenbindung

Wie ihr wisst, besteht Wasser aus Wassermolekülen. Aufgrund der unterschiedlich starken Elektronegativität von Wasserstoff bzw. Sauerstoff tragen die Sauerstoffatome eine negative Teilladung und die Wasserstoffatome eine positive Teilladung (siehe Abb. 2a). Zwischen den einzelnen Wassermolekülen wirken so genannte Wasserstoffbrückenbindungen. Sie sorgen für eine Anziehung der Wassermoleküle untereinander (siehe Abb. 2b).

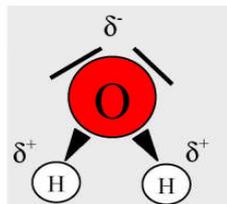


Abb. 2a: Elektronegativitätsunterschiede in einem Wassermolekül

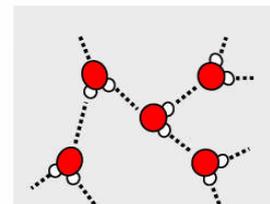


Abb. 2b: Wasserstoffbrückenbindungen (gestrichelte Linien)

Entstehung der Tropfenform: Ein Wassermolekül in der Mitte eines Wassertropfens wird von mehreren Wasserteilchen, die es umgeben, in alle Richtungen angezogen (siehe Abb. 3a). Insgesamt sind diese Anziehungskräfte in alle Richtungen gleich groß, wodurch sie sich gegenseitig aufheben.

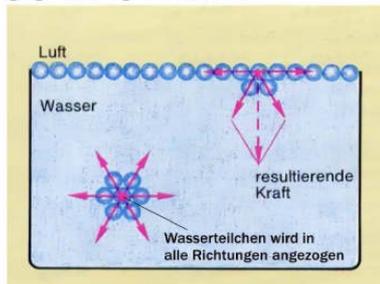


Abb. 3a: Kräfte auf Wassermoleküle an der Oberfläche und innerhalb einer Flüssigkeit

Ein Wasserteilchen in der äußeren Schicht besitzt hingegen keinen oberen Nachbarn. Folglich gleichen sich die Kräfte hier nicht aus und die äußeren Moleküle werden nur von den seitlichen Wasserteilchen und von den Wasserteilchen, die weiter innerhalb des Wassertropfens liegen, angezogen. Da alle äußeren Moleküle gleich stark nach innen gezogen werden, bildet sich die kugelförmige Form eines Tropfens (siehe Abb. 3b). Gleichzeitig rücken die Moleküle an der Oberfläche dichter zusammen, so dass die

Anziehungskräfte hier besonders stark wirken können. Es entsteht die so genannte Oberflächenspannung. Sie sorgt für die extreme Festigkeit der Wasseroberfläche, ähnlich der sich bildenden Haut auf einem frisch gekochten Pudding.

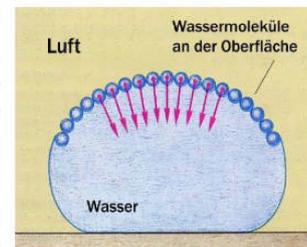


Abb. 3b: Tropfenbildung

2

Herabsetzung der Oberflächenspannung durch Seife:

Befüllt man ein normales Glas über den Rand mit Wasser, so bildet sich ein Wasserberg. Dieser bricht augenblicklich zusammen, sobald Seife hinzugegeben wird.

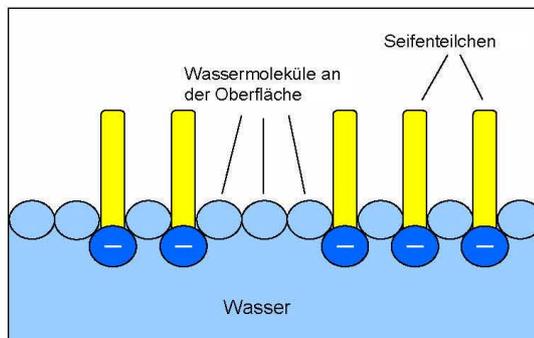


Abb. 4: Seifenteilchen auf einer Wasseroberfläche. Der Schwanz ragt ebenfalls etwas ins Wasser hinein und trennt praktisch die Wasserteilchen an der Oberfläche.

Erklärung: die Seifenteilchen schieben sich zwischen die Wassermoleküle an der Oberfläche des Wassers (siehe Abb. 4) Dadurch verringern sich die Anziehungskräfte (Wasserstoffbrückenbindungen) der Wasserteilchen untereinander, die Oberflächenspannung sinkt. Vergleichen kann man diesen Effekt mit zwei Magneten, die sich weniger stark anziehen, sobald man ihren Abstand vergrößert.

Stoffe (z. B. Seife), welche die Oberflächenspannung herabsetzen, nennt man Tenside. Den Aufbau der Seifenteilchen erklären euch eure Mitschüler noch.

Auswirkung auf das Wäschewaschen:

Eine verringerte Oberflächenspannung des Wassers führt dazu, dass das Wasser nicht mehr „dicke Tropfen“ bildet, sondern insgesamt fließfähiger wird. Dadurch kann es besser in den zu waschenden Stoff eindringen, man spricht von „Benetzen“. Dieses Benetzen ist eine Grundvoraussetzung, damit Wäsche mit Wasser gewaschen werden kann.

Lernziele:

- Du weißt, wie Wasserstoffbrückenbindungen zustande kommen.
- Dir ist klar, warum ein Wassertropfen rund ist!
- Du kannst das Zustandekommen der Oberflächenspannung erklären.
- Du weißt, warum Seife die Oberflächenspannung herabsetzt.
- Du kennst die Bedeutung der Oberflächenspannung für das Wäschewaschen.

Wichtige Begriffe:

Tropfen, Elektronegativität, Teilladung, Wasserstoffbrückenbindung, Anziehung der Teilchen untereinander, Kräfte heben sich auf/heben sich nicht auf, alle äußeren Moleküle nach innen, Moleküle an Oberfläche dichter, Anziehungskräfte besonders stark, Oberflächenspannung, Zwischenschieben der Seifenteilchen, Schwächung der Wasserstoffbrückenbindungen, Verringerung der Oberflächenspannung, Tensid, Benetzen.

Datum:	Thema 3	Vorname:
Klasse:	Expertenmaterial	Erster Buchstabe des Nachnamens:

Was ist Seife und die Nachteile von Seife

Vielleicht hast du das auch schon einmal zu Hause gesehen: Wenn man das Waschbecken nicht regelmäßig reinigt, sieht es nach einiger Zeit richtig schmutzig aus. Es hat sich ein hässlicher grauer Belag gebildet, der das Waschbecken und die Armaturen „stumpf“ erscheinen lässt (siehe Bild). Dieser Belag ist auf das Vorhandensein von hartem Wasser zurückzuführen. Was aber ist „hartes Wasser“?



Der Begriff der Wasserhärte:

Das in der Natur vorkommende Wasser ist nicht „chemisch rein“. Neben Gasen wie Sauerstoff und Kohlenstoffdioxid sind vor allem Calcium und Magnesium darin gelöst. Zur Erinnerung: Calcium und Magnesium liegen in gelöster Form als Ca^{2+} - und Mg^{2+} -Ionen vor. Je nach Menge dieser beiden Ionen bezeichnet man Wasser als weich oder hart. Merke:

Die Wasserhärte ist der Gehalt an Ca^{2+} - und Mg^{2+} -Ionen

Hartes Wasser enthält demnach sehr viele Ca^{2+} - und Mg^{2+} -Ionen. Diese Calcium- bzw. Magnesiumionen reagieren mit der Seife. Um zu verstehen, wie sie reagieren, müssen wir uns einmal klar machen, was Seife eigentlich ist.

Was ist Seife?

Chemisch betrachtet ist Seife ein Salz. Ähnlich wie Kochsalz, NaCl , besteht Seife aus einem positiv geladenen Kation und einem negativ geladenen Anion (siehe Abb. 1).

Dabei gibt es zwei verschiedene Kationen: Natrium und Kalium. Natrium-Seifen nennt man auch „Kernseife“, die ihr vielleicht zu Hause auch zum Händewaschen benutzt. Kalium-Seifen bezeichnet man auch als „Schmierseife“. Sie ist flüssig und wird heute - wenn auch seltener - z. B. zum Bodenwischen verwendet.

Der waschaktive Teil einer Seife aber ist das Seifenanion. Dieses bezeichnet man auch als Fettsäureanion. Fettsäuren sind wichtige Bestandteile von Fetten. Sicherlich hast du schon von so genannten „ungesättigten Fettsäuren“ auf einer Margarine-Packung gelesen. Früher hat man Seife aus Fett hergestellt. Wie genau eine Fettsäure aufgebaut ist, interessiert uns erst einmal nicht.

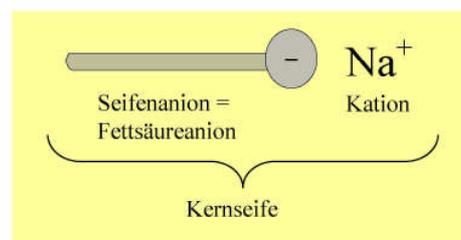


Abb. 1: Aufbau von Seife

Datum:	<u>Thema 1</u>	Vorname:
Klasse:	Arbeitsblatt Erarbeitungsphase	Erster Buchstabe des Nachnamens:

Aufbau und Löseverhalten von Seife/Waschwirkung

Zur Erinnerung:

In der Erarbeitungsphase soll sich jeder Schüler selbstständig und alleine in sein Thema einarbeiten. Dafür hast du ungefähr 20 min Zeit. Ziel ist es, am Ende dieser Phase Experte für das entsprechende Thema zu sein. Zur Überprüfung, ob du schon Experte bist, gibt es eine kurze Lernkontrolle. Kannst du alle Fragen fehlerfrei beantworten, bist du fit für die Expertenrunde!

Arbeitsaufträge:

1. Lies den Text einmal und führe am Anfang den Pfefferversuch durch.
2. Sieh dir die Lernziele an.
3. Lies den Text ein zweites Mal durch, wobei du wichtige Wörter markierst/herausschreibst. Behalte dabei deine Lernziele im Hinterkopf.
4. Beantworte die untenstehenden Fragen. Bitte schriftlich!

Lernkontrolle:

1. Warum wird Pfeffer auf einer Wasseroberfläche an den Rand gedrängt, sobald Seife hinzu gegeben wird?
2. Was passiert mit den Seifenteilchen, wenn kein Platz mehr auf der Wasseroberfläche ist?
3. Was besagt die Löslichkeits-Regel?
4. Beschreibe die Waschwirkung von Seife!

Viel Erfolg!

Datum:	<u>Thema 2</u>	Vorname:
Klasse:	Arbeitsblatt Erarbeitungsphase	Erster Buchstabe des Nachnamens:

Herabsetzung der Oberflächenspannung durch Seife

Zur Erinnerung:

In der Erarbeitungsphase soll sich jeder Schüler selbstständig und alleine in sein Thema einarbeiten. Dafür hast du ungefähr 20 min Zeit. Ziel ist es, am Ende dieser Phase Experte für das entsprechende Thema zu sein. Zur Überprüfung, ob du schon Experte bist, gibt es eine kurze Lernkontrolle. Kannst du alle Fragen fehlerfrei beantworten, bist du fit für die Expertenrunde!

Arbeitsaufträge:

1. Lies den Text einmal durch.
2. Sieh dir die Lernziele an.
3. Lies den Text ein zweites Mal durch, wobei du wichtige Wörter markierst/herausschreibst. Behalte dabei deine Lernziele im Hinterkopf.
4. Beantworte die untenstehenden Fragen. Bitte schriftlich!

Lernkontrolle:

1. Welche Anziehungskräfte wirken zwischen den Wassermolekülen und wie kommen sie zustande?
2. Erkläre, warum ein Wassertropfen rund ist!
3. Erkläre das Zustandekommen der Oberflächenspannung!
4. Warum verringert Seife die Oberflächenspannung von Wasser?
5. Warum muss zum Wäschewaschen die Oberflächenspannung des Wassers verringert werden?

Viel Erfolg!

Datum:	<u>Thema 3</u>	Vorname:
Klasse:	Arbeitsblatt Erarbeitungsphase	Erster Buchstabe des Nachnamens:

Was ist Seife und die Nachteile von Seife

Zur Erinnerung:

In der Erarbeitungsphase soll sich jeder Schüler selbstständig und alleine in sein Thema einarbeiten. Dafür hast du ungefähr 20 min Zeit. Ziel ist es, am Ende dieser Phase Experte für das entsprechende Thema zu sein. Zur Überprüfung, ob du schon Experte bist, gibt es eine kurze Lernkontrolle. Kannst du alle Fragen fehlerfrei beantworten, bist du fit für die Expertenrunde!

Arbeitsaufträge:

1. Lies den Text einmal durch.
2. Sieh dir die Lernziele an.
3. Lies den Text ein zweites Mal durch, wobei du wichtige Wörter markierst/herausschreibst. Behalte dabei deine Lernziele im Hinterkopf.
4. Beantworte die untenstehenden Fragen. Bitte schriftlich!

Lernkontrolle:

1. Wie kommt der graue Belag in nicht geputzten Waschbecken zustande?
2. Was versteht man unter hartem Wasser?
3. Was ist Kern- bzw. Schmierseife?
4. Warum ist die Verwendung von Seife bei hartem Wasser nachteilig? (2 Gründe)

Viel Erfolg!

Datum:	<u>Thema 1</u>	Vorname:
Klasse:	Arbeitsblatt Expertenphase	Erster Buchstabe des Nachnamens:

Aufbau und Löseverhalten von Seife/Waschwirkung

Zur Erinnerung:

In der Expertenphase setzen sich alle Schüler, die das gleiche Thema erarbeitet haben, in einer neuen Gruppe zusammen. Ziel ist die Vorbereitung auf eure Vermittlerrolle in der nächsten Stunde.

Arbeitsaufträge:

1. Vergleicht die Antworten eurer Lernkontrolle und findet eine gemeinsame Lösung.
2. Jeder Schüler erklärt seinen Mitschülern einmal sein Thema. Die anderen achten darauf, dass alle wichtigen Begriffe verwendet werden und die Erklärungen vollständig und fehlerfrei sind. Versucht, möglichst frei zu reden!

Falls ihr in dieser Stunde nicht fertig werdet, macht in der nächsten Stunde da weiter, wo ihr aufgehört habt.

Hausaufgabe:

1. Lest das Arbeitsblatt Vermittlungsphase durch und spielt den Versuch in Gedanken durch. Was erwartet ihr?
2. Überlegt euch, was ihr in der Vermittlungsphase euren Mitschülern erzählt. Achtet darauf, dass alle Lernziele und wichtigen Begriffe verwendet werden. Macht euch stichpunktartig eine Gliederung für euren Vortrag (halbe DIN A4 Seite). Das Expertenmaterial kann euch als Vorlage dienen. Denkt daran: Nur ihr selbst seid für euren Lernerfolg verantwortlich, also bereitet euch gut vor!

Datum:	<u>Thema 2</u>	Vorname:
Klasse:	Arbeitsblatt Expertenphase	Erster Buchstabe des Nachnamens:

Herabsetzung der Oberflächenspannung durch Seife

Zur Erinnerung:

In der Expertenphase setzen sich alle Schüler, die das gleiche Thema erarbeitet haben, in einer neuen Gruppe zusammen. Ziel ist die Vorbereitung auf eure Vermittlerrolle in der nächsten Stunde.

Arbeitsaufträge:

1. Vergleicht die Antworten eurer Lernkontrolle und findet eine gemeinsame Lösung.
2. Jeder Schüler erklärt seinen Mitschülern einmal sein Thema. Die anderen achten darauf, dass alle wichtigen Begriffe verwendet werden und die Erklärungen vollständig und fehlerfrei sind. Versucht, möglichst frei zu reden!

Falls ihr in dieser Stunde nicht fertig werdet, macht in der nächsten Stunde da weiter, wo ihr aufgehört habt.

Hausaufgabe:

1. Lest das Arbeitsblatt Vermittlungsphase durch und spielt den Versuch in Gedanken durch. Was erwartet ihr?
2. Überlegt euch, was ihr in der Vermittlungsphase euren Mitschülern erzählt. Achtet darauf, dass alle Lernziele und wichtigen Begriffe verwendet werden. Macht euch stichpunktartig eine Gliederung für euren Vortrag (halbe DIN A4 Seite). Das Expertenmaterial kann euch als Vorlage dienen. Denkt daran: Nur ihr selbst seid für euren Lernerfolg verantwortlich, also bereitet euch gut vor!

Datum:	<u>Thema 3</u>	Vorname:
Klasse:	Arbeitsblatt Expertenphase	Erster Buchstabe des Nachnamens:

Was ist Seife und die Nachteile von Seife

Zur Erinnerung:

In der Expertenphase setzen sich alle Schüler, die das gleiche Thema erarbeitet haben, in einer neuen Gruppe zusammen. Ziel ist die Vorbereitung auf eure Vermittlerrolle in der nächsten Stunde.

Arbeitsaufträge:

1. Vergleicht die Antworten eurer Lernkontrolle und findet eine gemeinsame Lösung.
2. Jeder Schüler erklärt seinen Mitschülern einmal sein Thema. Die anderen achten darauf, dass alle wichtigen Begriffe verwendet werden und die Erklärungen vollständig und fehlerfrei sind. Versucht, möglichst frei zu reden!

Falls ihr in dieser Stunde nicht fertig werdet, macht in der nächsten Stunde da weiter, wo ihr aufgehört habt.

Hausaufgabe:

1. Lest das Arbeitsblatt Vermittlungsphase durch und spielt den Versuch in Gedanken durch. Was erwartet ihr?
2. Überlegt euch, was ihr in der Vermittlungsphase euren Mitschülern erzählt. Achtet darauf, dass alle Lernziele und wichtigen Begriffe verwendet werden. Macht euch stichpunktartig eine Gliederung für euren Vortrag (halbe DIN A4 Seite). Das Expertenmaterial kann euch als Vorlage dienen. Denkt daran: Nur ihr selbst seid für euren Lernerfolg verantwortlich, also bereitet euch gut vor!

Datum:	<u>Thema 1</u>	Vorname:
Klasse:	Arbeitsblatt Vermittlungsphase	Erster Buchstabe des Nachnamens:

Aufbau und Löseverhalten von Seife/Waschwirkung

Geht in eure alten Gruppen zurück. In der Vermittlungsphase erklärt jeder Schüler seinen Mitschülern das von ihm erarbeitete Thema. Bei deinen Erklärungen sollst du nicht nur die unten aufgeführten Lernziele berücksichtigen, sondern auch folgende Begriffe verwenden. Benutze auch Stift und Papier zur Erklärung. Sei dir bewusst, dass nur du für den Lernerfolg der anderen verantwortlich bist und die anderen für deinen. Am Ende der Vermittlungsphase sollen alle Schülerinnen und Schüler alle Lernziele sowie die durchgeführten Versuche verstanden haben.

Lernziele:

- Ihr wisst, wie ein Seifenteilchen aufgebaut ist.
- Euch ist klar, was passiert, wenn man ein bisschen Seife in Wasser gibt.
- Ihr könnt erklären, was Mizellen sind und ab wann sie sich bilden.
- Ihr wisst, warum sich öliger Schmutz bzw. Öl nicht in Wasser löst.
- Ihr habt verstanden, warum Schmutz von Seife in Wasser gelöst wird.

Wichtige Begriffe:

polar/unpolar, hydrophob/hydrophil, Streichholzmodell, Kopf, Schwanz, Löslichkeits-Regel, senkrecht, monomolekulare Schicht, viel Platz, Pfefferteilchen werden beiseite geschoben, Mizelle, öliger Schmutz bzw. Öl, neue polare Oberfläche.

Arbeitsaufträge:

1. Erklärt euch gegenseitig in ca. 15 min eure Themen. Verwendet dabei die wichtigen Begriffe und achtet auf die Lernziele. Es ist sinnvoll, die Erklärungen der Mitschüler schriftlich festzuhalten.
2. Führt gemeinsam den Pfeffer-Versuch und folgendes Experiment durch und beantwortet die dazugehörigen Fragen.

Experiment:

In zwei Reagenzgläsern befindet sich jeweils ein Öl/Wasser-Gemisch. In das leerere Reagenzglas gibst du soviel Seifenlösung hinzu, bis fast keine Luft mehr drin ist. Verschließe mit einem Stopfen, schüttele beide kurz (Daumen auf Stopfen!) und lasse stehen.

Fragen:

- Welche Beobachtungen erwartest du?
- Wie kannst du dieses Verhalten erklären?
- Würde sich Seife auch in Öl (ohne Wasser!) lösen? Begründe! Falls ja: Wie würden die Mizellen dann aussehen?

Datum:	<u>Thema 2</u>	Vorname:
Klasse:	Arbeitsblatt Vermittlungsphase	Erster Buchstabe des Nachnamens:

Herabsetzung der Oberflächenspannung durch Seife

Geht in eure alten Gruppen zurück. In der Vermittlungsphase erklärt jeder Schüler seinen Mitschülern das von ihm erarbeitete Thema. Bei deinen Erklärungen sollst du nicht nur die unten aufgeführten Lernziele berücksichtigen, sondern auch folgende Begriffe verwenden. Benutze auch Stift und Papier zur Erklärung. Sei dir bewusst, dass nur du für den Lernerfolg der anderen verantwortlich bist und die anderen für deinen. Am Ende der Vermittlungsphase sollen alle Schülerinnen und Schüler alle Lernziele sowie die durchgeführten Versuche verstanden haben.

Lernziele:

- Ihr wisst, wie Wasserstoffbrückenbindungen zustande kommen.
- Euch ist klar, warum ein Wassertropfen rund ist!
- Ihr könnt das Zustandekommen der Oberflächenspannung erklären.
- Ihr wisst, warum Seife die Oberflächenspannung herabsetzt.
- Ihr kennt die Bedeutung der Oberflächenspannung für das Wäschewaschen.

Wichtige Begriffe:

Tropfen, Elektronegativität, Teilladung, Wasserstoffbrückenbindung, Anziehung der Teilchen untereinander, Kräfte heben sich auf/heben sich nicht auf, alle äußeren Moleküle nach innen, Moleküle an Oberfläche dichter, Anziehungskräfte besonders stark, Oberflächenspannung, Zwischenschieben der Seifenteilchen, Schwächung der Wasserstoffbrückenbindungen, Verringerung der Oberflächenspannung, Tensid, Benetzen.

Arbeitsaufträge:

1. Erklärt euch gegenseitig in ca. 15 min eure Themen. Verwendet dabei die wichtigen Begriffe und achtet auf die Lernziele. Es ist sinnvoll, die Erklärungen der Mitschüler schriftlich festzuhalten.
2. Führt gemeinsam folgendes Experiment durch und beantwortet die dazugehörigen Fragen.

Experiment:

Fülle ein Filmdöschen (Standzylinder) bis zum Rand mit Leitungswasser. Gib mit der Spritzflasche behutsam Wasser zu, bis sich ein deutlich sichtbarer Wasserberg gebildet hat (Abb. 1). Dann lege vorsichtig ein kleines Stückchen Kernseife auf die Wasseroberfläche und warte einen Augenblick.

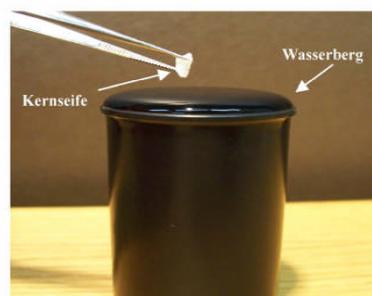


Abb. 1: Filmdöschen mit Wasserberg

Fragen:

- Welche Beobachtungen erwartest du?
- Wie kannst du dieses Verhalten erklären?
- Öl besitzt ebenfalls eine Oberflächenspannung (Ursache sind hier andere zwischenmolekulare Anziehungskräfte als die H-Brücken beim Wasser). Was würde passieren, wenn man im Versuch Öl statt Wasser verwendet? Begründe! (Der Aufbau eines Seifenteilchens spielt hier eine Rolle, siehe Thema 1)

Datum:	<u>Thema 3</u>	Vorname:
Klasse:	Arbeitsblatt Vermittlungsphase	Erster Buchstabe des Nachnamens:

Was ist Seife und die Nachteile von Seife

Geht in eure alten Gruppen zurück. In der Vermittlungsphase erklärt jeder Schüler seinen Mitschülern das von ihm erarbeitete Thema. Bei deinen Erklärungen sollst du nicht nur die unten aufgeführten Lernziele berücksichtigen, sondern auch folgende Begriffe verwenden. Benutze auch Stift und Papier zur Erklärung. Sei dir bewusst, dass nur du für den Lernerfolg der anderen verantwortlich bist und die anderen für deinen. Am Ende der Vermittlungsphase sollen alle Schülerinnen und Schüler alle Lernziele sowie die durchgeführten Versuche verstanden haben.

Lernziele:

- Ihr wisst, was man unter Wasserhärte versteht.
- Ihr kennt den waschaktiven Teil der Seife.
- Ihr habt verstanden, dass Seifen aus dem Anion einer Fettsäure und einem Kation bestehen.
- Ihr wisst, was Kern- bzw. Schmierseife ist.
- Ihr könnt erklären, wie Seife mit Ca^{2+} -Ionen reagiert (mit obiger Reaktionsgleichung!).
- Euch sind die beiden Nachteile von Seife klar.

Wichtige Begriffe:

grauer Belag, Wasserhärte, Ca^{2+} - und Mg^{2+} -Ionen, Salz, Fettsäureanion/Seifenanion, Natrium/Kalium-Salz einer Fettsäure, Kernseife, Schmierseife, Calcium-Ionen, Kalkseife, Wäsche wird grau, erhöhter Seifenverbrauch.

Arbeitsaufträge:

1. Erklärt euch gegenseitig in ca. 15 min eure Themen. Verwendet dabei die wichtigen Begriffe und achtet auf die Lernziele. Es ist sinnvoll, die Erklärungen der Mitschüler schriftlich festzuhalten.
2. Vergleicht den Aufbau von Seife mit dem Aufbau der Seifenteilchen aus Thema 1. Wo gibt es Gemeinsamkeiten/Unterschiede?
3. Führt gemeinsam folgendes Experiment durch und beantwortet die dazugehörigen Fragen.

Experiment:

Gib zu den gleichen Mengen

1. dest. Wasser
2. Leitungswasser
3. mit Calciumionen angereichertes Wasser

jeweils fünf Tropfen Seifenlösung. Schwenke kurz hin und her und beobachte. Dann schüttele kräftig und beobachte.

Fragen:

- Welche Beobachtungen erwartest du?
- Wie kannst du dieses Verhalten erklären?
- Was würde passieren, wenn statt Calcium- bzw. Magnesiumionen Natriumionen (gleiche Menge) im Wasser enthalten wären? Würde auch etwas ausfallen? (Denk an den Aufbau von Kernseife!)

Datum:	Protokollbogen Expertenphase	Klasse:
--------	-------------------------------------	---------

Folgende „wichtige Begriffe“ wurden von deinen Mitschülern in ihrem Vortrag erwähnt (bitte Anfangsbuchstaben des jeweiligen Schülers angeben):

Thema 1

Begriff	Schüler
polar/unpolar	
hydrophob/hydrophil	
Streichholzmodell	
Kopf	
Schwanz	
Löslichkeits-Regel	
senkrecht	
monomolekulare Schicht	
viel Platz	
Pfefferteilchen werden beiseite geschoben	
Mizelle	
öliger Schmutz bzw. Öl	
neue polare Oberfläche	

Sage deinen Mitschülern, welcher Begriff eventuell gefehlt hat und lasse diesen erklären.

Datum:	Protokollbogen Expertenphase	Klasse:
--------	-------------------------------------	---------

Folgende „wichtige Begriffe“ wurden von deinen Mitschülern in ihrem Vortrag erwähnt (bitte Anfangsbuchstaben des jeweiligen Schülers angeben):

Thema 2

Begriff	Schüler
Tropfen	
Elektronegativität	
Teilladung	
Wasserstoffbrückenbindung	
Anziehung der Teilchen untereinander	
Kräfte heben sich auf/heben sich nicht auf	
alle äußeren Moleküle nach innen	
Moleküle an Oberfläche dichter	
Anziehungskräfte besonders stark	
Oberflächenspannung	
Zwischenschieben der Seifenteilchen	
Schwächung der Wasserstoffbrückenbindungen	
Verringerung der Oberflächenspannung	
Tensid	
Benetzen	

Sage deinen Mitschülern, welcher Begriff eventuell gefehlt hat und lasse diesen erklären.

Datum:	Protokollbogen Expertenphase	Klasse:
--------	-------------------------------------	---------

Folgende „wichtige Begriffe“ wurden von deinen Mitschülern in ihrem Vortrag erwähnt (bitte Anfangsbuchstaben des jeweiligen Schülers angeben):

Thema 3

Begriff	Schüler
grauer Belag	
Wasserhärte	
Ca ²⁺ - und Mg ²⁺ -Ionen	
Salz	
Fettsäure-/Seifenanion	
Natrium/Kaliumsalz einer Fettsäure	
Kernseife	
Schmierseife	
Calciumionen	
Kalkseife	
Wäsche wird grau	
erhöhter Seifenverbrauch	

Sage deinen Mitschülern, welcher Begriff eventuell gefehlt hat und lasse diesen erklären.

Materialliste für GP-Versuche (6 Gruppen)

Pfefferversuch Thema 1

- 6 Pfefferstreuer mit schwarzem Pfeffer
- 6 Kunststoff-Petrischalen
- 6 Filmdöschen mit Kernseifestückchen

Emulgierversuch Thema I

- 6 x 2 Reagenzgläser
- 12 RG-Stopfen
- 6 RG-Ständer
- 6 Fläschchen mit Seifenlösung
- 6 Pipetten für Seifenlösung

Vorbereiten: In zwei Reagenzgläser jeweils 2 ml Öl geben, ein RG bis oben mit dest. Wasser befüllen, in anderem RG 2 ml Platz für Seifenlösung lassen.

Wasserberg-Versuch Thema II

- 6 Filmdöschen (alternativ: kleine Standzylinder)
- 6 Spritzflaschen mit (dest.) - H₂O
- kleine Stückchen Kernseife

Kalkseife-Versuch Thema III

- 6 x 3 Reagenzgläser
- 6 RG-Ständer
- 6 Fläschchen mit Seifenlösung
- 6 Pipetten für Seifenlösung

**Vorbereiten: jeweils 3 RG ca. 3 cm hoch befüllen mit:
a.) dest. Wasser b.) Leitungswasser c.) Ca²⁺-Wasser.**

Für alle:

- Große Kunststoffflasche mit Seifenlösung (0,5 %, mörsern, warm lösen und abkühlen lassen. Nicht höherprozentig, da sonst zu viskös wegen Kolloidbildung)
- Große Kunststoffflasche mit Olivenöl, Pipette
- Große Kunststoffflasche mit CaCl₂-Lösung (0,1 mol/l, ca. 2,8 g auf 250 ml) für Simulation von Leitungswasser (falls Leitungswasser zu weich)
- Große Kunststoffflasche mit CaCl₂-Lösung (4 %) für Simulation von hartem Wasser
- Spritzflasche mit dest. Wasser
- Leitungswasser

Legende:

Bereits für anderen Versuch vorhanden

Erklärung des Gruppenpuzzles

Begrüßen, Namen anschreiben, heute beginnt euer Teil der Methoden-Untersuchung zum Thema Seife. Ich erzähl euch jetzt fünf Minuten, was wir machen, danach bitte ich einen von euch, das noch einmal kurz zusammenzufassen. Die Methode heißt Gruppenpuzzle → lieber Expertenunterricht, da ihr bei dieser Methode zu Experten werdet. Das funzt folgendermaßen:

Zuerst setzt ihr euch in so genannten **Stammgruppen** zusammen, die ihr selbst bestimmen könnt. Jede Stammgruppe besteht aus drei Leuten, wovon **jeder ein anderes Thema erhält (Folie)**. Jeder Schüler soll sich selbstständig und alleine in sein Thema einarbeiten, man nennt diese Phase deshalb auch **Erarbeitungsphase**. Dafür habt ihr ungefähr 20 min Zeit. Ziel ist es, am Ende dieser Phase Experte für das entsprechende Thema zu sein. Zur Überprüfung, ob ihr schon Experten seid, gibt es eine kurze Lernkontrolle, die sich auf dem angehefteten Arbeitsblatt befindet.

Für die nächste Phase, die **Expertenphase**, setzen sich alle Schüler, die das gleiche Thema erarbeitet haben, in einer neuen Gruppe, der so genannten Expertengruppe, zusammen. **(Folie) Klärt offene Fragen.** 20min. Es ist ganz wichtig, dass hier alle Unklarheiten beseitigt werden, da ihr mit diesem Wissen die anderen später unterrichten sollt. Wenn fertig, erklärt jeder Schüler seinen Mitschülern einmal sein Thema. Die anderen achten darauf, dass alle wichtigen Begriffe verwendet werden.

Vermittlungsphase: 30 min 2. Stunde + 45 min 3. Stunde. Evtl. ca. 10 min in 4. Stunde + Test!
Geht in eure alten Stammgruppen zurück. In der Vermittlungsphase erklärt jeder Schüler seinen Mitschülern das von ihm erarbeitete Thema. **Bei euren Erklärungen sollt ihr nicht nur die unten aufgeführten Lernziele berücksichtigen, sondern auch die aufgeführten wichtigen Begriffe.** Benutzt - wenn nötig- Stift und Papier zur Erklärung. Am Ende der Vermittlungsphase sollen alle Schülerinnen und Schüler alle Lernziele sowie die durchgeführten Versuche verstanden haben. **Die Zettel der anderen bekommt ihr nicht** → Seid euch bewusst, dass nur ihr selbst für den Lernerfolg der anderen verantwortlich seid und die anderen für euren. Ihr übernehmt komplett die Lehrerrolle. Falls ihr glaubt, dass euch etwas nicht gut von euren Mitschülern erklärt worden ist: nachhaken. Falls ihr etwas nicht verstanden habt: nachfragen. Und zwar so lange, bis ihr es verstanden habt. Benutzt Hände/Füße, Stifte, Bilder, Zeichnungen zum Erklären. **Es ist sinnvoll, die Erklärungen der anderen schriftlich festzuhalten, da ihr die Expertenmaterialien der anderen nicht bekommt. Nur mit den Erklärungen der anderen schreiben wir darüber einen Test.**

Wiederholen lassen

Stammgruppen bilden lassen (beide Geschlechter müssen vertreten sein!)

Das Expertenmaterial wird am Anfang der nächsten Stunde ausgegeben.

Bitte euer Material nicht den anderen geben → Untersuchung kaputt.

Saubere Sache – ein Gruppenpuzzle zum Thema Seifen

- Ablauf und Aufbau -

- Beispiel für 14 Schüler
- Dauer: 3 x 45 min (plus Einführung in der Stunde vorher)
- Themen:

1

2

3

- In Grau jeweils doppelt besetzte Themen

Phase	Gruppentyp	Gruppeneinteilung																														
Erarbeitungsphase ca. 30 min	Stammgruppe à 3 SuS	<table style="margin-left: auto; margin-right: auto;"> <tr><td>1</td><td>2</td><td>1</td><td>2</td></tr> <tr><td>3</td><td>1</td><td>3</td><td>2</td></tr> <tr><td>1</td><td>2</td><td>1</td><td>2</td></tr> <tr><td>3</td><td></td><td>3</td><td></td></tr> </table>	1	2	1	2	3	1	3	2	1	2	1	2	3		3															
1	2	1	2																													
3	1	3	2																													
1	2	1	2																													
3		3																														
Expertenphase ca. 2 x 15 min	Expertengruppe (max. 5 SuS)	<table style="margin-left: auto; margin-right: auto;"> <tr><td>1</td><td>1</td><td></td><td></td><td>3</td><td>3</td></tr> <tr><td>1</td><td>1</td><td></td><td></td><td>3</td><td>3</td></tr> <tr><td>1</td><td></td><td>2</td><td>2</td><td></td><td></td></tr> <tr><td></td><td></td><td>2</td><td>2</td><td></td><td></td></tr> <tr><td></td><td></td><td>2</td><td></td><td></td><td></td></tr> </table>	1	1			3	3	1	1			3	3	1		2	2					2	2					2			
1	1			3	3																											
1	1			3	3																											
1		2	2																													
		2	2																													
		2																														
Vermittlungsphase ca. 30 + 45 min	Stammgruppe	<table style="margin-left: auto; margin-right: auto;"> <tr><td>1</td><td>2</td><td>1</td><td>2</td></tr> <tr><td>3</td><td>1</td><td>3</td><td>2</td></tr> <tr><td>1</td><td>2</td><td>1</td><td>2</td></tr> <tr><td>3</td><td></td><td>3</td><td></td></tr> </table>	1	2	1	2	3	1	3	2	1	2	1	2	3		3															
1	2	1	2																													
3	1	3	2																													
1	2	1	2																													
3		3																														

Erklärungen während des Gruppenpuzzles

jeweils vor der entsprechenden Phase

Alle Arbeitsblätter werden vor der entsprechenden Phase ausgehändigt, das AB Vermittlungsphase am Ende der ersten Stunde, da es als HA durchzulesen ist.

Erarbeitungsphase (ca. 30 min) (Nur Pfefferversuch Thema I)

Bildung von Stammgruppen. Jeder Schüler soll sich selbstständig und in Einzelarbeit in sein Thema einarbeiten. Ziel ist es, am Ende dieser Phase Experte für das entsprechende Thema zu sein. Zur Überprüfung, ob die SuS schon Experten sind, gibt es eine kurze Lernkontrolle, die sich auf dem angehefteten Arbeitsblatt befindet.

Der Lehrer darf nur Fragen zum Text bzw. Ablauf gefragt werden. Fragen zur jeweiligen Thematik sollen in der Expertenrunde geklärt werden.

Das Expertenmaterial bitte nicht den Mitschülern geben → nicht im Sinne der Methode!

Expertenphase (15 min + 15 min am Anfang der 2. GP-Stunde, alle Versuche vorbereiten, siehe Materialliste für GP-Versuche)

Alle Schüler, die das gleiche Thema erarbeitet haben, setzen sich in einer neuen Gruppe, der so genannten Expertengruppe, zusammen und klären offene Fragen.

Es ist ganz wichtig, dass hier alle Unklarheiten beseitigt werden, da mit diesem Wissen die anderen später unterrichten werden sollen!

Wenn alle Fragen geklärt sind, erklärt jeder Schüler seinen Mitschülern einmal sein Thema. Die Zuhörer achten darauf, dass alle wichtigen Begriffe verwendet werden und ob die Erklärungen vollständig und fehlerfrei sind. Das Bildmaterial auf den Materialien darf bzw. soll zu Hilfe genommen werden. Nicht nur vorlesen sondern möglichst auswendig erklären. Einer aus der Gruppe erhält einen **Protokollbogen, auf dem er abhakt, ob alle wichtigen Begriffe im Erklärungs Vortrag erwähnt wurden.** Der Protokollführer wird aus der Gruppe gewählt.

→ HA: Struktur für Vortrag machen

Vermittlungsphase 30 min 2. Stunde + 45 min 3. Stunde. 4. Stunde: Test!

Zurückgehen in alte Stammgruppen. In der Vermittlungsphase erklärt jeder Schüler seinen Mitschülern das von ihm erarbeitete Thema. Bei den Erklärungen sollen nicht nur die unten aufgeführten Lernziele berücksichtigt werden, sondern auch die wichtigen Begriffe. Wenn nötig Stift und Papier zur Erklärung benutzen! Am Ende der Vermittlungsphase sollen alle Schülerinnen und Schüler alle Lernziele sowie die durchgeführten Versuche verstanden haben.

Seid euch bewusst, dass nur ihr selbst für den Lernerfolg der anderen verantwortlich seid und die anderen für euren. Ihr übernehmt komplett die Lehrerrolle.

Gleichzeitig sind die Zuhörer dafür verantwortlich, dass sie auch die Informationen einfordern: Falls etwas nicht gut von den Mitschülern erklärt worden ist: nachhaken. Falls etwas nicht verstanden wurde: nachfragen. Und zwar so lange, bis alle es verstanden haben.

Benutzt Hände/Füße, Stifte, Zeichnungen zum Erklären. Besonders die **Bilder des Expertenmaterials** sollen bei den Erklärungen **verwendet** werden!

Themen nacheinander behandeln, Versuche immer zum Thema. Es ist sinnvoll, die **Erklärungen** der Mitschüler **schriftlich festzuhalten**, da die Expertenmaterialien der anderen nicht verteilt werden und über alle Themen in der **vierten Stunde** ein **Test** geschrieben wird!

Das Expertenmaterial bitte nicht den Mitschülern geben → nicht im Sinne der Methode!

12.2.2 Ergänzendes Material für den Frontalunterricht

Wichtige Begriffe für die Lehrer

Wichtige Begriffe Thema 1

polar/unpolar, hydrophob/hydrophil, Streichholzmodell, Kopf, Schwanz, Löslichkeits-Regel, senkrecht, monomolekulare Schicht, viel Platz, Pfefferteilchen werden beiseite geschoben, Mizelle, öliger Schmutz bzw. Öl, neue polare Oberfläche.

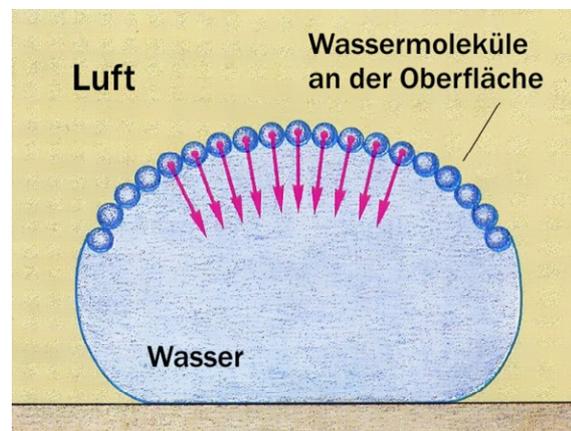
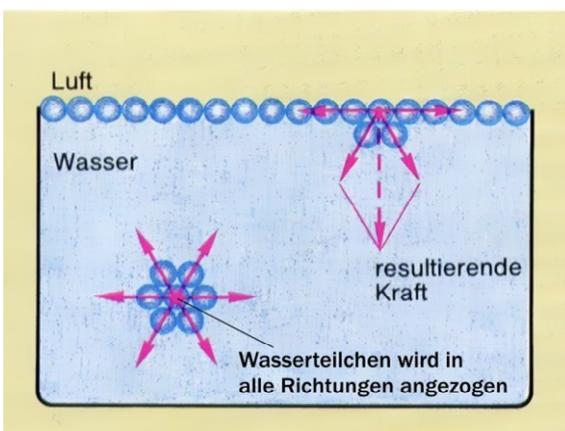
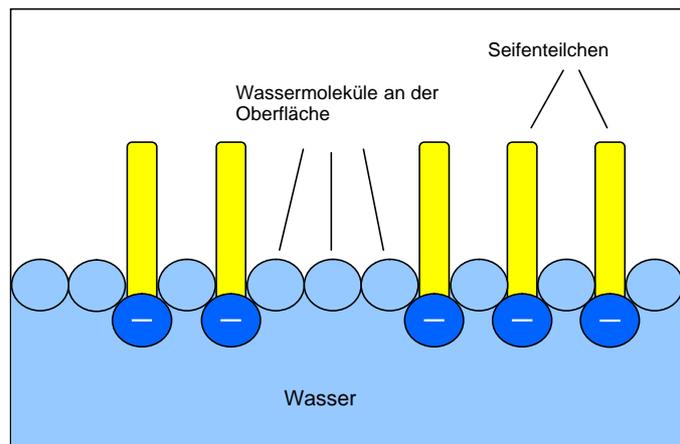
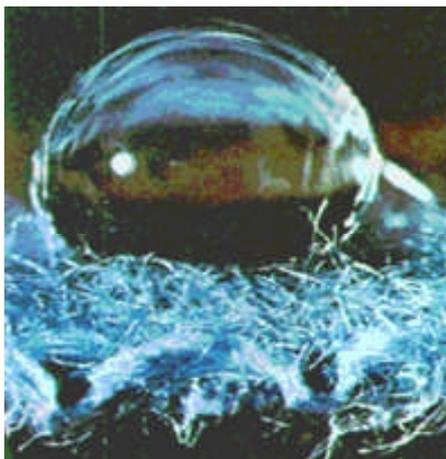
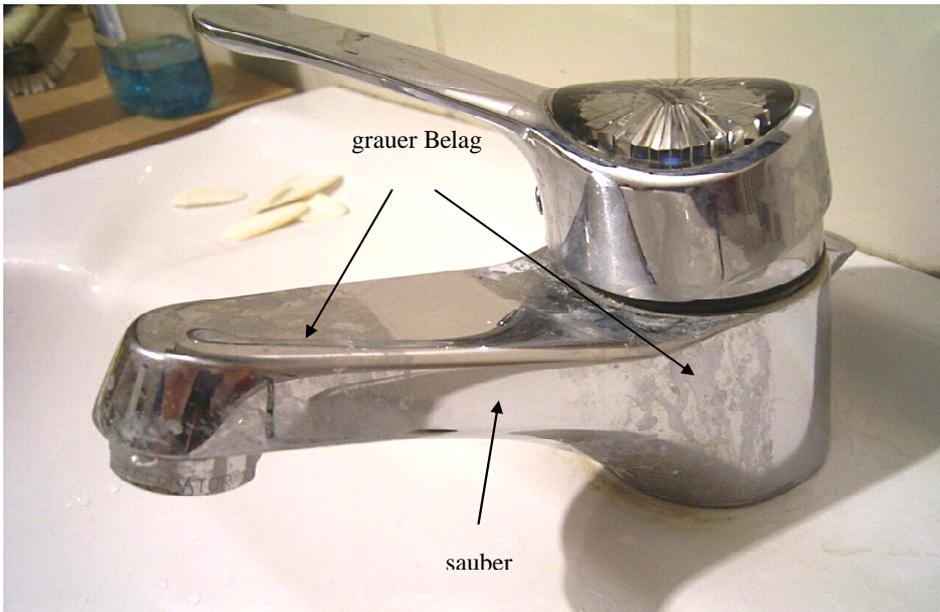
Wichtige Begriffe Thema 2

Tropfen, Elektronegativität, Teilladung, Wasserstoffbrückenbindung, Anziehung der Teilchen untereinander, Kräfte heben sich auf/heben sich nicht auf, alle äußeren Moleküle nach innen, Moleküle an Oberfläche dichter, Anziehungskräfte besonders stark, Oberflächenspannung, Zwischenschieben der Seifenmoleküle, Schwächung der Wasserstoffbrückenbindungen, Verringerung der Oberflächenspannung, Tensid, Benetzen.

Wichtige Begriffe Thema 3

grauer Belag, Wasserhärte, Ca^{2+} - und Mg^{2+} -Ionen, Salz, Fettsäureanion/Seifenanion, Natrium/Kalium-Salz einer Fettsäure, Kernseife, Schmierseife, Calcium-Ionen, Kalkseife, Wäsche wird grau, erhöhter Seifenverbrauch.

Folie Bilder GP



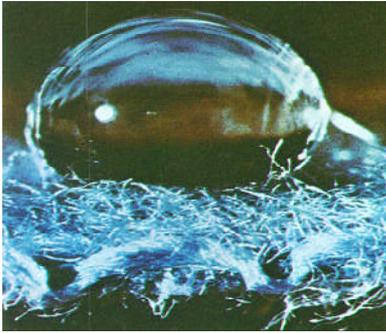
Bildquellen nicht selbst erstellter Grafiken des GP-Materials (Thema 2):

Abb. 1: Wassertropfen auf Wolle

Elemente Chemie I. Klett 1999, S. 330.

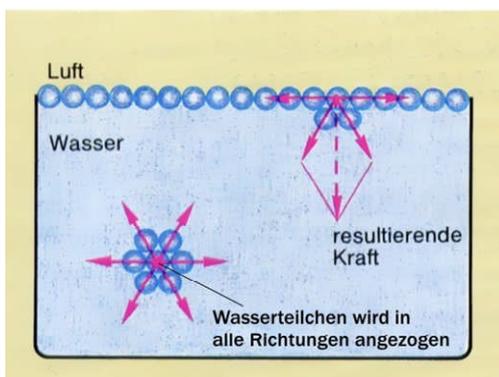


Abb. 3a: Kräfte auf Wassermoleküle an der Oberfläche und innerhalb einer Flüssigkeit

Natur und Technik - Chemie für Realschulen Cornelsen 1992, S. 193, leicht verändert.

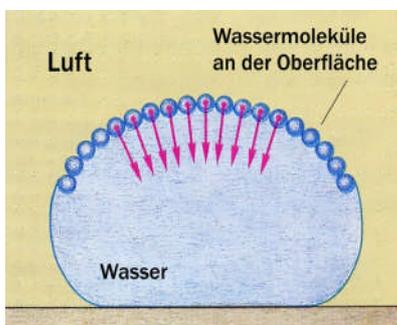


Abb. 3b: Tropfenbildung

Natur und Technik - Chemie für Realschulen. Cornelsen 1992, S. 193, leicht verändert.

Experimentieranleitungen für FU (falls gewünscht)

Experimentieranleitungen

Pfefferversuch

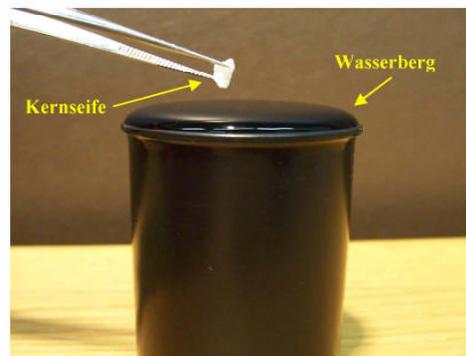
Bedecke eine Wasseroberfläche mit Hilfe eines Pfefferstreuers leicht mit gemahlenem Pfeffer und gib anschließend ein kleines Stück Kernseife in die Mitte der Wasseroberfläche. Beobachte!

Emulsionsversuch

In zwei Reagenzgläsern befindet sich jeweils ein Öl/Wasser-Gemisch. In ein Reagenzglas gibst du soviel Seifenlösung hinzu, bis fast keine Luft mehr drin ist. Verschließe mit einem Stopfen, schüttele beide kurz (Daumen auf Stopfen!) und lasse stehen. Beobachte!

Wasserbergversuch

Fülle ein Filmdöschen (Standzylinder) bis zum Rand mit Leitungswasser. Gib behutsam Wasser zu, bis sich ein deutlich sichtbarer Wasserberg gebildet hat (Abb. 1). Dann lege vorsichtig ein kleines Stückchen Kernseife auf die Wasseroberfläche und warte einen Augenblick.



Kalkseifeversuch

Gib zu den gleichen Mengen

1. dest. Wasser
2. Leitungswasser
3. mit Calciumionen angereichertes Wasser

jeweils fünf Tropfen Seifenlösung. Schwenke kurz hin und her und beobachte. Dann schüttele kräftig und beobachte.

12.2.3 Wissenstest

Datum:	Was wisst ihr über Seifen?	Vorname:
Klasse:		Erster Buchstabe des Nachnamens:

Dieser Test dürfte dir in ähnlicher Form bekannt vorkommen. Nun zeige, was du über Seife gelernt hast! Kreuze die richtige Aussage bzw. Antwort an. Es ist jeweils nur eine Lösung richtig. Raten ist in jedem Fall besser als nichts anzukreuzen. Viel Erfolg!!!

Ein Seifenteilchen ist aufgebaut wie ein Streichholz, ...

- ...der Kopf ist unpolar.
- ...der Schwanz ist geladen.
- ...der Kopf ist positiv geladen.
- ...der Schwanz ist unpolar.
- ...der Kopf ist ungeladen.

Kernseife ist das...

- ...Kaliumsalz einer Fettsäure.
- ...Calciumsalz einer Fettsäure.
- ...Natriumsalz einer Fettsäure.
- ...Kochsalz einer Fettsäure.
- ...Magnesiumsalz einer Fettsäure.

Zwischen Molekülen treten immer dann Wasserstoffbrückenbindungen auf, wenn...

- ...in diesen Molekülen Sauerstoff mit einem stark elektronegativen Atom verbunden ist.
- ...in diesen Molekülen zwei gleich stark elektronegative Atome vorkommen.
- ...Wasserstoff ein Elektron komplett an den Sauerstoff abgibt.
- ...die Moleküle geladen sind.
- ...in diesen Molekülen Wasserstoff mit einem stark elektronegativen Atom verbunden ist.

Die Ursache für den grauen Belag in nicht geputzten Waschbecken ist eine Reaktion von...

- ...Seifenanionen mit Kalk zu Kalkseife.
- ...Seifenkationen mit Magnesiumkationen zu Kalkseife.
- ...Seifenanionen mit Natriumkationen zu Kalkseife.
- ...Seifenkationen mit Calciumkationen zu Kalkseife.
- ...Seifenanionen mit Calciumkationen zu Kalkseife.

1

Blatt bitte wenden

Seife verdrängt Pfeffer auf einer Wasseroberfläche, weil...

- ...sich sofort Mizellen bilden, welche die Pfefferkörnchen beiseite schieben.
- ...sich zunächst eine monomolekulare Schicht bildet, die viel Platz benötigt.
- ...Seifenteilchen negativ geladen sind und die Pfefferteilchen abstoßen.
- ...sich zunächst eine mehrlagige Molekülschicht bildet, die viel Platz benötigt.
- ...Seifenteilchen positiv geladen sind und die Pfefferteilchen abstoßen.

Seife wäscht, weil sich die Seifenteilchen...

- ...mit ihrem Schwanz an den Schmutz lagern und ihn mit einer neuen unpolaren Oberfläche versehen.
- ...mit ihrem Kopf an den Schmutz lagern und ihn mit einer neuen unpolaren Oberfläche versehen.
- ...mit ihrem Kopf an den Schmutz lagern und ihn mit einer neuen polaren Oberfläche versehen.
- ...mit ihrem Kopf an den Schmutz lagern, das zusätzliche Gewicht zieht ihn auf den Boden.
- ...mit ihrem Schwanz an den Schmutz lagern und ihn mit einer neuen polaren Oberfläche versehen.

Seifenblasen kann man nicht ohne Seife machen, weil...

- ...die Oberflächenspannung von reinem Wasser zu hoch ist.
- ...für die Bildung von Seifenblasen Mizellen benötigt werden.
- ...reines Wasser zu schwer ist.
- ...die Oberflächenspannung von reinem Wasser zu niedrig ist.
- ...selbst reines Wasser immer etwas Schmutz enthält.

Kreuze die richtige Aussage an:

- Alle Stoffe, die die Oberflächenspannung herabsetzen, nennt man Seifen.
- Alle Tenside sind Seifen.
- Nicht alle Seifen sind Tenside.
- Alle Stoffe, die die Oberflächenspannung erhöhen, nennt man Tenside.
- Alle Seifen sind Tenside.

Als "hartes Wasser" bezeichnet man Wasser, das besonders...

- ...viele Mg^+ - Ionen enthält.
- ...viele Ca^{2+} - und Mg^+ - Ionen enthält.
- ...viele Ca^+ - und Mg^+ - Ionen enthält.
- ...viele Ca^{2+} - und Mg^{2+} - Ionen enthält.
- ...viele Ca^+ - Ionen enthält.

Wasserstoffbrückenbindungen kommen zustande, weil...

- ...Wasserstoff elektronegativer ist als Sauerstoff.
- ...Sauerstoff polar ist.
- ...Wasserstoff ein Dipol ist.
- ...Wasser elektronegativer ist als Sauerstoff.
- ...Sauerstoff elektronegativer ist als Wasserstoff.

Die Löslichkeitsregel besagt:

- Polare Stoffe lösen sich in unpolaren Stoffen, unpolare Stoffe in polaren.
- Polare Atome lösen sich in polaren Stoffen, unpolare Atome in unpolaren.
- Polare Stoffe lösen sich in polaren Stoffen, unpolare Stoffe in unpolaren.
- Gleiches löst nicht Gleiches.
- Unpolare Stoffe lösen sich in polaren Stoffen, polare Stoffe in unpolaren.

Seife verringert die Oberflächenspannung von Wasser, weil...

- ...Seife ein unpolarer Stoff ist.
- ...sich Seifenteilchen zwischen die Wassermoleküle an der Oberfläche schieben.
- ...Seifenteilchen die Wassermoleküle umschließen.
- ...Seife ein polarer Stoff ist.
- ...Seife an der Oberfläche Mizellen bildet.

Sobald kein Platz mehr für die sich ausdehnende Seife auf der Wasseroberfläche ist...

- ...klumpen die Seifenteilchen zusammen und schwimmen auf der Wasseroberfläche.
- ...bilden sich Kugeln aus Seifenteilchen, so genannte Mizellen.
- ...sinkt die jeweils unterste Seifenschicht auf den Grund.
- ...weichen die überzähligen Seifenteilchen zur Seite aus.
- ...türmen sich die Seifenteilchen schichtweise auf der Wasseroberfläche auf.

Bei einer Mizelle, die sich in Öl gelöst hat,...

- ...ragen alle polaren Schwänze nach außen zum Öl.
- ...ragen alle polaren Köpfe nach außen zum Öl.
- ...ragen alle unpolaren Schwänze nach innen.
- ...ragen alle unpolaren Schwänze nach außen zum Öl.
- ...ragen alle polaren Schwänze nach innen.

Blatt bitte wenden

Schmierseife ist das...

- ...Kaliumsalz einer Fettsäure.
- ...Magnesiumsalz einer Fettsäure.
- ...Calciumsalz einer Fettsäure.
- ...Kochsalz einer Fettsäure.
- ...Natriumsalz einer Fettsäure.

Der waschaktive Teil der Seife ist...

- ...sowohl Kation als auch Anion.
- ...der Kopf.
- ...das Seifenanion.
- ...der Schwanz.
- ...das Seifenkation.

Würde Kalkseife auch in sehr weichem Regenwasser ausfallen?

- Ja, wenn man erhitzt.
- Ja, wenn man schüttelt.
- Nein, außer man erhitzt.
- Ja, wenn man schüttelt und erhitzt.
- Nein.

In hartem Wasser benötigt man zum Waschen mehr Seife, weil...

- ... sich zuerst eine monomolekulare Schicht bildet.
- ... die Seifenanionen durch Calcium- und Magnesiumionen inaktiviert sind.
- ... erst die Oberflächenspannung herabgesetzt werden muss.
- ... hartes Wasser mehr Schmutz enthält.
- ... Calcium- und Magnesiumionen die Wassermoleküle an sich binden.

Ein Wassertropfen ist rund, weil...

- ...die Moleküle an der Oberfläche seitlich angezogen werden.
- ...sich zwischen den Molekülen Wasserstoffbrückenbindungen bilden.
- ...die resultierende Kraft gleich null ist.
- ...die Moleküle an der Oberfläche nach innen gezogen werden.
- ...an der Oberfläche zusätzliche Kräfte wirken.

Seifenteilchen auf einer Wasseroberfläche...

- ...bilden schwimmende Kugeln.
- ...tauchen mit dem Schwanz komplett ins Wasser.
- ...stehen senkrecht nebeneinander.
- ...haben keine bestimmte Ausrichtung.
- ...liegen waagrecht nebeneinander.

Löst sich Seife auch in Öl?

- Ja.
- Nur wenn auch etwas Wasser dabei ist.
- Nur Flüssigseife.
- Nur in kochendem Öl.
- Nur wenn man schüttelt.

Größere Mengen Wasser besitzen keine Tropfenform, obwohl alle äußeren Teilchen ins Innere gezogen werden wollen. Der Grund ist, dass...

- ...größere Mengen Wasser oft Spuren von Seife enthalten, welche die Oberflächenspannung herabsetzt.
- ...die Anziehungskraft der Erde stärker ist als die Kräfte ins Innere des Tropfens, sie „drückt“ die Tropfen flach, so dass sie zusammenfließen.
- ...bei großen Mengen Wasser mehr Calcium- und Magnesiumionen vorhanden sind.
- ...bei großen Mengen Wasser kaum Wasserstoffbrückenbindungen existieren, welche die Teilchen ins Innere ziehen könnten.
- ...die Luft das Wasser von außen flach drückt.

Öl löst sich nicht in Wasser,...

- ...weil Öl unpolar ist.
- ...weil Öl ungeladen ist.
- ...weil Öl polar ist.
- ...weil Öl geladen ist.
- ...weil Öl hydrophil ist.

Bei einer Wasserstoffbrückenbindung...

- ...ist der Wasserstoff komplett positiv geladen.
- ...besitzen Wasserstoff und Sauerstoff nur Teilladungen.
- ...ist der Sauerstoff komplett negativ geladen.
- ...sind die einzelnen Wasserteilchen unterschiedlich geladen.
- ...liegen Wasserstoff und Sauerstoff als Ionen vor.

Die Oberflächenspannung entsteht, weil...

- ...die Moleküle im Inneren einer Flüssigkeit dichter zusammenrücken und sich daher stärker anziehen.
- ...die Erdanziehungskraft auf Oberflächenmoleküle besonders stark wirkt.
- ...die resultierende Kraft auf ein Teilchen im Inneren einer Flüssigkeit gleich null ist.
- ...die Moleküle an der Oberfläche dichter zusammenrücken und sich daher stärker anziehen.
- ...die Moleküle an der Oberfläche mehr Nachbarn haben als im Inneren einer Flüssigkeit.

Wie viele Seifenanionen (Fettsäureanionen) reagieren mit einem einzigen Ca^{2+} -Ion zu Kalkseife?

- 1/2
- 2
- alle
- 1
- 3

Wenn man Natrium-Ionen in eine schwach konzentrierte Kernseifenlösung gibt,...

- ...ändert sich die Wasserhärte.
- ...steigt die Oberflächenspannung des Wassers.
- ...passiert nichts, die Kernseife bleibt gelöst.
- ...fällt Kalkseife aus.
- ...sinkt die Oberflächenspannung des Wassers.

Seifen bestehen aus...

- ...dem Kation einer Fettsäure und einem Anion.
- ...dem Anion einer Fettsäure und einem Kation.
- ...einer Fettsäure und Magnesium.
- ...einer Fettsäure und Kochsalz.
- ...einer Fettsäure und Wasser.

Die Oberflächenspannung des Wassers hat zur Folge, dass...

- ...Kalkseife ausfällt.
- ...Seife schlechter in Stoffe eindringen kann.
- ...Wasser schlechter in Stoffe eindringen kann.
- ...Seife sich schlechter in Wasser löst.
- ...Wäsche grau wird.

Seife setzt auch die Oberflächenspannung von Öl herab, weil sich...

- ...Quatsch! Öl besitzt keine Oberflächenspannung.
- ...Seifenteilchen zwischen die Ölmoleküle schieben, mit dem Schwanz zuerst.
- ...Seifenteilchen zwischen die Ölmoleküle schieben, mit dem Kopf zuerst.
- ...mit Hilfe der Seife Luft im Öl löst.
- ...Mizellen im Inneren der Flüssigkeit bilden.

Wenn man zu einem Öl/Wasser-Gemisch Seife gibt und schüttelt,...

- ...verschwindet die Ölschicht (zumindest für eine Zeit lang).
- ...fällt Kalkseife aus.
- ...entsteht eine klare Lösung.
- ...bildet sich eine weitere Schicht.
- ...passiert nichts anderes, als wenn keine Seife zugegeben worden wäre.

Der hydrophobe Teil des Seifenteilchens ist...

- ...der kurze Anteil.
- ...polar.
- ...der Kopf.
- ...unpolar.
- ...negativ geladen.

Das Seifenkation ist...

- ...entscheidend für den Seifentyp.
- ...waschaktiv.
- ...die Fettsäure.
- ...hydrophob.
- ...immer Na^+ .

Die resultierende Kraft auf ein Wasserteilchen...

- ...im Inneren ist gleich eins.
- ...im Inneren ist gleich null.
- ...sorgt dafür, dass Wassermoleküle gewinkelt sind.
- ...sorgt dafür, dass sich alle Wassermoleküle in eine bestimmte Richtung bewegen.
- ...an der Oberfläche ist gleich eins.

Richtige Antworten WT VU - HU II

Änderungen von VU zu HU gelb bzw. grau

Frage	Richtige Antwort	Subtest	entspricht	VU	entspricht	Frage
Frage 1	4	1	LÖ	LÖ	LÖ	Frage 1
Frage 2	3	3	NA	NA		Frage 5
Frage 3	5	4	TR	OB		Frage 6
Frage 4	5	3	NA	NA		Frage 11
Frage 5	2	1	LÖ	LÖ		Frage 13
Frage 6	5	1	LÖ	LÖ		Frage 14
Frage 7	1	4	TR	OB		Frage 20
Frage 8	5	2	OB	OB		Frage 21
Frage 9	4	3	NA	NA		Frage 23
Frage 10	5	2	OB	OB		Frage 31
Frage 11	3	1	LÖ	LÖ		Frage 32
Frage 12	2	2	OB	OB		
Frage 13	2	1	LÖ	LÖ	OB	Frage 8
Frage 14	4	1	LÖ	LÖ		Frage 10
Frage 15	1	3	NA	NA		Frage 12
Frage 16	3	3	NA	NA		Frage 19
Frage 17	5	3	NA	NA		Frage 24
Frage 18	2	3	NA	NA		Frage 25
Frage 19	4	2	OB	OB		Frage 29
Frage 20	3	1	LÖ	LÖ		Frage 34
Frage 21	1	1	LÖ	LÖ		
Frage 22	2	2	TR	TR	NA	Frage 2
Frage 23	1	1	LÖ	LÖ		Frage 4
Frage 24	2	2	OB	OB		Frage 9
Frage 25	4	2	OB	OB		Frage 15
Frage 26	2	3	NA	NA		Frage 16
Frage 27	3	4	TR	TR		Frage 17
Frage 28	2	3	NA	NA		Frage 18
Frage 29	3	2	OB	OB		Frage 26
Frage 30	2	2	TR	TR		Frage 28
Frage 31	1	1	LÖ	LÖ		Frage 33
Frage 32	4	1	LÖ			
Frage 33	1	3	NA		TR	Frage 3
Frage 34	2	2	OB			Frage 7
						Frage 22
						Frage 27
						Frage 30

LÖ = Löslichkeit, Thema 1

OB = Oberflächenspannung, Thema 2

NA = Nachteile, Thema 3

TR = Transfer

	schr zu- treffend	zu- treffend	mittel	unzu- treffend	schr unzu- treffend
Im Chemieunterricht lerne ich etwas	<input type="checkbox"/>				
Ich glaube, dass ich im Fach Chemie etwas fürs Leben lernen kann	<input type="checkbox"/>				
Chemische Kenntnisse kann ich auch in anderen Fächern gebrauchen	<input type="checkbox"/>				
Chemie finde ich schwieriger als andere Fächer	<input type="checkbox"/>				
Der Chemieunterricht ist abwechslungsreich	<input type="checkbox"/>				
Im Chemieunterricht habe ich ein ungutes Gefühl im Bauch	<input type="checkbox"/>				
Ich kann im Chemieunterricht etwas leisten	<input type="checkbox"/>				
Ich glaube, dass chemische Kenntnisse wichtig sind im Leben	<input type="checkbox"/>				
Ich finde die Benotung im Chemieunterricht ungerechter als in anderen Fächern	<input type="checkbox"/>				
Ich mache lieber Hausaufgaben in Chemie als in anderen Fächern	<input type="checkbox"/>				
Wenn ich könnte, würde ich Chemie abwählen	<input type="checkbox"/>				
Chemie finde ich zu theoretisch	<input type="checkbox"/>				

	sehr zu- treffend	zu- treffend	mittel	unzu- treffend	sehr unzu- treffend
Im Chemieunterricht lerne ich Dinge, die ich später gebrauchen kann	<input type="checkbox"/>				
Ich gehe ungern zum Chemieunterricht	<input type="checkbox"/>				
Ich finde den Chemieunterricht interessant	<input type="checkbox"/>				
Wenn ich mich im Chemieunterricht anstrengte, bekomme ich auch gute Noten	<input type="checkbox"/>				
Das Fach Chemie finde ich spannend	<input type="checkbox"/>				
Ich hätte gerne mehr Stunden Chemie in der Woche	<input type="checkbox"/>				
Der Sinn des Faches Chemie ist mir unklar	<input type="checkbox"/>				
Ich finde den Chemieunterricht langweilig	<input type="checkbox"/>				
Mir macht der Chemieunterricht Spaß	<input type="checkbox"/>				
Im Chemieunterricht werde ich zu viel gefordert	<input type="checkbox"/>				
Unter Chemie kann ich mir sehr wenig vorstellen	<input type="checkbox"/>				
Ich könnte mir vorstellen, Chemielehrer zu werden	<input type="checkbox"/>				



Fragebogen zur Attraktivität der letzten drei Stunden

sehr zu- treffend	zu- treffend	mittel	unzu- treffend	sehr unzu- treffend
-------------------------	-----------------	--------	-------------------	---------------------------

Ich fand den Unterricht der letzten 3 Stunden angenehm	<input type="checkbox"/>				
Ich habe in den letzten 3 Stunden das meiste verstanden	<input type="checkbox"/>				
Ich fand den Unterricht der letzten 3 Stunden abwechslungsreich	<input type="checkbox"/>				
Ich war im Unterricht der letzten 3 Stunden entspannt	<input type="checkbox"/>				
Ich fand den Unterricht der letzten 3 Stunden stressig	<input type="checkbox"/>				
Ich fühlte mich in den letzten 3 Stunden von meinen Mitschülern ernst genommen	<input type="checkbox"/>				
Mir hat der Chemieunterricht der letzten 3 Stunden gefallen	<input type="checkbox"/>				
So wie in den letzten 3 Stunden sollten auch andere Themen unterrichtet werden	<input type="checkbox"/>				
Ich bin froh, dass diese Unterrichtseinheit nun vorbei ist	<input type="checkbox"/>				
Die Unterrichtsmethode der letzten 3 Stunden fand ich gut	<input type="checkbox"/>				
In den letzten 3 Stunden war ich unaufmerksam	<input type="checkbox"/>				
Ich habe in den letzten 3 Stunden wenig Neues gelernt	<input type="checkbox"/>				

	sehr zu- treffend	zu- treffend	mittel	unzu- treffend	sehr unzu- treffend
Ich hätte in den letzten 3 Stunden lieber ein anderes Thema gemacht	<input type="checkbox"/>				
Ich fand die Versuche der letzten 3 Stunden gut	<input type="checkbox"/>				
Mir hat das Klassenklima in den letzten 3 Stunden gefallen	<input type="checkbox"/>				
Ich konnte in den letzten 3 Stunden selbstständig arbeiten	<input type="checkbox"/>				
Die Versuche der letzten 3 Stunden fand ich weniger gut	<input type="checkbox"/>				
Der Lehrer/die Lehrerin hat in den letzten 3 Stunden zu viel geredet	<input type="checkbox"/>				
Ich wurde in den letzten 3 Stunden von meinen Mitschülern abgelenkt	<input type="checkbox"/>				
In den letzten 3 Stunden war ich müde	<input type="checkbox"/>				
Das Thema Seife war mir zu schwierig	<input type="checkbox"/>				
Das Thema Seifen fand ich interessant	<input type="checkbox"/>				
Ich konnte mich in den letzten 3 Stunden gut konzentrieren	<input type="checkbox"/>				
Der Unterricht der letzten 3 Stunden hat mir Spaß gemacht	<input type="checkbox"/>				



Fragebogen zur Attraktivität des Gruppenpuzzles

sehr zu- treffend	zu- treffend	mittel	unzu- treffend	sehr unzu- treffend
-------------------------	-----------------	--------	-------------------	---------------------------

Beim Gruppenpuzzle fand ich es gut, dass nicht immer der Lehrer erklärt	<input type="checkbox"/>				
Im Gruppenpuzzle habe ich viel intensiver gearbeitet als sonst	<input type="checkbox"/>				
Im Gruppenpuzzle konnte ich viel selbstständiger arbeiten als sonst	<input type="checkbox"/>				
Ich habe durch das Gruppenpuzzle mehr gelernt als im normalen Unterricht	<input type="checkbox"/>				
Im Gruppenpuzzle konnte ich endlich zeigen, was ich wirklich kann	<input type="checkbox"/>				
Ich finde normalen Unterricht effektiver als das Gruppenpuzzle	<input type="checkbox"/>				
Beim Gruppenpuzzle blieben meine Beiträge unberücksichtigt	<input type="checkbox"/>				
Beim Gruppenpuzzle fühlte ich mich vom Lehrer allein gelassen	<input type="checkbox"/>				
Ich finde normalen Unterricht besser als das Gruppenpuzzle	<input type="checkbox"/>				
Beim Gruppenpuzzle fühle ich mich ungerecht beurteilt	<input type="checkbox"/>				
Die Verantwortung beim Gruppenpuzzle war mir zu groß	<input type="checkbox"/>				
Ich würde gerne noch einmal Unterricht mit dem Gruppenpuzzle machen	<input type="checkbox"/>				

	sehr zu- treffend	zu- treffend	mittel	unzu- treffend	sehr unzu- treffend
Beim Gruppenpuzzle war ich unsicher, wie ich vorgehen sollte	<input type="checkbox"/>				
Den Unterricht mit dem Gruppenpuzzle empfand ich als weniger stressig	<input type="checkbox"/>				
Durch das Gruppenpuzzle habe ich die Lerninhalte viel besser verstanden	<input type="checkbox"/>				
Ich war beim Gruppenpuzzle mit den Erklärungen der Experten zufrieden	<input type="checkbox"/>				
Ich hatte beim Gruppenpuzzle Probleme, mich verständlich auszudrücken	<input type="checkbox"/>				
Einige Mitschüler aus meiner Gruppenpuzzle-Gruppe sind unsympathisch	<input type="checkbox"/>				
Beim Gruppenpuzzle hatte ich Probleme, den Text zu verstehen	<input type="checkbox"/>				
Das Gruppenpuzzle hat meinen Lernerfolg positiv beeinflusst	<input type="checkbox"/>				
Ich würde gerne weniger Verantwortung für mein Lernen übernehmen	<input type="checkbox"/>				
Ich finde, dass der Experte zu großen Einfluss auf meinen Lernerfolg hatte	<input type="checkbox"/>				
Das Expertenmaterial des Gruppenpuzzles hat mir gut gefallen	<input type="checkbox"/>				

12.2.5 Individuelle Schülerbeurteilungen

Individuelle Schülerbeurteilung

Liebe Lehrerinnen und Lehrer,

bitte füllen Sie den Bogen für jeden Schüler klassenweise in alphabetischer Reihenfolge aus. Dabei sollten die Kreuzchen in der Nähe des Begriffes gemacht werden, der Ihrer Meinung nach den Schüler am besten beschreibt. Bitte denken Sie nicht zu lange über jeden Schüler nach. Wenn Sie eine Eigenschaft nicht nach kurzer Zeit (5 s) einschätzen können, kreuzen Sie bitte „weiß nicht“ (Spalte ganz rechts) an. Sie erhalten diesen Rating-Bogen nach der Untersuchung noch einmal. Sollten Sie dann im Wiederholungsbogen etwas ankreuzen können, wäre das auch ein schönes Ergebnis!

Ich bin mir bewusst, dass Sie das Ausfüllen der Bögen Zeit und Mühe kosten wird und weiß Ihren Einsatz sehr zu würdigen. Im Voraus schon einmal ganz herzlichen Dank.

Anmerkung: positive und negative Items wechseln sich ab!

Klasse: 10 _____

Name des Schülers: _____

Leistungsbezogene Eigenschaften							weiß nicht	
meldet sich von alleine	<input type="checkbox"/>	muss angesprochen werden	<input type="checkbox"/>					
passiv	<input type="checkbox"/>	aktiv	<input type="checkbox"/>					
selbstständig	<input type="checkbox"/>	unselbstständig	<input type="checkbox"/>					
Ausdrucksweise weitschweifig	<input type="checkbox"/>	Ausdrucksweise prägnant	<input type="checkbox"/>					
verwendet häufig Fachbegriffe	<input type="checkbox"/>	verwendet kaum Fachbegriffe	<input type="checkbox"/>					
verwendet Fachbegriffe falsch	<input type="checkbox"/>	verwendet Fachbegriffe richtig	<input type="checkbox"/>					
erklärt anschaulich	<input type="checkbox"/>	erklärt abstrakt	<input type="checkbox"/>					
umgangssprachliche Artikulation	<input type="checkbox"/>	angemessene Fachsprache	<input type="checkbox"/>					
beginnt von alleine mit der Arbeit	<input type="checkbox"/>	muss aufgefordert werden	<input type="checkbox"/>					
ängstlich	<input type="checkbox"/>	selbstbewusst	<input type="checkbox"/>					
hinterfragend	<input type="checkbox"/>	unkritisch	<input type="checkbox"/>					
chaotisch	<input type="checkbox"/>	strukturiert	<input type="checkbox"/>					
intelligent	<input type="checkbox"/>	dumm	<input type="checkbox"/>					
fachlich inkompetent	<input type="checkbox"/>	fachlich kompetent	<input type="checkbox"/>					
überträgt Tafelanschrieb ins Heft	<input type="checkbox"/>	schreibt nicht von der Tafel ab	<input type="checkbox"/>					
destruktiv	<input type="checkbox"/>	konstruktiv	<input type="checkbox"/>					
Nicht leistungsbezogene Eigenschaften								
vorlaut	<input type="checkbox"/>	schüchtern	<input type="checkbox"/>					
aufmerksam	<input type="checkbox"/>	verträumt	<input type="checkbox"/>					
stört	<input type="checkbox"/>	hält Regeln ein	<input type="checkbox"/>					
hört gut zu	<input type="checkbox"/>	unterbricht schnell	<input type="checkbox"/>					
unbeliebt	<input type="checkbox"/>	beliebt	<input type="checkbox"/>					
hilfsbereit	<input type="checkbox"/>	egoistisch	<input type="checkbox"/>					
ungeduldig	<input type="checkbox"/>	geduldig	<input type="checkbox"/>					
freundlich	<input type="checkbox"/>	frech	<input type="checkbox"/>					
faul	<input type="checkbox"/>	fleißig	<input type="checkbox"/>					
starr	<input type="checkbox"/>	beweglich	<input type="checkbox"/>					
leise	<input type="checkbox"/>	laut	<input type="checkbox"/>					
redselig	<input type="checkbox"/>	still	<input type="checkbox"/>					
frisch	<input type="checkbox"/>	müde	<input type="checkbox"/>					

Beobachtungsbogen Expertenphase

Datum:	Klasse 10__	Thema Nr.:	Beobachter:
	Vorname und erster Buchstabe des Nachnamens:		
Qualität der Erklärung (nicht Fragen klären!)	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler erwähnt festgelegte Schlüsselbegriffe	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>
	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler erklärt ohne abzulesen	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>
	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler erklärt fachlich richtig	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>
	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Kann gut erklären (im Vergleich, z.B. strukturiert, nicht zu schnell, stellt Zusammenhänge her)	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>
	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> S. wirkt motiviert (im Vergleich, z.B. wiederholt auf Nachfrage/ist geduldig/fragt nach, ob verstanden)	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Sehr zureichend <input type="checkbox"/> Zureichend <input type="checkbox"/> Mittel <input type="checkbox"/> Unzureichend <input type="checkbox"/> Sehr unzureichend <input type="checkbox"/>
	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler benutzt Hilfsmittel (Abbildung, Zettel, Stift)	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Oft <input type="checkbox"/> Selten <input type="checkbox"/> Gar nicht <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Oft <input type="checkbox"/> Selten <input type="checkbox"/> Gar nicht <input type="checkbox"/>
Art der Kommunikation	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler lässt andere austreden	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Ja <input type="checkbox"/> Überwiegend <input type="checkbox"/> Zum Teil <input type="checkbox"/> Nein <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Ja <input type="checkbox"/> Überwiegend <input type="checkbox"/> Zum Teil <input type="checkbox"/> Nein <input type="checkbox"/>
Motivation	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler hat Hausaufgabe gemacht (nachfragen!)	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Ja <input type="checkbox"/> Überwiegend <input type="checkbox"/> Zum Teil <input type="checkbox"/> Nein <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Ja <input type="checkbox"/> Überwiegend <input type="checkbox"/> Zum Teil <input type="checkbox"/> Nein <input type="checkbox"/>
Rollerverteilung (am Schluss der Stunde)	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Schüler ist (im Vergleich zu anderen Gruppenmitgliedern):	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Wortführer <input type="checkbox"/> Durchschnittlich <input type="checkbox"/> Rezipient <input type="checkbox"/> Weiß nicht <input type="checkbox"/>	Alle <input type="checkbox"/> Viele <input type="checkbox"/> Wenige <input type="checkbox"/> Keinen <input type="checkbox"/> Wortführer <input type="checkbox"/> Durchschnittlich <input type="checkbox"/> Rezipient <input type="checkbox"/> Weiß nicht <input type="checkbox"/>

Beobachtungsbogen Vermittlungsphase

Datum:	Klasse 10 ___				Beobachter:			
Vorname und erster Buchstabe Nachnamens:	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>		
Qualität der Wissensaufnahme Qualität der Erklärung (nicht Fragen klären!)	Alle <input type="checkbox"/>	Viele <input type="checkbox"/>	Wenige <input type="checkbox"/>	Keinen <input type="checkbox"/>	Alle <input type="checkbox"/>	Viele <input type="checkbox"/>	Wenige <input type="checkbox"/>	Keinen <input type="checkbox"/>
Schüler macht sich Notizen	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Schüler erwähnt festgelegte Schlüsselbegriffe	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Schüler erklärt ohne abzulesen	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Schüler erklärt fachlich richtig	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Kann gut erklären (im Vergleich, z.B. strukturiert, nicht zu schnell, stellt Zusammenhänge her)	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Schüler wirkt motiviert (im Vergleich, z.B. wiederholt auf Nachfrage/ist geduldig/fragt nach, ob verstanden)	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Schüler benutzt Hilfsmittel (Abbildung, Zettel, Stift) zum Erklären	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>	Sehr zutreffend <input type="checkbox"/>	Zutreffend <input type="checkbox"/>	Mittel <input type="checkbox"/>	Unzutreffend <input type="checkbox"/>
Schüler lässt andere ausreden	Oft <input type="checkbox"/>	Selten <input type="checkbox"/>	Gar nicht <input type="checkbox"/>	Oft <input type="checkbox"/>	Selten <input type="checkbox"/>	Gar nicht <input type="checkbox"/>	Oft <input type="checkbox"/>	Selten <input type="checkbox"/>
Schüler ist (im Vergleich zu anderen Gruppenmitgliedern)	Überwiegend <input type="checkbox"/>	Zum Teil <input type="checkbox"/>	Nein <input type="checkbox"/>	Überwiegend <input type="checkbox"/>	Zum Teil <input type="checkbox"/>	Nein <input type="checkbox"/>	Überwiegend <input type="checkbox"/>	Zum Teil <input type="checkbox"/>
Art der Kommunikation	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
Rollenverteilung (am Schluss der Stunde)	Wortführer <input type="checkbox"/>	Durchschnittlich <input type="checkbox"/>	Rezipient <input type="checkbox"/>	Weiß nicht <input type="checkbox"/>	Wortführer <input type="checkbox"/>	Durchschnittlich <input type="checkbox"/>	Rezipient <input type="checkbox"/>	Weiß nicht <input type="checkbox"/>

Datum:	Wichtige Begriffe im GP	Klasse: 10
Phase im GP:		Beobachter:

Folgende „wichtige Begriffe“ wurden von den Schülern erwähnt (bitte Anfangsbuchstaben des Schülers angeben):

Thema 1

Begriff	Schüler	Begriff	Schüler	Begriff	Schüler
polar/unpolar		Löslichkeits-Regel		Pfefferteilchen werden beiseite geschoben	
hydrophob/hydrophil		senkrecht		Mizelle	
Streichholzmodell		monomolekulare Schicht		öliger Schmutz bzw. Öl	
Kopf/Schwanz		viel Platz		neue polare Oberfläche	

Thema 2

Begriff	Schüler	Begriff	Schüler	Begriff	Schüler
Tropfen		Kräfte heben sich auf / heben sich nicht auf		Zwischenschieben der Seifenteilchen	
Elektronegativität		alle äußeren Moleküle nach innen		Schwächung der Wasserstoffbrückenbindungen	
Teilladung		Moleküle an Oberfläche dichter		Verringerung der Oberflächenspannung	
Wasserstoffbrückenbindung		Anziehungskräfte besonders stark		Tensid	
Anziehung der Teilchen untereinander		Oberflächenspannung		Benetzen	

Thema 3

Begriff	Schüler	Begriff	Schüler	Begriff	Schüler
grauer Belag		Fettsäure-/Seifenanion		Calciumionen	
Wasserhärte		Natrium/Kaliumsalz einer Fettsäure		Kalkseife	
Ca ²⁺ - und Mg ²⁺ -Ionen		Kernseife		Wäsche wird grau	
Salz		Schmierseife		erhöhter Seifenverbrauch	

12.2.7 Stundenprotokoll Frontalunterricht

Datum:	Stundenprotokoll FU	Klasse: 10
Studentyp:		LehrerIn:

Fehlende SchülerInnen:

Folgende „wichtige Begriffe“ wurden vom Lehrer/der Lehrerin (L) bzw. den Schülern (S) erwähnt:

polar/unpolar(), hydrophob/hydrophil(), Streichholzmodell(), Kopf(), Schwanz(), Löslichkeits-Regel(), senkrecht(), monomolekulare Schicht(), viel Platz(), Pfefferteilchen werden beiseite geschoben(), Mizelle(), öliges Schmutz bzw. Öl(), neue polare Oberfläche().

Tropfen(), Elektronegativität(), Teilladung(), Wasserstoffbrückenbindung(), Anziehung der Teilchen untereinander(), Kräfte heben sich auf/heben sich nicht auf(), alle äußeren Moleküle nach innen(), Moleküle an Oberfläche dichter(), Anziehungskräfte besonders stark(), Oberflächenspannung(), Zwischenschieben der Seifenteilchen(), Schwächung der Wasserstoffbrückenbindungen(), Verringerung der Oberflächenspannung(), Tensid(), Benetzen().

grauer Belag(), Wasserhärte(), Ca^{2+} - und Mg^{2+} -Ionen(), Salz(), Fettsäure-/Seifenanion(), Natrium/Kaliumsalz einer Fettsäure(), Kernseife(), Schmierseife(), Calciumionen(), Kalk-seife(), Wäsche wird grau(), erhöhter Seifenverbrauch().

Folgende Lernziele wurden von den SuS erreicht:



	sehr zu- treffend	zu- treffend	mittel	Unzu- treffend	sehr unzu- treffend
Du weißt, wie ein Seifenteilchen aufgebaut ist.	<input type="checkbox"/>				
Dir ist klar, was passiert, wenn man ein bisschen Seife in Wasser gibt.	<input type="checkbox"/>				
Du kannst erklären, was Mizellen sind und ab wann sie sich bilden.	<input type="checkbox"/>				
Du weißt, warum sich öliges Schmutz bzw. Öl nicht in Wasser löst.	<input type="checkbox"/>				
Du hast verstanden, warum Schmutz von Seife in Wasser gelöst wird.	<input type="checkbox"/>				
Du weißt, wie Wasserstoffbrückenbindungen zustande kommen.	<input type="checkbox"/>				
Dir ist klar, warum ein Wassertropfen rund ist!	<input type="checkbox"/>				
Du kannst das Zustandekommen der Oberflächenspannung erklären.	<input type="checkbox"/>				
Du weißt, warum Seife die Oberflächenspannung herabsetzt.	<input type="checkbox"/>				
Du kennst die Bedeutung der Oberflächenspannung für das Wäsche- waschen.	<input type="checkbox"/>				
Du weißt, was man unter Wasserhärte versteht.	<input type="checkbox"/>				
Du kennst den waschaktiven Teil der Seife.	<input type="checkbox"/>				
Du hast verstanden, dass Seifen aus dem Anion einer Fettsäure und einem Kation bestehen.	<input type="checkbox"/>				
Du weißt, was Kern- bzw. Schmierseife ist.	<input type="checkbox"/>				
Du kannst erklären, wie Seife mit Ca^{2+} - Ionen reagiert (mit obiger Re- aktionsgleichung!).	<input type="checkbox"/>				
Dir sind die beiden Nachteile von Seife klar.	<input type="checkbox"/>				

Sachverhalte, die die SuS wahrscheinlich nicht verstanden haben:

Legende:

sehr zutreffend: Sachverhalte, die von den SuS erwähnt wurden bzw. schon bekannt waren.

zutreffend: Sachverhalte, die vom Lehrer ausführlich erklärt wurden.

mittel: Sachverhalte, die vom Lehrer nur kurz/in einem Nebensatz erwähnt wurden.

unzutreffend: Sachverhalte, die vom Lehrer nicht erwähnt wurden.

sehr unzutreffend: Sachverhalte, die vom Lehrer falsch erklärt bzw. nicht korrigiert wurden.

12.2.8 Erfahrungen

Auf Basis der Erfahrungen während der Entwicklung und Erprobung des Gruppenpuzzles sollen an dieser Stelle einige Tipps zur Erstellung und Durchführung aufgelistet werden. Diese wurden im Rahmen zweier Lehrerfortbildungen (Neue Methoden im Chemieunterricht) in den Jahren 2005 und 2006 zusammengestellt und mögen für weitere Einsätze des Gruppenpuzzles hilfreich sein.

Tipps zur Erstellung

- Überlegen, welches **Oberthema** sich gut in Unterthemen **unterteilen** lässt
- **Nicht zu viele Unterthemen**, schließlich muss jedes erklärt werden
- Verwendung von bereits vorhandenem **Material** - z. B. **aus Schulbüchern**
- Die **Unterthemen** sollten in etwa **gleich schwierig** und voneinander **unabhängig** sein (Ist das Material heterogen, kann dieser Umstand zur inneren Differenzierung genutzt werden)
- Alle **neuen Begriffe** sollten im Text **erklärt** werden
- **Klare Strukturierung** (präzise Überschrift, Absätze, Teilüberschriften)
- **Altersgemäße Wortwahl** (trotzdem Fachsprache)
- **Vorwissen** der Schüler **berücksichtigen** (Texte selbsterklärend)
- **Altersgemäße** Begrenzung der **Textlänge** (max. 20 min, Faustregel: ca. 1 Din A4 Seite Text, 0,5 Seiten Bilder)
- Verwendung unterstützender **Illustrationen**
- **Keine** (noch so interessanten) **Zusatzinfos** → lenken ab
- Für die Schüler **transparente Lernziele**
- Auflistung der **wichtigen Begriffe** (Strukturierungshilfe für Vortrag, altersabhängig)
- Evtl. **Lernkontrollen** als **Zwischensicherungen** auf Expertenmaterial
- Möglichst **1 Versuch** pro Unterthema
- Pro Phase ein **Arbeitsblatt**

Tipps für die Durchführung

- **Ausführliche Erklärung** des GPs im Vorfeld
- Erinnerung **vor jeder Phase, was genau zu tun ist**
- **Schülervortrag** bzw. **Test** am Ende des GPs um Beliebigkeit vorzubeugen
- **Kontrolle** ob alle **Aufgaben** erledigt worden sind (während des gesamten GPs)
- Durch **kritische Fragen** zum **erneuten Überlegen anregen**
- **Keine unmittelbaren Verbesserungen von** falschen Erklärungen → Schüler sollen selbstständig Überlegen
- Auf jeden Fall **anwesend sein** während der Experten- und Vermittlungsphase
- Die **Stammgruppen** sollten aufgelöst bzw. **vom Lehrer gebildet** werden, wenn das Klassenklima in Ordnung ist, da dieses nachweislich effektiver ist. Nur anfangs sollten die Schüler ihre Stammgruppen selbst bilden, um eine weitere Verunsicherung zur neuen Lehrerrolle zu vermeiden.
- **Keine Zusammenfassung der Ergebnisse** am Ende des Gruppenpuzzles (Schwächung der Lehrerwartung). Eine Besprechung kann nach dem obligatorischen Test erfolgen.
- Evtl. Aufteilen der Note: 50 % Einzelleistung, 50 % Gruppenleistung

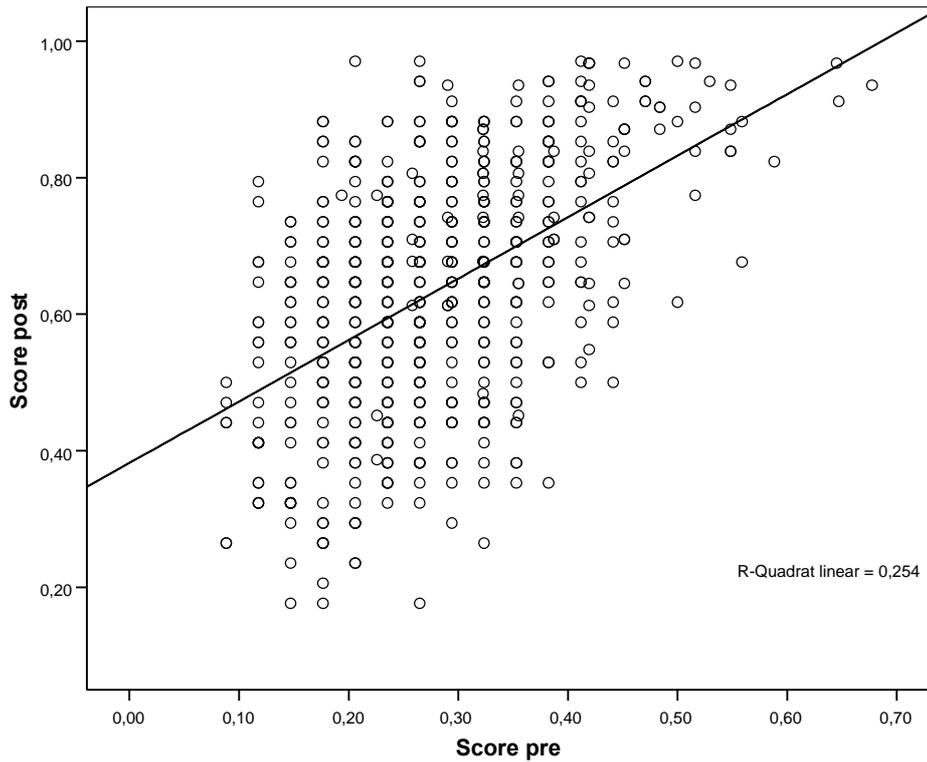
Wichtig: Deutlich machen, dass der Lehrer seinen Schüler die **Lehrerrolle zutraut** (Schaffung des nötigen Selbstvertrauens).

12.3 Statistische Berechnungen Wissenstest

12.3.1 Linearitätsprüfung

Linearitätsprüfung für den Wissenstest

Streudiagramm der Scores für gesamte Untersuchung



Verarbeitete Fälle

	Fälle					
	Eingeschlossen		Ausgeschlossen		Insgesamt	
	N	Prozent	N	Prozent	N	Prozent
Score WT post * Score WT prae	641	96,1%	26	3,9%	667	100,0%

ANOVA-Tabelle

			Quadrat summe	df	Mittel der Quadrate	F	Signifika nz
Score WT post * Score WT prae	Zwischen den Gruppen	(Kombiniert) Linearität	5,463	32	,171	7,676	,000
		Abweichung von der Linearität	4,827	1	4,827	217,038	,000
			,636	31	,021	,922	,590
	Innerhalb der Gruppen		13,522	608	,022		
	Insgesamt		18,985	640			

Zusammenhangsmaße

	R	R-Quadrat	Eta	Eta-Quadrat
Score WT post * Score WT prae	,504	,254	,536	,288

12.3.2 Test auf Normalverteilung

Test auf Normalverteilung der Residuen WT, nach Gruppen VU

Kolmogorov-Smirnov-Anpassungstest			Residuen prae post
FU	N		33
	Parameter der Normalverteilung(a,b)	Mittelwert	-0,3130787
		Standardabweichung	1,11581330
	Extremste Differenzen	Absolut	0,076
		Positiv	0,064
		Negativ	-0,076
	Kolmogorov-Smirnov-Z		0,436
Asymptotische Signifikanz (2-seitig)		0,991	
GP	N		35
	Parameter der Normalverteilung(a,b)	Mittelwert	0,2951885
		Standardabweichung	0,76455232
	Extremste Differenzen	Absolut	0,096
		Positiv	0,096
		Negativ	-0,076
	Kolmogorov-Smirnov-Z		0,566
Asymptotische Signifikanz (2-seitig)		0,905	

HU I

Kolmogorov-Smirnov-Anpassungstest			Residuen prae post
FU	N		78
	Parameter der Normalverteilung(a,b)	Mittelwert	-0,2336567
		Standardabweichung	1,00472558
	Extremste Differenzen	Absolut	0,067
		Positiv	0,067
		Negativ	-0,044
	Kolmogorov-Smirnov-Z		0,594
Asymptotische Signifikanz (2-seitig)		0,872	
GP	N		80
	Parameter der Normalverteilung(a,b)	Mittelwert	0,2278152
		Standardabweichung	0,94037506
	Extremste Differenzen	Absolut	0,146
		Positiv	0,063
		Negativ	-0,146
	Kolmogorov-Smirnov-Z		1,306
Asymptotische Signifikanz (2-seitig)		0,066	

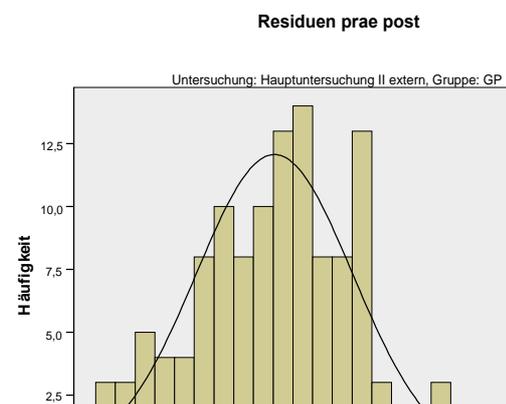
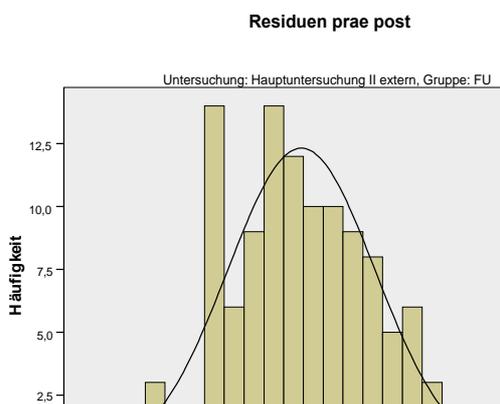
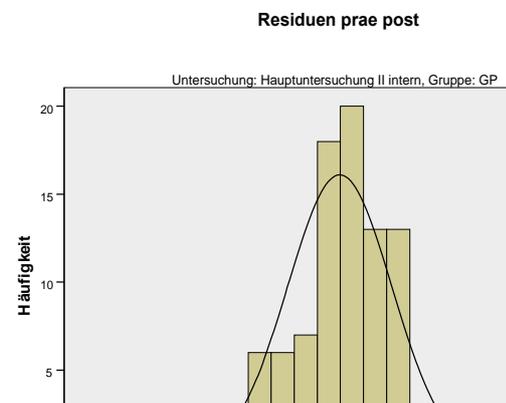
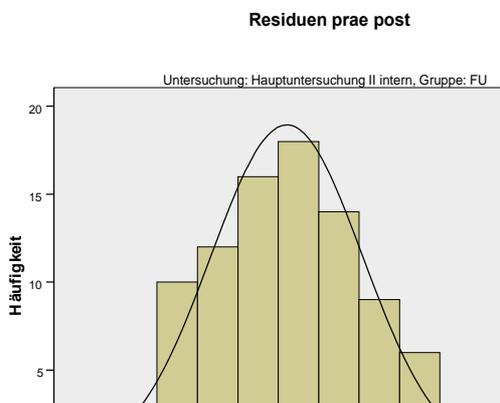
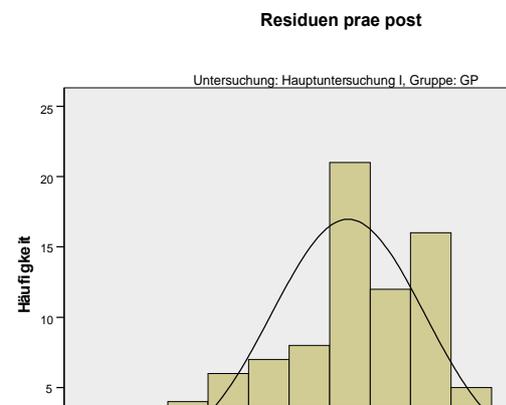
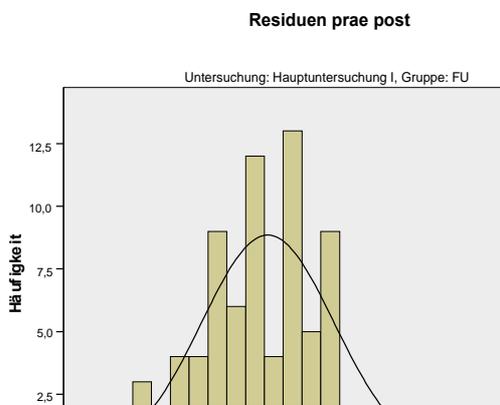
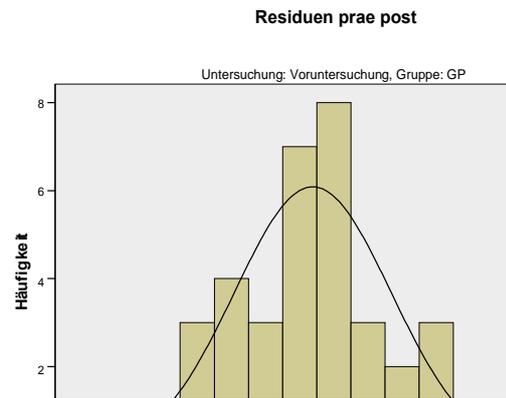
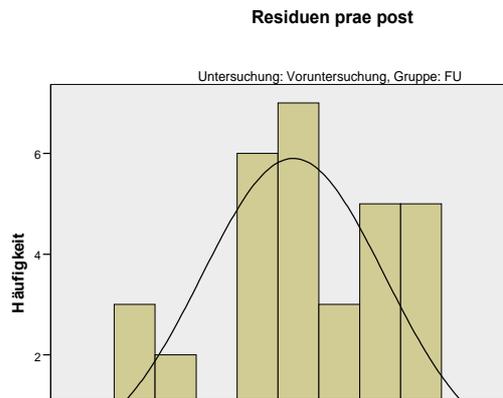
HU II intern

Kolmogorov-Smirnov-Anpassungstest			Residuen prae post
FU	N		89
	Parameter der Normalverteilung(a,b)	Mittelwert	-0,3949130
		Standardabweichung	0,93746016
	Extremste Differenzen	Absolut	0,063
		Positiv	0,063
		Negativ	-0,047
	Kolmogorov-Smirnov-Z		0,598
Asymptotische Signifikanz (2-seitig)		0,867	
GP	N		91
	Parameter der Normalverteilung(a,b)	Mittelwert	0,3862335
		Standardabweichung	0,90189720
	Extremste Differenzen	Absolut	0,093
		Positiv	0,062
		Negativ	-0,093
	Kolmogorov-Smirnov-Z		0,892
Asymptotische Signifikanz (2-seitig)		0,404	

HU II extern

Kolmogorov-Smirnov-Anpassungstest			Residuen prae post
FU	N		115
	Parameter der Normalverteilung(a,b)	Mittelwert	-0,2767387
		Standardabweichung	0,93066282
	Extremste Differenzen	Absolut	0,046
		Positiv	0,046
		Negativ	-0,045
	Kolmogorov-Smirnov-Z		0,493
Asymptotische Signifikanz (2-seitig)		0,968	

Normalverteilungskurven der Residuen



12.3.3 Reliabilität

Reliabilität Wissenstest post			
VU			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	67	90,5
	Ausgeschlossen(a)	7	9,5
	Insgesamt	74	100,0
Variablen in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,793	31		
HU I			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	144	89,4
	Ausgeschlossen(a)	17	10,6
	Insgesamt	161	100,0
Variablen in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,728	34		
HU II intern			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	180	98,9
	Ausgeschlossen(a)	2	1,1
	Insgesamt	182	100,0
Variablen in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,798	34		
HU II extern			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	215	86,0
	Ausgeschlossen(a)	35	14,0
	Insgesamt	250	100,0
Variablen in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,818	34		

Reliabilität WT follow up			
VU			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	66	89,2
	Ausgeschlossen(a)	8	10,8
	Insgesamt	74	100,0
in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,79	31		
HU I			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	139	86,3
	Ausgeschlossen(a)	22	13,7
	Insgesamt	161	100,0
in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,82	34		
HU II intern			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	172	94,5
	Ausgeschlossen(a)	10	5,5
	Insgesamt	182	100,0
in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,81	34		
HU II extern			
Zusammenfassung der Fallverarbeitung			
		Anzahl	%
Fälle	Gültig	214	85,6
	Ausgeschlossen(a)	36	14,4
	Insgesamt	250	100,0
in der Prozedur.			
Zuverlässigkeitsstatistik			
Cronbachs Alpha	Anzahl der Items		
0,80	34		

12.3.4 Effektstärken

Effektgrößen Wissenstest

VU				Effektgröße d
Deskriptive Statistik	N	Mittelwert	Standardabweichung	
Score WT prae	70	0,40	0,101849377	$d = \frac{M_{IG} - M_{KG}}{S_{gesamt}}$ 0,22 0,69 0,62 0,61 0,57 0,18
Score WT post	71	0,77	0,138693601	
Score WT follow up	71	0,70	0,156230091	
Residuen prae post	68	0,00	0,992509258	
Residuen prae follow up	67	0,00	0,992395327	
Residuen post follow up	70	0,00	0,992727176	
Gültige Werte (Listenweise)	67			

Gruppenstatistiken					
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	34	0,384250474	0,10212254	0,017513871
	GP	36	0,406810036	0,101798244	0,016966374
Score WT post	FU	35	0,725345622	0,154381724	0,026095274
	GP	36	0,821684588	0,103019988	0,017169998
Score WT follow up	FU	36	0,653225806	0,170932959	0,028488826
	GP	35	0,750230415	0,123541341	0,020882298
Residuen prae post	FU	33	-0,313078737	1,115813304	0,194238165
	GP	35	0,295188523	0,764552319	0,129232929
Residuen prae follow up	FU	33	-0,286010484	0,970273733	0,168902977
	GP	34	0,277598411	0,946359348	0,162299289
Residuen post follow up	FU	35	-0,089668614	1,11864605	0,189085694
	GP	35	0,089668614	0,855615058	0,144625341

HU I				Effektgröße d
Deskriptive Statistik	N	Mittelwert	Standardabweichung	
Score WT prae	158	0,256142964	0,085107792	0,08 0,42 0,04 0,46 0,08 0,16
Score WT post	158	0,624162323	0,135966993	
Score WT follow up	149	0,517173312	0,178482853	
Residuen prae post	158	1,7009E-15	0,996810199	
Residuen prae follow up	149	-7,63278E-17	0,996615896	
Residuen post follow up	149	-1,71304E-15	0,996615896	
Gültige Werte (Listenweise)	149			

Gruppenstatistiken					
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	78	0,259426848	0,087627598	0,009921872
	GP	80	0,252941176	0,083004549	0,009280191
Score WT post	FU	78	0,595399698	0,143320389	0,016227838
	GP	80	0,652205882	0,122906899	0,013741409
Score WT follow up	FU	72	0,513480392	0,183572737	0,021634255
	GP	77	0,520626432	0,174724654	0,019911722
Residuen prae post	FU	78	-0,233656652	1,004725578	0,11376277
	GP	80	0,227815235	0,940375062	0,105137128
Residuen prae follow up	FU	72	-0,041114835	0,967139825	0,113978521
	GP	77	0,03844504	1,028245575	0,117179458
Residuen post follow up	FU	72	0,083359644	1,01971479	0,120174541
	GP	77	-0,07794668	0,974711566	0,111078692

HU II intern				Effektgröße d
Deskriptive Statistik	N	Mittelwert	Standardabweichung	
Score WT prae	181	0,292167696	0,083339228	0,18 0,69 0,58 0,78 0,69 0,18
Score WT post	180	0,679411765	0,160496779	
Score WT follow up	180	0,588562092	0,175206031	
Residuen prae post	180	-2,71051E-15	0,997202792	
Residuen prae follow up	180	-9,83588E-16	0,997202792	
Residuen post follow up	180	1,35135E-15	0,997202792	
Gültige Werte (Listenweise)	180			

Gruppenstatistiken					
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	89	0,299735625	0,077166262	0,008179607
	GP	92	0,284846547	0,088711208	0,009248783
Score WT post	FU	89	0,623265036	0,145842322	0,015459255
	GP	91	0,734324499	0,1557837	0,016330569
Score WT follow up	FU	89	0,537012558	0,162463736	0,017221122
	GP	91	0,638978668	0,173341869	0,018171165
Residuen prae post	FU	89	-0,394912951	0,937460159	0,099370578
	GP	91	0,386233545	0,901897201	0,094544516
Residuen prae follow up	FU	89	-0,347668066	0,938034604	0,099431469
	GP	91	0,340027009	0,938610316	0,098393096
Residuen post follow up	FU	89	-0,089939898	0,974728307	0,103320994
	GP	91	0,087963197	1,016352759	0,106542719

HU II extern				Effektgröße d
Deskriptive Statistik	N	Mittelwert	Standardabweichung	
Score WT prae	240	0,247303922	0,084783225	0,05 0,48 0,45 0,54 0,48 0,18
Score WT post	235	0,559574468	0,174573416	
Score WT follow up	225	0,461176471	0,172295678	
Residuen prae post	235	-7,13405E-17	0,99786096	
Residuen prae follow up	225	-1,26982E-15	0,99776536	
Residuen post follow up	225	-3,02536E-15	0,99776536	
Gültige Werte (Listenweise)	225			

Gruppenstatistiken					
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	120	0,245098039	0,088878307	0,008113442
	GP	120	0,249509804	0,08079374	0,007375426
Score WT post	FU	115	0,516879795	0,159915828	0,014912228
	GP	120	0,600490196	0,178826884	0,016324586
Score WT follow up	FU	111	0,422098569	0,1639329	0,015559822
	GP	114	0,499226006	0,17240948	0,016147627
Residuen prae post	FU	115	-0,276738671	0,930662818	0,086784755
	GP	120	0,265207893	0,991617748	0,090521901
Residuen prae follow up	FU	111	-0,243328392	0,935346252	0,088779137
	GP	114	0,236925013	1,003501091	0,093986488
Residuen post follow up	FU	111	-0,091989055	0,997576663	0,094685787
	GP	114	0,08956829	0,994136105	0,093109377

12.3.5 t-Test & U-Test

Voruntersuchung

Deskriptive Statistik			
	N	Mittelwert	Standardabweichung
Score WT prae	70	0,3959	0,10185
Score WT post	71	0,7742	0,13869
Score WT follow up	71	0,7010	0,15623
Residuen prae post	68	0,000000	0,99250926
Residuen prae follow up	67	0,000000	0,99239533
Residuen post follow up	70	0,000000	0,99272718
Gültige Werte (Listenweise)	67		

T-Test

Gruppenstatistiken				
Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	34	0,38	0,10212
	GP	36	0,41	0,10180
Score WT post	FU	35	0,73	0,15438
	GP	36	0,82	0,10302
Score WT follow up	FU	36	0,65	0,17093
	GP	35	0,75	0,12354
Residuen prae post	FU	33	-0,313	1,11581330
	GP	35	0,295	0,76455232
Residuen prae follow up	FU	33	-0,286	0,97027373
	GP	34	0,278	0,94635935
Residuen post follow up	FU	35	-0,090	1,11864605
	GP	35	0,090	0,85561506

Test bei unabhängigen Stichproben											
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit							
		F	Signifikanz	t	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der		
										Untere	Obere
Score WT prae	Varianzen sind gleich	0,305	0,583	-0,925	68	0,358	-0,02256	0,02438	-0,07121	0,02609	
	Varianzen sind nicht gleich			-0,925	67,746	0,358	-0,02256	0,02438	-0,07122	0,02610	
Score WT post	Varianzen sind gleich	5,510	0,022	-3,101	69	0,003	-0,09634	0,03107	-0,15831	-0,03436	
	Varianzen sind nicht gleich			-3,084	59,059	0,003	-0,09634	0,03124	-0,15884	-0,03383	
Score WT follow up	Varianzen sind gleich	3,946	0,051	-2,734	69	0,008	-0,09700	0,03548	-0,16779	-0,02622	
	Varianzen sind nicht gleich			-2,746	63,765	0,008	-0,09700	0,03532	-0,16757	-0,02643	
Residuen prae post	Varianzen sind gleich	5,237	0,025	-2,635	66	0,010	-0,60826726	0,23079984	-1,06907400	-0,14746052	
	Varianzen sind nicht gleich			-2,607	56,231	0,012	-0,60826726	0,23330155	-1,07558411	-0,14095041	
Residuen prae follow up	Varianzen sind gleich	0,141	0,708	-2,407	65	0,019	-0,56360890	0,23415336	-1,03124544	-0,09597235	
	Varianzen sind nicht gleich			-2,406	64,802	0,019	-0,56360890	0,23424192	-1,03144338	-0,09576841	
Residuen post follow up	Varianzen sind gleich	3,786	0,056	-0,753	68	0,454	-0,17933723	0,23805438	-0,65436735	0,29569290	
	Varianzen sind nicht gleich			-0,753	63,636	0,454	-0,17933723	0,23805438	-0,65495765	0,29626320	

Mann-Whitney-Test

Ränge				
Gruppe	N	Mittlerer Rang	Rangsumme	
Score WT prae	FU	34	34,01	1.156,50
	GP	36	36,90	1.328,50
	Gesamt	70		
Score WT post	FU	35	29,40	1.029,00
	GP	36	42,42	1.527,00
	Gesamt	71		
Score WT follow up	FU	36	29,97	1.079,00
	GP	35	42,20	1.477,00
	Gesamt	71		
Residuen prae post	FU	33	29,06	959,00
	GP	35	39,63	1.387,00
	Gesamt	68		
Residuen prae follow up	FU	33	28,47	939,50
	GP	34	39,37	1.338,50
	Gesamt	67		
Residuen post follow up	FU	35	33,09	1.158,00
	GP	35	37,91	1.327,00
	Gesamt	70		

Statistik für Test(a)						
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	561,500	399,000	413,000	398,000	378,500	528,000
Wilcoxon-W	1.156,500	1.029,000	1.079,000	959,000	939,500	1.158,000
Z	-0,597	-2,665	-2,506	-2,203	-2,289	-0,993
Asymptotische Signifikanz (2-seitig)	0,550	0,008	0,012	0,028	0,022	0,321

a. Gruppenvariable: Gruppe

Explorative Datenanalyse

Verarbeitete Fälle							
Gruppe		Fälle					
		Gültig		Fehlend		Gesamt	
		N	Prozent	N	Prozent	N	Prozent
Residuen prae post	FU	33	89,2%	4	10,8%	37	100,0%
	GP	35	94,6%	2	5,4%	37	100,0%

Tests auf Normalverteilung							
Gruppe		Kolmogorov-Smirnov(a)			Shapiro-Wilk		
		Statistik	#	Signifikanz	Statistik	#	Signifikanz
Residuen prae post	FU	0,076	33	0,200	0,961	33	0,274
	GP	0,096	35	0,200	0,987	35	0,952

. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

t-Test & U-Test

HU I

Deskriptive Statistik			
	N	Mittelwert	Standardabweichung
Score WT prae	158	0,2561	0,08511
Score WT post	158	0,6242	0,13597
Score WT follow up	149	0,5172	0,17848
Residuen prae post	158	0,000000	0,99661590
Residuen prae follow up	149	0,000000	0,99661590
Residuen post follow up	149	0,000000	0,99661590
Gültige Werte (Listenweise)	149		

Gruppenstatistiken				
Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	78	0,2594	0,08763
	GP	80	0,2529	0,08300
Score WT post	FU	78	0,5954	0,14332
	GP	80	0,6522	0,12291
Score WT follow up	FU	72	0,5135	0,16357
	GP	77	0,5206	0,17472
Residuen prae post	FU	78	-0,2336667	1,00472558
	GP	80	0,2278152	0,94037506
Residuen prae follow up	FU	72	-0,0411148	0,96713982
	GP	77	0,0384450	1,02824557
Residuen post follow up	FU	72	0,0833596	1,01971479
	GP	77	-0,0779467	0,97471157

Test bei unabhängigen Stichproben											
		Levene-Test der Varianzgleichheit			T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sign. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der		
										Untere	Obere
Score WT prae	Varianzen sind gleich	0,194	0,660	0,478	156	0,634	0,00649	0,01358	-0,02033	0,03330	
	Varianzen sind nicht gleich			0,477	155,018	0,634	0,00649	0,01359	-0,02035	0,03332	
Score WT post	Varianzen sind gleich	1,428	0,234	-2,677	156	0,008	-0,05681	0,02122	-0,09873	-0,01488	
	Varianzen sind nicht gleich			-2,671	151,228	0,008	-0,05681	0,02126	-0,09882	-0,01479	
Score WT follow up	Varianzen sind gleich	0,090	0,764	-0,243	147	0,808	-0,00715	0,02935	-0,06516	0,05086	
	Varianzen sind nicht gleich			-0,243	145,020	0,808	-0,00715	0,02940	-0,06526	0,05097	
Residuen prae post	Varianzen sind gleich	0,222	0,638	-2,982	156	0,003	-0,46147189	0,15477535	-0,76719770	-0,15574608	
	Varianzen sind nicht gleich			-2,979	154,704	0,003	-0,46147189	0,15490572	-0,76747526	-0,15546851	
Residuen prae follow up	Varianzen sind gleich	0,524	0,470	-0,486	147	0,628	-0,07955987	0,16380733	-0,40328138	0,24416163	
	Varianzen sind nicht gleich			-0,487	146,994	0,627	-0,07955987	0,16346905	-0,40261295	0,24349321	
Residuen post follow up	Varianzen sind gleich	0,228	0,634	0,987	147	0,325	0,16130632	0,16339797	-0,16160619	0,48421884	
	Varianzen sind nicht gleich			0,986	145,159	0,326	0,16130632	0,16364717	-0,16213271	0,48474536	

Mann-Whitney-Test

Ränge			
Gruppe	N	Mittlerer Rang	Rangsumme
Score WT prae	FU	81,34	6.344,50
	GP	77,71	6.216,50
	Gesamt		12.561,00
Score WT post	FU	69,19	5.397,00
	GP	89,55	7.164,00
	Gesamt		12.561,00
Score WT follow up	FU	75,44	5.432,00
	GP	74,58	5.743,00
	Gesamt		11.175,00
Residuen prae post	FU	67,22	5.243,00
	GP	91,48	7.318,00
	Gesamt		12.561,00
Residuen prae follow up	FU	74,85	5.389,50
	GP	75,14	5.785,50
	Gesamt		11.175,00
Residuen post follow up	FU	78,26	5.634,50
	GP	71,95	5.540,50
	Gesamt		11.175,00

Statistik für Test(a)						
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	2.976,500	2.316,000	2.740,000	2.162,000	2.761,500	2.537,500
Wilcoxon-W	6.216,500	5.397,000	5.743,000	5.243,000	5.389,500	5.540,500
Z	-0,502	-2,803	-0,122	-3,332	-0,040	-0,891
Asymptotische Signifikanz (2-seitig)	0,616	0,005	0,903	0,001	0,968	0,373

a. Gruppenvariable: Gruppe

Explorative Datenanalyse

Verarbeitete Fälle							
Gruppe	Fälle						
	Gültig		Fehlend		Gesamt		
	N	Prozent	N	Prozent	N	Prozent	
Residuen prae post	FU	78	100,0%	0	0,0%	78	100,0%
	GP	80	96,4%	3	3,6%	83	100,0%

Tests auf Normalverteilung							
Gruppe	Kolmogorov-Smirnov(a)			Shapiro-Wilk			
	Statistik	df	Signifikanz	Statistik	df	Signifikanz	
Residuen prae post	FU	0,067	78	0,200	0,982	78	0,359
	GP	0,146	80	0,000	0,947	80	0,002

*. Dies ist eine untere Grenze der echten Signifikanz.
a. Signifikanzkorrektur nach Lilliefors

t-Test & U-Test

HU II intern

Deskriptive Statistik			
	N	Mittelwert	Standardabweichung
Score WT prae	181	0,2922	0,08334
Score WT post	180	0,6794	0,16050
Score WT follow up	180	0,5886	0,17521
Residuen prae post	180	0,000000	0,99720279
Residuen prae follow up	180	0,000000	0,99720279
Residuen post follow up	180	0,000000	0,99720279
Gültige Werte (Listenweise)	180		

Gruppenstatistiken						
Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes		
Score WT prae	FU	89	0,2997	0,07717	0,00818	
	GP	92	0,2848	0,08871	0,00925	
Score WT post	FU	89	0,6233	0,14584	0,01546	
	GP	91	0,7343	0,15578	0,01633	
Score WT follow up	FU	89	0,5370	0,16246	0,01722	
	GP	91	0,6390	0,17334	0,01817	
Residuen prae post	FU	89	-0,3949130	0,93746016	0,09937058	
	GP	91	0,3862335	0,90189720	0,09454452	
Residuen prae follow up	FU	89	-0,3476681	0,93803460	0,09943147	
	GP	91	0,3400270	0,93861032	0,09839310	
Residuen post follow up	FU	89	-0,0899399	0,97472831	0,10332099	
	GP	91	0,0879632	1,01639276	0,10664272	

Test bei unabhängigen Stichproben											
		Levene-Test der Varianzgleichheit			T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der		
									Untere	Obere	
Score WT prae	Varianzen sind gleich	1,834	0,177	1,203	179	0,231	0,01489	0,01238	-0,00953	0,03931	
	Varianzen sind nicht gleich			1,206	177,030	0,229	0,01489	0,01235	-0,00948	0,03926	
Score WT post	Varianzen sind gleich	0,221	0,639	-4,935	178	0,000	-0,11106	0,02250	-0,15547	-0,06665	
	Varianzen sind nicht gleich			-4,939	177,663	0,000	-0,11106	0,02249	-0,15544	-0,06668	
Score WT follow up	Varianzen sind gleich	0,024	0,878	-4,070	178	0,000	-0,10197	0,02505	-0,15141	-0,05253	
	Varianzen sind nicht gleich			-4,073	177,680	0,000	-0,10197	0,02504	-0,15137	-0,05256	
Residuen prae post	Varianzen sind gleich	0,621	0,432	-5,698	178	0,000	-0,78114650	0,13710191	-1,05170078	-0,51059221	
	Varianzen sind nicht gleich			-5,695	177,340	0,000	-0,78114650	0,13710114	-1,05182455	-0,51046844	
Residuen prae follow up	Varianzen sind gleich	0,275	0,601	-4,916	178	0,000	-0,68769508	0,13988598	-0,96374339	-0,41164676	
	Varianzen sind nicht gleich			-4,916	177,916	0,000	-0,68769508	0,13988502	-0,96374239	-0,41164776	
Residuen post follow up	Varianzen sind gleich	0,567	0,452	-1,198	178	0,232	-0,17790310	0,14848286	-0,47091633	0,11511014	
	Varianzen sind nicht gleich			-1,199	177,933	0,232	-0,17790310	0,14841354	-0,47078030	0,11497411	

Mann-Whitney-Test

Ränge					
Gruppe	N	Mittlerer Rang	Rangsumme		
Score WT prae	FU	89	95,53	8.502,50	
	GP	92	86,61	7.968,50	
	Gesamt	181			
Score WT post	FU	89	71,23	6.339,50	
	GP	91	109,35	9.950,50	
	Gesamt	180			
Score WT follow up	FU	89	75,16	6.689,00	
	GP	91	105,51	9.601,00	
	Gesamt	180			
Residuen prae post	FU	89	68,98	6.139,50	
	GP	91	111,54	10.150,50	
	Gesamt	180			
Residuen prae follow up	FU	89	72,42	6.445,00	
	GP	91	108,19	9.845,00	
	Gesamt	180			
Residuen post follow up	FU	89	86,29	7.679,50	
	GP	91	94,62	8.610,50	
	Gesamt	180			

Statistik für Test(a)						
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	3.690,500	2.334,500	2.684,000	2.134,500	2.440,000	3.674,500
Wilcoxon-W	7.968,500	6.339,500	6.689,000	6.139,500	6.445,000	7.679,500
Z	-1,153	-4,915	-3,913	-5,479	-4,605	-1,073
Asymptotische Signifikanz (2-seitig)	0,249	0,000	0,000	0,000	0,000	0,283

a. Gruppvariable: Gruppe

Explorative Datenanalyse

Verarbeitete Fälle							
Gruppe		Fälle					
		Gültig		Fehlend		Gesamt	
		N	Prozent	N	Prozent	N	Prozent
Residuen prae post	FU	89	98,9%	1	1,1%	90	100,0%
	GP	91	98,9%	1	1,1%	92	100,0%

Tests auf Normalverteilung							
Gruppe		Kolmogorov-Smirnov(a)			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
Residuen prae post	FU	0,063	89	0,200	0,981	89	0,216
	GP	0,093	91	0,048	0,956	91	0,003

*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

t-Test & U-Test

HU II extern

Deskriptive Statistik			
	N	Mittelwert	Standardabweichung
Score WT prae	240	0,2473	0,08478
Score WT post	235	0,5596	0,17457
Score WT follow up	225	0,4612	0,17230
Residuen prae post	235	0,000000	0,99786096
Residuen prae follow up	225	0,000000	0,99776536
Residuen post follow up	225	0,000000	0,99776536
Gültige Werte (Listenweise)	225		

T-Test

Gruppenstatistiken					
Gruppe		N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score WT prae	FU	120	0,25	0,08888	0,00811
	GP	120	0,25	0,08079	0,00738
Score WT post	FU	115	0,52	0,15992	0,01491
	GP	120	0,60	0,17883	0,01632
Score WT follow up	FU	111	0,42	0,16393	0,01556
	GP	114	0,50	0,17241	0,01615
Residuen prae post	FU	115	-0,277	0,93066282	0,08678476
	GP	120	0,265	0,99161775	0,09052190
Residuen prae follow up	FU	111	-0,243	0,93534625	0,08877914
	GP	114	0,237	1,00350109	0,09398649
Residuen post follow up	FU	111	-0,092	0,99757666	0,09468579
	GP	114	0,090	0,99413610	0,09310938

Test bei unabhängigen Stichproben											
		Levene-Test der Varianzgleichheit			T-Test für die Mittelwertgleichheit						
		z	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der		
									Untere	Oberer	
Score WT prae	Varianzen sind gleich	1,171	0,280	-0,402	238	0,688	-0,00441	0,01096	-0,02601	0,01719	
	Varianzen sind nicht gleich			-0,402	235,868	0,688	-0,00441	0,01096	-0,02601	0,01719	
Score WT post	Varianzen sind gleich	1,348	0,247	-3,773	233	0,000	-0,08361	0,02216	-0,12728	-0,03994	
	Varianzen sind nicht gleich			-3,782	231,902	0,000	-0,08361	0,02211	-0,12717	-0,04005	
Score WT follow up	Varianzen sind gleich	0,193	0,661	-3,437	223	0,001	-0,07713	0,02244	-0,12135	-0,03291	
	Varianzen sind nicht gleich			-3,439	222,876	0,001	-0,07713	0,02242	-0,12132	-0,03294	
Residuen prae post	Varianzen sind gleich	0,426	0,515	-4,316	233	0,000	-0,54194656	0,12557247	-0,78934914	-0,29454399	
	Varianzen sind nicht gleich			-4,322	232,900	0,000	-0,54194656	0,12540259	-0,78901498	-0,29487814	
Residuen prae follow up	Varianzen sind gleich	1,261	0,263	-3,711	223	0,000	-0,48025340	0,12940890	-0,73527422	-0,22523259	
	Varianzen sind nicht gleich			-3,715	222,579	0,000	-0,48025340	0,12928726	-0,73503713	-0,22546968	
Residuen post follow up	Varianzen sind gleich	0,019	0,891	-1,367	223	0,173	-0,18155735	0,13278977	-0,44324069	0,08012600	
	Varianzen sind nicht gleich			-1,367	222,796	0,173	-0,18155735	0,13273591	-0,44325411	0,08013941	

Mann-Whitney-Test

Ränge				
Gruppe		N	Mittlerer Rang	Rangsumme
Score WT prae	FU	120	118,72	14.246,00
	GP	120	122,28	14.674,00
	Gesamt	240		
Score WT post	FU	115	100,89	11.602,50
	GP	120	134,40	16.127,50
	Gesamt	235		
Score WT follow up	FU	111	97,78	10.853,50
	GP	114	127,82	14.571,50
	Gesamt	225		
Residuen prae post	FU	115	99,10	11.397,00
	GP	120	136,11	16.333,00
	Gesamt	235		
Residuen prae follow up	FU	111	97,27	10.796,50
	GP	114	128,32	14.628,50
	Gesamt	225		
Residuen post follow up	FU	111	109,57	12.162,50
	GP	114	116,34	13.262,50
	Gesamt	225		

Statistik für Test(a)						
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	6.986,000	4.932,500	4.637,500	4.727,000	4.580,500	5.946,500
Wilcoxon-W	14.246,000	11.602,500	10.853,500	11.397,000	10.796,500	12.162,500
Z	-0,401	-3,782	-3,466	-4,171	-3,578	-0,779
Asymptotische Signifikanz (2-seitig)	0,688	0,000	0,001	0,000	0,000	0,436

a. Gruppenvariable: Gruppe

Explorative Datenanalyse

Verarbeitete Fälle							
		Fälle					
		Gültig		Fehlend		Gesamt	
Gruppe		N	Prozent	N	Prozent	N	Prozent
Residuen prae post	FU	115	92,7%	9	7,3%	124	100,0%
	GP	120	95,2%	6	4,8%	126	100,0%

Tests auf Normalverteilung							
		Kolmogorov-Smirnov(a)			Shapiro-Wilk		
Gruppe		Statistik	df	Signifikanz	Statistik	df	Signifikanz
Residuen prae post	FU	0,046	115	0,200	0,992	115	0,706
	GP	0,064	120	0,200	0,986	120	0,245

*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

t-Test & U-Test Klassenvergleich

Klasse 1 (Ri 10b)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	56.5	54	39	61.500	41	45	
Wilcoxon-W	122.5	132	117	127.500	107	123	
Z	-0.589954755	-1.042864169	-1.915695725	-0.277	-1.539023878	-1.558845727	
Asymptotische Signifikanz (2-seitig)	0.555220983	0.297011207	0.055403837	0.782	0.123798467	0.119032898	
Exakte Signifikanz [2*(1-seitig Sig.)]	0.565783927	0.31858517	0.059657061	0.786	0.133548508	0.1276879	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	11	0.34	0.097	0.029296173	
	GP	12	0.39	0.121	0.034842692	
Score WT post	FU	12	0.68	0.193	0.055707756	
	GP	12	0.77	0.110	0.031630836	
Score WT follow up	FU	12	0.55	0.178	0.051287054	
	GP	12	0.68	0.138	0.039706933	
Residuen prae post	FU	11	-0.186	1.316	0.396685534	
	GP	12	0.171	0.518	0.149676951	
Residuen prae follow up	FU	11	-0.351	0.923	0.278288148	
	GP	12	0.322	0.948	0.273609062	
Residuen post follow up	FU	12	-0.238	0.849	0.245089112	
	GP	12	0.238	1.075	0.310278916	

Klasse 2 (Ri 10d)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	53	35.5	52.5	37.000	59	68.5	
Wilcoxon-W	119	113.5	130.5	103.000	125	146.5	
Z	-0.805687034	-2.121675235	-1.132749133	-1.787	-0.430928686	-0.202116537	
Asymptotische Signifikanz (2-seitig)	0.420423328	0.03386502	0.257319818	0.074	0.666521679	0.83982562	
Exakte Signifikanz [2*(1-seitig Sig.)]	0.44913681	0.033241425	0.265669584	0.079	0.694702525	0.84283599	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	11	0.38	0.115	0.034648811	
	GP	12	0.42	0.110	0.031630836	
Score WT post	FU	12	0.71	0.145	0.041833849	
	GP	12	0.84	0.105	0.030233899	
Score WT follow up	FU	12	0.69	0.178	0.051389423	
	GP	12	0.77	0.115	0.033329786	
Residuen prae post	FU	11	-0.413	0.988	0.282892481	
	GP	12	0.378	0.984	0.285306348	
Residuen prae follow up	FU	11	-0.172	1.029	0.310285379	
	GP	12	0.158	0.943	0.272283616	
Residuen post follow up	FU	12	0.078	1.271	0.368939582	
	GP	12	-0.078	0.609	0.175800316	

Klasse 3 (Vo 10b)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	62.5	38.5	33	33.500	26.5	45.5	
Wilcoxon-W	140.5	104.5	111	99.500	92.5	111.5	
Z	-0.553437752	-1.712075218	-2.050348139	-1.774	-2.007561464	-0.986089387	
Asymptotische Signifikanz (2-seitig)	0.579963694	0.086882808	0.04033047	0.076	0.044668913	0.324089251	
Exakte Signifikanz [2*(1-seitig Sig.)]	0.589875732	0.090842392	0.043879125	0.076	0.042963744	0.331649259	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	12	0.42	0.083	0.024098325	
	GP	12	0.42	0.076	0.021913886	
Score WT post	FU	11	0.79	0.100	0.030249377	
	GP	12	0.85	0.081	0.023322542	
Score WT follow up	FU	12	0.71	0.115	0.033329786	
	GP	11	0.81	0.078	0.023423725	
Residuen prae post	FU	11	-0.370	0.976	0.294154428	
	GP	11	0.370	0.865	0.260659549	
Residuen prae follow up	FU	11	-0.423	0.950	0.286463963	
	GP	10	0.465	0.806	0.254932971	
Residuen post follow up	FU	11	-0.239	1.121	0.337940193	
	GP	11	0.239	0.786	0.236978124	

Klasse 4 (De 10b+c)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	138.5	114.5	107.5	118.500	117	114	
Wilcoxon-W	291.5	267.5	243.5	271.500	253	250	
Z	-0.484685414	-1.273546437	-1.02988627	-1.139	-0.684530655	-0.792813267	
Asymptotische Signifikanz (2-seitig)	0.627899514	0.202824205	0.303016451	0.255	0.493640139	0.427886626	
Exakte Signifikanz [2*(1-seitig Sig.)]	0.636362121	0.206643257	0.309032541	0.258	0.510218928	0.443784935	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	17	0.21	0.065	0.015654811	
	GP	18	0.29	0.065	0.015374194	
Score WT post	FU	17	0.57	0.159	0.038506857	
	GP	18	0.63	0.119	0.028056252	
Score WT follow up	FU	17	0.36	0.149	0.037348309	
	GP	17	0.49	0.126	0.030467782	
Residuen prae post	FU	17	-0.187	1.106	0.268144744	
	GP	18	0.176	0.850	0.200425983	
Residuen prae follow up	FU	16	-0.181	1.085	0.271150564	
	GP	17	0.170	0.878	0.212991729	
Residuen post follow up	FU	16	-0.126	0.805	0.20117084	
	GP	17	0.119	1.140	0.276472928	

t-Test & U-Test Klassenvergleich Fortsetzung

Klasse 5 (Ki 10a)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	78,5	73	61	76.500	60	55,5	
Wilcoxon-W	169,5	164	152	167.500	151	146,5	
Z	-0,30971068	-0,592791381	-0,928615201	-0,410	-0,970970928	-1,224074081	
Asymptotische Signifikanz (2-seitig)	0,75678098	0,553320777	0,353088541	0,682	0,327544935	0,220924283	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,762310828	0,578767187	0,37602023	0,687	0,347492645	0,225413534	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	13	0,24	0,086	0,023818419	
	GP	13	0,25	0,080	0,02222498	
Score WT post	FU	13	0,53	0,175	0,048655686	
	GP	13	0,55	0,100	0,027632083	
Score WT follow up	FU	12	0,54	0,177	0,051006952	
	GP	13	0,49	0,135	0,037461491	
Residuen prae post	FU	13	-0,062	1,053	0,291934711	
	GP	13	0,062	0,940	0,260735616	
Residuen prae follow up	FU	12	0,184	1,062	0,30655686	
	GP	13	-0,170	0,904	0,25082862	
Residuen post follow up	FU	12	0,299	1,090	0,314647878	
	GP	13	-0,276	0,809	0,224507213	

Klasse 6 (Ru 10b)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	44,5	71,5	52	53.500	44	55,5	
Wilcoxon-W	135,5	162,5	143	144.500	135	146,5	
Z	-2,085796054	-0,671621651	-1,672970053	-1,590	-2,077278198	-1,487688185	
Asymptotische Signifikanz (2-seitig)	0,036997101	0,501824591	0,094333192	0,112	0,03777589	0,136833145	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,038694498	0,511438186	0,101430879	0,113	0,038694498	0,138868527	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	13	0,31	0,079	0,021935203	
	GP	13	0,25	0,062	0,017155827	
Score WT post	FU	13	0,66	0,136	0,037756378	
	GP	13	0,68	0,126	0,034903277	
Score WT follow up	FU	13	0,66	0,106	0,02951311	
	GP	13	0,72	0,133	0,037014755	
Residuen prae post	FU	13	-0,253	0,954	0,264497828	
	GP	13	0,253	0,975	0,270534932	
Residuen prae follow up	FU	13	-0,361	0,796	0,220735744	
	GP	13	0,361	1,042	0,288871103	
Residuen post follow up	FU	13	-0,224	1,017	0,282169611	
	GP	13	0,224	0,925	0,256648011	

Klasse 7 (Ru 10d)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	62	62	58	53.500	66	43,5	
Wilcoxon-W	140	140	136	131.500	144	121,5	
Z	-0,579622813	-0,580386313	-0,810584405	-1,069	-0,346485493	-1,646164146	
Asymptotische Signifikanz (2-seitig)	0,562169005	0,561654133	0,417604376	0,285	0,728977885	0,099729969	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,589875732	0,589875732	0,44283318	0,291	0,755284828	0,100530443	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	12	0,28	0,097	0,027916147	
	GP	12	0,26	0,086	0,024775741	
Score WT post	FU	12	0,61	0,124	0,035809029	
	GP	12	0,64	0,129	0,037244346	
Score WT follow up	FU	12	0,62	0,138	0,038796281	
	GP	12	0,58	0,194	0,056139563	
Residuen prae post	FU	12	-0,216	0,886	0,25569815	
	GP	12	0,216	1,055	0,304646562	
Residuen prae follow up	FU	12	0,082	0,863	0,249195372	
	GP	12	-0,082	1,114	0,321485841	
Residuen post follow up	FU	12	0,285	1,038	0,299597143	
	GP	12	-0,285	0,864	0,249337931	

Klasse 8 (Li 10b)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	77	14	46	11.000	49	62,5	
Wilcoxon-W	168	92	112	89.000	115	140,5	
Z	-0,054891224	-3,49528527	-1,235194239	-3,645	-1,046277673	-0,215463343	
Asymptotische Signifikanz (2-seitig)	0,956225123	0,000473555	0,216758223	0,000	0,29543285	0,829406089	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,978697383	0,000193835	0,235116613	0,000	0,316385593	0,832752252	

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	12	0,26	0,069	0,019980308	
	GP	13	0,27	0,088	0,02434975	
Score WT post	FU	12	0,59	0,082	0,023694091	
	GP	13	0,74	0,067	0,018679436	
Score WT follow up	FU	11	0,49	0,108	0,032550113	
	GP	12	0,54	0,120	0,034599018	
Residuen prae post	FU	12	-0,719	0,807	0,233061081	
	GP	13	0,664	0,569	0,157870122	
Residuen prae follow up	FU	11	-0,213	0,941	0,283733053	
	GP	12	0,195	1,009	0,291185114	
Residuen post follow up	FU	11	0,086	0,980	0,29548507	
	GP	12	-0,079	1,011	0,291763988	

t-Test & U-Test Klassenvergleich Fortsetzung

Klasse 9 (Li 10c)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	52,5	40,5	37	41.000	34	33	33
Wilcoxon-W	118,5	106,5	92	107.000	89	88	88
Z	-0,531662208	-1,319649208	-0,267108873	-1,281	-0,533389288	-0,621966322	-0,621966322
Asymptotische Signifikanz (2-seitig)	0,59495971	0,186952165	0,789385336	0,200	0,593764131	0,533964013	0,533964013
Exakte Signifikanz [2*(1-seitig Sig.)]	0,606317831	0,193186586	0,82857064	0,217	0,633438457	0,572603867	0,572603867

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	11	0,26	0,109	0,032899654	
	GP	11	0,28	0,124	0,037280054	
Score WT post	FU	11	0,61	0,149	0,044948425	
	GP	11	0,68	0,126	0,037926467	
Score WT follow up	FU	8	0,38	0,231	0,081690009	
	GP	10	0,37	0,126	0,039751245	
Residuen prae post	FU	11	-0,220	1,060	0,319718613	
	GP	11	0,220	0,877	0,264403231	
Residuen prae follow up	FU	8	0,059	1,340	0,473843945	
	GP	10	-0,047	0,613	0,193708625	
Residuen post follow up	FU	8	0,175	1,127	0,398358505	
	GP	10	-0,140	0,861	0,272258401	

Klasse 10 (Ti 10c)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	79	62	71	78.000	80	82	82
Wilcoxon-W	184	167	176	183.000	185	187	187
Z	-0,883836014	-1,412224775	-0,975599009	-0,631	-0,533869233	-0,436735427	-0,436735427
Asymptotische Signifikanz (2-seitig)	0,376784742	0,157883788	0,329263216	0,528	0,59343201	0,662303233	0,662303233
Exakte Signifikanz [2*(1-seitig Sig.)]	0,401263068	0,169385541	0,349884387	0,550	0,615950604	0,68484747	0,68484747

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	14	0,29	0,065	0,01744116	
	GP	14	0,32	0,104	0,027779287	
Score WT post	FU	14	0,64	0,170	0,045362138	
	GP	13	0,73	0,140	0,038759823	
Score WT follow up	FU	14	0,52	0,153	0,040923892	
	GP	13	0,59	0,100	0,027832039	
Residuen prae post	FU	14	-0,206	1,169	0,312538028	
	GP	13	0,222	0,706	0,19592224	
Residuen prae follow up	FU	14	-0,199	1,261	0,337008839	
	GP	13	0,215	0,514	0,142603197	
Residuen post follow up	FU	14	-0,099	1,250	0,333994697	
	GP	13	0,107	0,606	0,168155435	

Klasse 11 (Ti 10d)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	104,5	31,5	67,5	30.500	62	91	91
Wilcoxon-W	209,5	136,5	172,5	135.500	167	211	211
Z	-0,021984843	-3,23096432	-1,642714572	-3,252	-1,876673856	-0,611085354	-0,611085354
Asymptotische Signifikanz (2-seitig)	0,982460046	0,001233733	0,100442002	0,001	0,060562808	0,541143075	0,541143075
Exakte Signifikanz [2*(1-seitig Sig.)]	0,982864863	0,000764607	0,102308598	0,001	0,062953353	0,561290791	0,561290791

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	14	0,32	0,083	0,022118382	
	GP	15	0,31	0,080	0,020658145	
Score WT post	FU	14	0,66	0,120	0,03214724	
	GP	15	0,83	0,086	0,022270229	
Score WT follow up	FU	14	0,58	0,175	0,046744378	
	GP	15	0,70	0,158	0,04086181	
Residuen prae post	FU	14	-0,642	0,880	0,235091674	
	GP	15	0,599	0,643	0,165939157	
Residuen prae follow up	FU	14	-0,351	0,997	0,266467647	
	GP	15	0,328	0,876	0,226138626	
Residuen post follow up	FU	14	0,132	1,063	0,284116698	
	GP	15	-0,123	0,920	0,237415828	

Klasse 12 (Sc Diff)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	95,5	79,5	115,5	67.000	90	152	152
Wilcoxon-W	266,5	232,5	268,5	220.000	243	305	305
Z	-1,916803586	-2,437668727	-1,241340816	-2,839	-2,079601121	-0,033009542	-0,033009542
Asymptotische Signifikanz (2-seitig)	0,055262889	0,014782314	0,214479873	0,005	0,037562134	0,973666979	0,973666979
Exakte Signifikanz [2*(1-seitig Sig.)]	0,057029686	0,013908802	0,218846091	0,004	0,037992391	0,987005105	0,987005105

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	17	0,33	0,087	0,021136318	
	GP	18	0,27	0,079	0,018571214	
Score WT post	FU	17	0,61	0,124	0,030109523	
	GP	18	0,71	0,178	0,042037703	
Score WT follow up	FU	17	0,53	0,140	0,033930561	
	GP	18	0,59	0,165	0,038871762	
Residuen prae post	FU	17	-0,386	0,760	0,184301237	
	GP	18	0,365	1,052	0,248065782	
Residuen prae follow up	FU	17	-0,312	0,887	0,215177338	
	GP	18	0,295	1,006	0,237012981	
Residuen post follow up	FU	17	-0,010	0,919	0,222913963	
	GP	18	0,010	1,070	0,252314424	

t-Test & U-Test Klassenvergleich Fortsetzung

Klasse 13 (Kü 10a)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	79	19	48,5	20.000	45,5	92	
Wilcoxon-W	199	124	153,5	125.000	150,5	197	
Z	-1,151299351	-3,762626707	-2,474716271	-3,711	-2,598713836	-0,5674364	
Asymptotische Signifikanz (2-seitig)	0,249609104	0,000168138	0,013334206	0,000	0,009357374	0,570417725	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,270298777	5,27858E-05	0,012021182	0,000	0,007903066	0,590744927	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	14	0,25	0,073	0,019403816	
	GP	15	0,23	0,060	0,015510216	
Score WT post	FU	14	0,45	0,084	0,022574166	
	GP	15	0,68	0,150	0,038708201	
Score WT follow up	FU	14	0,41	0,127	0,033915098	
	GP	15	0,57	0,165	0,042618475	
Residuen prae post	FU	14	-0,686	0,493	0,131751329	
	GP	15	0,640	0,891	0,230000769	
Residuen prae follow up	FU	14	-0,494	0,731	0,195270387	
	GP	15	0,461	0,980	0,25303573	
Residuen post follow up	FU	14	-0,053	0,792	0,211570115	
	GP	15	0,050	1,158	0,299021699	

Klasse 14 (Li 10b2)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	97,5	84,5	76	93.000	71,5	56,5	
Wilcoxon-W	217,5	204,5	196	213.000	191,5	176,5	
Z	-0,628209609	-1,164629093	-1,51987825	-0,809	-1,701736912	-2,323288714	
Asymptotische Signifikanz (2-seitig)	0,529866637	0,244169162	0,128541578	0,419	0,088804698	0,020163648	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,53929854	0,249556195	0,136974224	0,436	0,089194773	0,018553726	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	15	0,28	0,063	0,016186017	
	GP	15	0,26	0,094	0,024332693	
Score WT post	FU	15	0,70	0,120	0,031108836	
	GP	15	0,64	0,171	0,044168971	
Score WT follow up	FU	15	0,51	0,164	0,042398824	
	GP	15	0,60	0,200	0,051574858	
Residuen prae post	FU	15	0,170	0,918	0,236939476	
	GP	15	-0,170	1,047	0,270335404	
Residuen prae follow up	FU	15	-0,314	0,846	0,218456012	
	GP	15	0,314	1,036	0,267479494	
Residuen post follow up	FU	15	-0,419	0,801	0,206735246	
	GP	15	0,419	0,991	0,255894503	

Klasse 15 (Wi 10a)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	106	39	60	39.000	54,5	103	
Wilcoxon-W	226	159	180	159.000	174,5	223	
Z	-0,272812623	-3,056809464	-2,183435331	-3,049	-2,406531135	-0,394041644	
Asymptotische Signifikanz (2-seitig)	0,784997256	0,002237064	0,029003766	0,002	0,016104831	0,693550294	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,806333778	0,001665215	0,029496475	0,002	0,014519172	0,712965125	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	15	0,31	0,074	0,019211684	
	GP	15	0,32	0,087	0,022393202	
Score WT post	FU	15	0,67	0,123	0,031764014	
	GP	15	0,82	0,099	0,025490196	
Score WT follow up	FU	15	0,66	0,124	0,031945056	
	GP	15	0,78	0,147	0,037934245	
Residuen prae post	FU	15	-0,566	0,901	0,232615236	
	GP	15	0,566	0,709	0,183082378	
Residuen prae follow up	FU	15	-0,443	0,969	0,250277155	
	GP	15	0,443	0,800	0,206644585	
Residuen post follow up	FU	15	-0,107	0,961	0,24816601	
	GP	15	0,107	1,026	0,264809964	

Klasse 16 (Kn 10c)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	71	14	18	18.000	21	69	
Wilcoxon-W	162	105	109	109.000	112	160	
Z	-0,984444702	-3,750853486	-3,553818334	-3,546	-3,397868461	-1,067738466	
Asymptotische Signifikanz (2-seitig)	0,324896908	0,000176234	0,000379681	0,000	0,000679131	0,285638502	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,349884387	5,05526E-05	0,000155247	0,000	0,000330138	0,302163095	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	13	0,25	0,058	0,016157078	
	GP	14	0,28	0,077	0,020616908	
Score WT post	FU	13	0,62	0,113	0,031240294	
	GP	14	0,81	0,089	0,023803978	
Score WT follow up	FU	13	0,42	0,127	0,035146848	
	GP	14	0,65	0,129	0,034451417	
Residuen prae post	FU	13	-0,660	0,845	0,234391568	
	GP	14	0,613	0,650	0,173740392	
Residuen prae follow up	FU	13	-0,634	0,788	0,218508903	
	GP	14	0,589	0,758	0,202560716	
Residuen post follow up	FU	13	-0,231	1,121	0,311040227	
	GP	14	0,215	0,812	0,216949683	

t-Test & U-Test Klassenvergleich Fortsetzung

Klasse 17 (Wh 10d)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	38	26,5	29,5	31.500	30	41,5	
Wilcoxon-W	93	81,5	84,5	86.500	85	96,5	
Z	-0,915061426	-1,787896175	-1,556102711	-1,399	-1,511857892	-0,642781297	
Asymptotische Signifikanz (2-seitig)	0,360159384	0,07379276	0,11968367	0,162	0,130570018	0,520366021	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,393048128	0,075256013	0,123005477	0,165	0,143140142	0,52884886	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	10	0,31	0,106	0,033405343	
	GP	10	0,36	0,112	0,035402883	
Score WT post	FU	10	0,70	0,162	0,051215521	
	GP	10	0,82	0,106	0,033520237	
Score WT follow up	FU	10	0,66	0,159	0,050230151	
	GP	10	0,78	0,141	0,044540458	
Residuen prae post	FU	10	-0,334	1,111	0,351457901	
	GP	10	0,334	0,719	0,227322778	
Residuen prae follow up	FU	10	-0,275	0,942	0,297951255	
	GP	10	0,275	0,972	0,307221837	
Residuen post follow up	FU	10	-0,058	0,799	0,252775773	
	GP	10	0,058	1,163	0,367859358	

Gruppe 18 (Tw parallel)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	91,5	68,5	50,5	78.500	66	61	
Wilcoxon-W	244,5	173,5	141,5	183.500	157	152	
Z	-1,371707035	-1,600100436	-1,981402608	-1,157	-1,21352445	-1,455784758	
Asymptotische Signifikanz (2-seitig)	0,170154658	0,109576304	0,047546141	0,247	0,224929372	0,145452131	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,176456045	0,112069791	0,048172876	0,252	0,238822532	0,154582293	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	17	0,19	0,064	0,015558915	
	GP	15	0,23	0,063	0,016152048	
Score WT post	FU	14	0,33	0,107	0,028497185	
	GP	15	0,41	0,145	0,037365309	
Score WT follow up	FU	13	0,25	0,075	0,020694454	
	GP	14	0,31	0,094	0,025162906	
Residuen prae post	FU	14	-0,235	0,908	0,242773553	
	GP	15	0,219	1,028	0,265311334	
Residuen prae follow up	FU	13	-0,249	0,875	0,242687952	
	GP	14	0,231	1,047	0,279912411	
Residuen post follow up	FU	13	-0,280	0,835	0,231462701	
	GP	14	0,260	1,063	0,284021773	

Klasse 19 (Tw mp)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	97,5	79	43,5	75.000	43	46	
Wilcoxon-W	202,5	184	134,5	180.000	134	137	
Z	-0,023410367	-0,878310066	-2,113430139	-1,057	-2,130026442	-1,974696558	
Asymptotische Signifikanz (2-seitig)	0,981322935	0,379775475	0,034563969	0,290	0,03316943	0,048302594	
Exakte Signifikanz [2*(1-seitig Sig.)]	0,981968711	0,401263068	0,033827664	0,306	0,033827664	0,050140184	

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	14	0,23	0,084	0,022483749	
	GP	14	0,23	0,062	0,016654531	
Score WT post	FU	14	0,51	0,130	0,034770214	
	GP	14	0,55	0,127	0,03383492	
Score WT follow up	FU	13	0,33	0,120	0,033379029	
	GP	13	0,45	0,140	0,038858742	
Residuen prae post	FU	14	-0,183	1,023	0,273495908	
	GP	14	0,183	0,938	0,250770608	
Residuen prae follow up	FU	13	-0,421	0,827	0,229317221	
	GP	13	0,421	0,966	0,267917805	
Residuen post follow up	FU	13	-0,375	0,912	0,252903991	
	GP	13	0,375	0,929	0,257716833	

Gruppe 20 (Dr parallel)

U-Test							
	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up	
Mann-Whitney-U	208,5	164	196,5	158.500	204	149	
Wilcoxon-W	398,5	440	472,5	434.500	480	320	
Z	-0,258182623	-1,382410545	-0,277438282	-1,517	-0,078839515	-1,524543563	
Asymptotische Signifikanz (2-seitig)	0,796265969	0,16684568	0,781443959	0,129	0,937161069	0,127372983	

a. Gruppenvariable: Gruppe

T-Test						
Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	23	0,27	0,109	0,022758099	
	GP	19	0,24	0,060	0,013651968	
Score WT post	FU	23	0,50	0,145	0,030288285	
	GP	19	0,54	0,127	0,029170886	
Score WT follow up	FU	23	0,49	0,140	0,029170691	
	GP	18	0,47	0,106	0,024925981	
Residuen prae post	FU	23	-0,186	1,012	0,210976648	
	GP	19	0,225	0,935	0,214439709	
Residuen prae follow up	FU	23	0,000	1,059	0,220851846	
	GP	18	0,001	0,918	0,216323748	
Residuen post follow up	FU	23	0,171	0,942	0,1964566	
	GP	18	-0,219	1,027	0,242083602	

t-Test & U-Test Klassenvergleich Fortsetzung

Klasse 21 (Dr mp)

U-Test

	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	39	33,5	58	56,500	65	60
Wilcoxon-W	175	169,5	194	192,500	201	105
Z	-2,177328275	-2,187641252	-0,794890193	-0,878	-0,396296962	-0,67938622
Asymptotische Signifikanz (2-seitig)	0,029456083	0,028695744	0,426677379	0,380	0,691885968	0,496905848
Exakte Signifikanz [2*(1-seitig Sig.)]	0,030914946	0,02715158	0,452302157	0,388	0,718231011	0,522456711

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	10	0,28	0,071	0,022442202	
	GP	16	0,21	0,081	0,020353839	
Score WT post	FU	9	0,60	0,106	0,035386287	
	GP	16	0,47	0,143	0,035767732	
Score WT follow up	FU	9	0,55	0,083	0,027633226	
	GP	16	0,50	0,157	0,039357129	
Residuen prae post	FU	9	0,261	1,031	0,343627175	
	GP	16	-0,147	0,950	0,237516753	
Residuen prae follow up	FU	9	0,138	0,671	0,223792413	
	GP	16	-0,078	1,129	0,282302396	
Residuen post follow up	FU	9	-0,108	0,641	0,213568987	
	GP	16	0,061	1,142	0,285419734	

Klasse 22 (Pa parallel)

U-Test

	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	143,5	29	97,5	34,000	103,5	142
Wilcoxon-W	353,5	239	287,5	244,000	293,5	262
Z	-1,082964029	-4,428356809	-1,572737925	-4,269	-1,353415498	-0,017343524
Asymptotische Signifikanz (2-seitig)	0,27882439	9,49537E-06	0,115779514	0,000	0,175922929	0,986162563
Exakte Signifikanz [2*(1-seitig Sig.)]	0,28960235	1,3351E-06	0,119556101	0,000	0,178537705	1

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	20	0,21	0,066	0,014783079	
	GP	18	0,23	0,055	0,012981473	
Score WT post	FU	20	0,43	0,096	0,021390814	
	GP	18	0,62	0,109	0,025771694	
Score WT follow up	FU	19	0,36	0,109	0,024949881	
	GP	15	0,44	0,134	0,034618494	
Residuen prae post	FU	20	-0,626	0,686	0,153475653	
	GP	18	0,695	0,786	0,18533315	
Residuen prae follow up	FU	19	-0,274	0,852	0,195542212	
	GP	15	0,347	1,059	0,273353494	
Residuen post follow up	FU	19	-0,032	0,945	0,21677918	
	GP	15	0,041	1,065	0,275025207	

Klasse 23 (Pö GO 12)

U-Test

	Score WT prae	Score WT post	Score WT follow up	Residuen prae post	Residuen prae follow up	Residuen post follow up
Mann-Whitney-U	83	47	46	36,500	32	60
Wilcoxon-W	188	125	112	114,500	98	126
Z	-0,394149428	-1,912909726	-1,709631038	-2,445	-2,465892248	-0,930842158
Asymptotische Signifikanz (2-seitig)	0,69347072	0,05575961	0,087334122	0,014	0,013667244	0,351935219
Exakte Signifikanz [2*(1-seitig Sig.)]	0,720314184	0,059595556	0,09542693	0,013	0,012838875	0,373028671

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Gruppe

T-Test

Gruppenstatistiken						
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Score WT prae	FU	13	0,27	0,087	0,024209202	
	GP	14	0,27	0,070	0,018618052	
Score WT post	FU	12	0,59	0,125	0,035991574	
	GP	14	0,69	0,111	0,029687506	
Score WT follow up	FU	11	0,39	0,139	0,041868618	
	GP	14	0,48	0,097	0,026057717	
Residuen prae post	FU	12	-0,492	0,924	0,266839095	
	GP	14	0,421	0,842	0,22503219	
Residuen prae follow up	FU	11	-0,513	0,977	0,294651223	
	GP	14	0,403	0,798	0,213338424	
Residuen post follow up	FU	11	-0,220	1,277	0,384918771	
	GP	14	0,173	0,665	0,177719956	

12.3.6 Art der Wissensvermittlung

U-Test selbst - fremd vermittelt Thema 1

VU

Ränge

AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Selbst	12	50,08	601,00
	Fremd	56	31,16	1.745,00
	Gesamt	68		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	149,000
Wilcoxon-W	1.745,000
Z	-3,012
Asymptotische Signifikanz (2-seitig)	0,003

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik

AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 1	12	0,7082622
	Gültige Werte (Listenweise)	12	
Fremd	Residuen Thema 1	56	-0,1517705
	Gültige Werte (Listenweise)	56	

HU I

Ränge

AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Selbst	27	118,33	3.195,00
	Fremd	130	70,83	9.208,00
	Gesamt	157		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	693,000
Wilcoxon-W	9.208,000
Z	-4,945
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik(a)

AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 1	27	0,7960873
	Gültige Werte (Listenweise)	27	
Fremd	Residuen Thema 1	130	-0,1530534
	Gültige Werte (Listenweise)	130	

gültigen Fälle vorliegen.

HU II intern

Ränge

AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Selbst	30	121,88	3.656,50
	Fremd	150	84,22	12.633,50
	Gesamt	180		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	1.308,500
Wilcoxon-W	12.633,500
Z	-3,617
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik

AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 1	30	0,6059676
	Gültige Werte (Listenweise)	30	
Fremd	Residuen Thema 1	150	-0,1211935
	Gültige Werte (Listenweise)	150	

HU II extern

Ränge

AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Selbst	41	161,77	6.632,50
	Fremd	194	108,75	21.097,50
	Gesamt	235		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	2.182,500
Wilcoxon-W	21.097,500
Z	-4,540
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik(a)

AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 1	41	0,6299148
	Gültige Werte (Listenweise)	41	
Fremd	Residuen Thema 1	194	-0,1331263
	Gültige Werte (Listenweise)	194	

gültigen Fälle vorliegen.

U-Test Mitschüler-Lehrer Thema 1

VU

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Mitschüler	23	31,70	729,00
	Lehrer	33	26,27	867,00
	Gesamt	56		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	306,000
Wilcoxon-W	867,000
Z	-1,226
Asymptotische Signifikanz (2-seitig)	0,220

a. Gruppenvariable: Art der Wissensvermittlung

HU I

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Mitschüler	53	64,26	3.406,00
	Lehrer	77	66,35	5.109,00
	Gesamt	130		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	1.975,000
Wilcoxon-W	3.406,000
Z	-0,311
Asymptotische Signifikanz (2-seitig)	0,756

a. Gruppenvariable: Art der Wissensvermittlung

HU II intern

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Mitschüler	61	85,65	5.224,50
	Lehrer	89	68,54	6.100,50
	Gesamt	150		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	2.095,500
Wilcoxon-W	6.100,500
Z	-2,370
Asymptotische Signifikanz (2-seitig)	0,018

a. Gruppenvariable: Art der Wissensvermittlung

HU II extern

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 1	Mitschüler	79	107,41	8.485,50
	Lehrer	115	90,69	10.429,50
	Gesamt	194		

Statistik für Test(a)

Residuen Thema 1	
Mann-Whitney-U	3.759,500
Wilcoxon-W	10.429,500
Z	-2,039
Asymptotische Signifikanz (2-seitig)	0,041

a. Gruppenvariable: Art der Wissensvermittlung

Ergebnisse Jonckheere-Terpstra Thema 1

VU

Jonckheere-Terpstra-Test(a)

	Residuen Thema 1
Anzahl der Stufen in Art der Wissensvermittlung	3
N	68
Beobachtete J-T-Statistik	455,000
Mittelwert der J-T-Statistik	715,500
Standardabweichung der J-T-Statistik	86,302
Standardisierte J-T-Statistik	-3,018
Asymptotische Signifikanz (2-seitig)	0,003

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 1	12	0,7082622
	Gültige Werte (Listenweise)	12	
Mitschüler	Residuen Thema 1	23	0,1124500
	Gültige Werte (Listenweise)	23	
Lehrer	Residuen Thema 1	33	-0,3359241
	Gültige Werte (Listenweise)	33	

HU I

Jonckheere-Terpstra-Test(a)

	Residuen Thema 1
Anzahl der Stufen in Art der Wissensvermittlung	3
N	157
Beobachtete J-T-Statistik	2.799,000
Mittelwert der J-T-Statistik	3.795,500
Standardabweichung der J-T-Statistik	300,939
Standardisierte J-T-Statistik	-3,311
Asymptotische Signifikanz (2-seitig)	0,001

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik(a)

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 1	27	0,7960873
	Gültige Werte (Listenweise)	27	
Mitschüler	Residuen Thema 1	53	-0,2247558
	Gültige Werte (Listenweise)	53	
Lehrer	Residuen Thema 1	77	-0,1036998
	Gültige Werte (Listenweise)	77	

Fälle vorliegen.

HU II intern

Jonckheere-Terpstra-Test(a)

	Residuen Thema 1
Anzahl der Stufen in Art der Wissensvermittlung	3
N	180
Beobachtete J-T-Statistik	3.404,000
Mittelwert der J-T-Statistik	4.964,500
Standardabweichung der J-T-Statistik	368,697
Standardisierte J-T-Statistik	-4,232
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 1	30	0,6059676
	Gültige Werte (Listenweise)	30	
Mitschüler	Residuen Thema 1	61	0,0945596
	Gültige Werte (Listenweise)	61	
Lehrer	Residuen Thema 1	89	-0,2690693
	Gültige Werte (Listenweise)	89	

HU II extern

Jonckheere-Terpstra-Test(a)

	Residuen Thema 1
Anzahl der Stufen in Art der Wissensvermittlung	3
N	235
Beobachtete J-T-Statistik	5.942,000
Mittelwert der J-T-Statistik	8.519,500
Standardabweichung der J-T-Statistik	551,054
Standardisierte J-T-Statistik	-4,677
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik(a)

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 1	41	0,6299148
	Gültige Werte (Listenweise)	41	
Mitschüler	Residuen Thema 1	79	0,0365750
	Gültige Werte (Listenweise)	79	
Lehrer	Residuen Thema 1	115	-0,2497037
	Gültige Werte (Listenweise)	115	

Fälle vorliegen.

U-Test selbst-fremd vermittelt Thema 2

VU

		Ränge		
AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 2	Selbst	11	50,86	559,50
	Fremd	57	31,34	1.786,50
	Gesamt	68		

Statistik für Test(a)	
Residuen Thema 2	
Mann-Whitney-U	133,500
Wilcoxon-W	1.786,500
Z	-3,011
Asymptotische Signifikanz (2-seitig)	0,003

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 2	11	0,7830551
	Gültige Werte (Listenweise)	11	
Fremd	Residuen Thema 2	57	-0,1511159
	Gültige Werte (Listenweise)	57	

HU I

		Ränge		
AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 2	Selbst	28	100,91	2.825,50
	Fremd	130	74,89	9.735,50
	Gesamt	158		

Statistik für Test(a)	
Residuen Thema 2	
Mann-Whitney-U	1.220,500
Wilcoxon-W	9.735,500
Z	-2,736
Asymptotische Signifikanz (2-seitig)	0,006

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik(a)			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 2	28	0,4600146
	Gültige Werte (Listenweise)	28	
Fremd	Residuen Thema 2	130	-0,0990801
	Gültige Werte (Listenweise)	130	

gültigen Fälle vorliegen.

HU II intern

		Ränge		
AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 2	Selbst	30	120,83	3.625,00
	Fremd	150	84,43	12.665,00
	Gesamt	180		

Statistik für Test(a)	
Residuen Thema 2	
Mann-Whitney-U	1.340,000
Wilcoxon-W	12.665,000
Z	-3,496
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 2	30	0,5425602
	Gültige Werte (Listenweise)	30	
Fremd	Residuen Thema 2	150	-0,1085120
	Gültige Werte (Listenweise)	150	

HU II extern

		Ränge		
AdW 2stufig		N	Mittlerer Rang	Rangsumme
Residuen Thema 2	Selbst	40	171,16	6.846,50
	Fremd	195	107,09	20.883,50
	Gesamt	235		

Statistik für Test(a)	
Residuen Thema 2	
Mann-Whitney-U	1.773,500
Wilcoxon-W	20.883,500
Z	-5,434
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik(a)			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 2	40	0,7577492
	Gültige Werte (Listenweise)	40	
Fremd	Residuen Thema 2	195	-0,1554357
	Gültige Werte (Listenweise)	195	

gültigen Fälle vorliegen.

U-Test Mitschüler Lehrer Thema 2

VU

Ränge

		N	Mittlerer Rang	Rangsumme
Art der Wissensvermittlung				
Residuen Thema 2	Mitschüler	24	27,50	660,00
	Lehrer	33	30,09	993,00
	Gesamt	57		

Statistik für Test(a)

		Residuen Thema 2
Mann-Whitney-U		360,000
Wilcoxon-W		660,000
Z		-0,584
Asymptotische Signifikanz (2-seitig)		0,559

a. Gruppenvariable: Art der Wissensvermittlung

HU I

Ränge

		N	Mittlerer Rang	Rangsumme
Art der Wissensvermittlung				
Residuen Thema 2	Mitschüler	52	62,93	3.272,50
	Lehrer	78	67,21	5.242,50
	Gesamt	130		

Statistik für Test(a)

		Residuen Thema 2
Mann-Whitney-U		1.894,500
Wilcoxon-W		3.272,500
Z		-0,636
Asymptotische Signifikanz (2-seitig)		0,525

a. Gruppenvariable: Art der Wissensvermittlung

HU II intern

Ränge

		N	Mittlerer Rang	Rangsumme
Art der Wissensvermittlung				
Residuen Thema 2	Mitschüler	61	82,19	5.013,50
	Lehrer	89	70,92	6.311,50
	Gesamt	150		

Statistik für Test(a)

		Residuen Thema 2
Mann-Whitney-U		2.306,500
Wilcoxon-W		6.311,500
Z		-1,562
Asymptotische Signifikanz (2-seitig)		0,118

a. Gruppenvariable: Art der Wissensvermittlung

HU II extern

Ränge

		N	Mittlerer Rang	Rangsumme
Art der Wissensvermittlung				
Residuen Thema 2	Mitschüler	80	112,85	9.028,00
	Lehrer	115	87,67	10.082,00
	Gesamt	195		

Statistik für Test(a)

		Residuen Thema 2
Mann-Whitney-U		3.412,000
Wilcoxon-W		10.082,000
Z		-3,067
Asymptotische Signifikanz (2-seitig)		0,002

a. Gruppenvariable: Art der Wissensvermittlung

Ergebnisse Jonckheere-Terpstra Thema 2

VU

Jonckheere-Terpstra-Test(a)

		Residuen Thema 2
Anzahl der Stufen in Art der Wissensvermittlung		3
N		68
Beobachtete J-T-Statistik		565,500
Mittelwert der J-T-Statistik		709,500
Standardabweichung der J-T-Statistik		85,838
Standardisierte J-T-Statistik		-1,678
Asymptotische Signifikanz (2-seitig)		0,093

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 2	11	0,7830551
	Gültige Werte (Listenweise)	11	
Mitschüler	Residuen Thema 2	24	-0,1896482
	Gültige Werte (Listenweise)	24	
Lehrer	Residuen Thema 2	33	-0,1230924
	Gültige Werte (Listenweise)	33	

HU I

Jonckheere-Terpstra-Test(a)

		Residuen Thema 2
Anzahl der Stufen in Art der Wissensvermittlung		3
N		158
Beobachtete J-T-Statistik		3.382,000
Mittelwert der J-T-Statistik		3.848,000
Standardabweichung der J-T-Statistik		303,500
Standardisierte J-T-Statistik		-1,535
Asymptotische Signifikanz (2-seitig)		0,125

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik(a)

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 2	28	0,4600146
	Gültige Werte (Listenweise)	28	
Mitschüler	Residuen Thema 2	52	-0,2172693
	Gültige Werte (Listenweise)	52	
Lehrer	Residuen Thema 2	78	-0,0202872
	Gültige Werte (Listenweise)	78	

Fälle vorliegen.

HU II intern

Jonckheere-Terpstra-Test(a)

		Residuen Thema 2
Anzahl der Stufen in Art der Wissensvermittlung		3
N		180
Beobachtete J-T-Statistik		3.646,500
Mittelwert der J-T-Statistik		4.964,500
Standardabweichung der J-T-Statistik		368,746
Standardisierte J-T-Statistik		-3,574
Asymptotische Signifikanz (2-seitig)		0,000

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 2	30	0,5425602
	Gültige Werte (Listenweise)	30	
Mitschüler	Residuen Thema 2	61	0,0510050
	Gültige Werte (Listenweise)	61	
Lehrer	Residuen Thema 2	89	-0,2178440
	Gültige Werte (Listenweise)	89	

HU II extern

Jonckheere-Terpstra-Test(a)

		Residuen Thema 2
Anzahl der Stufen in Art der Wissensvermittlung		3
N		235
Beobachtete J-T-Statistik		5.185,500
Mittelwert der J-T-Statistik		8.500,000
Standardabweichung der J-T-Statistik		550,613
Standardisierte J-T-Statistik		-6,020
Asymptotische Signifikanz (2-seitig)		0,000

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik(a)

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 2	40	0,7577492
	Gültige Werte (Listenweise)	40	
Mitschüler	Residuen Thema 2	80	0,0830943
	Gültige Werte (Listenweise)	80	
Lehrer	Residuen Thema 2	115	-0,3213697
	Gültige Werte (Listenweise)	115	

Fälle vorliegen.

U-Test selbst-fremd vermittelt Thema 3

VU

Ränge		AdW 2stufig	N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Selbst		12	42,54	510,50
	Fremd		56	32,78	1.835,50
	Gesamt		68		

Statistik für Test(a)		Residuen Thema 3
Mann-Whitney-U		239,500
Wilcoxon-W		1.835,500
Z		-1,555
Asymptotische Signifikanz (2-seitig)		0,120

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 3	12	0,3936018
	Gültige Werte (Listenweise)	12	
Fremd	Residuen Thema 3	56	-0,0843433
	Gültige Werte (Listenweise)	56	

HU I

Ränge		AdW 2stufig	N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Selbst		25	126,88	3.172,00
	Fremd		133	70,59	9.389,00
	Gesamt		158		

Statistik für Test(a)		Residuen Thema 3
Mann-Whitney-U		478,000
Wilcoxon-W		9.389,000
Z		-5,647
Asymptotische Signifikanz (2-seitig)		0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik(a)			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 3	25	1,0360210
	Gültige Werte (Listenweise)	25	
Fremd	Residuen Thema 3	133	-0,1947408
	Gültige Werte (Listenweise)	133	

a. Es werden für eine oder mehrere aufgeteilte Dateien keine Statistiken berechnet, da keine gültigen Fälle vorliegen.

HU II intern

Ränge		AdW 2stufig	N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Selbst		31	131,02	4.061,50
	Fremd		149	82,07	12.228,50
	Gesamt		180		

Statistik für Test(a)		Residuen Thema 3
Mann-Whitney-U		1.053,500
Wilcoxon-W		12.228,500
Z		-4,761
Asymptotische Signifikanz (2-seitig)		0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 3	31	0,7475438
	Gültige Werte (Listenweise)	31	
Fremd	Residuen Thema 3	149	-0,1555292
	Gültige Werte (Listenweise)	149	

HU II extern

Ränge		AdW 2stufig	N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Selbst		39	156,15	6.090,00
	Fremd		196	110,41	21.640,00
	Gesamt		235		

Statistik für Test(a)		Residuen Thema 3
Mann-Whitney-U		2.334,000
Wilcoxon-W		21.640,000
Z		-3,840
Asymptotische Signifikanz (2-seitig)		0,000

a. Gruppenvariable: AdW 2stufig

Deskriptive Statistik(a)			
AdW 2stufig		N	Mittelwert
Selbst	Residuen Thema 3	39	0,5266486
	Gültige Werte (Listenweise)	39	
Fremd	Residuen Thema 3	196	-0,1047923
	Gültige Werte (Listenweise)	196	

a. Es werden für eine oder mehrere aufgeteilte Dateien keine Statistiken berechnet, da keine gültigen Fälle vorliegen.

U-Test Mitschüler Lehrer Thema 3

VU

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Mitschüler	23	31,35	721,00
	Lehrer	33	26,52	875,00
	Gesamt	56		

Statistik für Test(a)

	Residuen Thema 3
Mann-Whitney-U	314,000
Wilcoxon-W	875,000
Z	-1,093
Asymptotische Signifikanz (2-seitig)	0,275

a. Gruppenvariable: Art der Wissensvermittlung

HU I

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Mitschüler	55	67,35	3.704,50
	Lehrer	78	66,75	5.206,50
	Gesamt	133		

Statistik für Test(a)

	Residuen Thema 3
Mann-Whitney-U	2.125,500
Wilcoxon-W	5.206,500
Z	-0,089
Asymptotische Signifikanz (2-seitig)	0,929

a. Gruppenvariable: Art der Wissensvermittlung

HU II intern

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Mitschüler	60	94,93	5.695,50
	Lehrer	89	61,57	5.479,50
	Gesamt	149		

Statistik für Test(a)

	Residuen Thema 3
Mann-Whitney-U	1.474,500
Wilcoxon-W	5.479,500
Z	-4,630
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: Art der Wissensvermittlung

HU II extern

Ränge

Art der Wissensvermittlung		N	Mittlerer Rang	Rangsumme
Residuen Thema 3	Mitschüler	81	94,86	7.683,50
	Lehrer	115	101,07	11.622,50
	Gesamt	196		

Statistik für Test(a)

	Residuen Thema 3
Mann-Whitney-U	4.362,500
Wilcoxon-W	7.683,500
Z	-0,755
Asymptotische Signifikanz (2-seitig)	0,450

a. Gruppenvariable: Art der Wissensvermittlung

Ergebnisse Jonckheere-Terpstra Thema 3

VU

Jonckheere-Terpstra-Test(a)

	Residuen Thema 3
Anzahl der Stufen in Art der Wissensvermittlung	3
N	68
Beobachtete J-T-Statistik	553,500
Mittelwert der J-T-Statistik	715,500
Standardabweichung der J-T-Statistik	86,296
Standardisierte J-T-Statistik	-1,877
Asymptotische Signifikanz (2-seitig)	0,060

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 3	12	0,3936018
	Gültige Werte (Listenweise)	12	
Mitschüler	Residuen Thema 3	23	0,0579406
	Gültige Werte (Listenweise)	23	
Lehrer	Residuen Thema 3	33	-0,1835108
	Gültige Werte (Listenweise)	33	

HU I

Jonckheere-Terpstra-Test(a)

	Residuen Thema 3
Anzahl der Stufen in Art der Wissensvermittlung	3
N	158
Beobachtete J-T-Statistik	2.603,500
Mittelwert der J-T-Statistik	3.807,500
Standardabweichung der J-T-Statistik	303,044
Standardisierte J-T-Statistik	-3,973
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik(a)

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 3	25	1,0360210
	Gültige Werte (Listenweise)	25	
Mitschüler	Residuen Thema 3	55	-0,1625059
	Gültige Werte (Listenweise)	55	
Lehrer	Residuen Thema 3	78	-0,2174705
	Gültige Werte (Listenweise)	78	

Fälle vorliegen.

HU II intern

Jonckheere-Terpstra-Test(a)

	Residuen Thema 3
Anzahl der Stufen in Art der Wissensvermittlung	3
N	180
Beobachtete J-T-Statistik	2.528,000
Mittelwert der J-T-Statistik	4.979,500
Standardabweichung der J-T-Statistik	369,153
Standardisierte J-T-Statistik	-6,641
Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 3	31	0,7475438
	Gültige Werte (Listenweise)	31	
Mitschüler	Residuen Thema 3	60	0,3060366
	Gültige Werte (Listenweise)	60	
Lehrer	Residuen Thema 3	89	-0,4666972
	Gültige Werte (Listenweise)	89	

HU II extern

Jonckheere-Terpstra-Test(a)

	Residuen Thema 3
Anzahl der Stufen in Art der Wissensvermittlung	3
N	235
Beobachtete J-T-Statistik	7.286,500
Mittelwert der J-T-Statistik	8.479,500
Standardabweichung der J-T-Statistik	550,407
Standardisierte J-T-Statistik	-2,167
Asymptotische Signifikanz (2-seitig)	0,030

a. Gruppenvariable: Art der Wissensvermittlung

Deskriptive Statistik(a)

Art der Wissensvermittlung		N	Mittelwert
Selbst	Residuen Thema 3	39	0,5266486
	Gültige Werte (Listenweise)	39	
Mitschüler	Residuen Thema 3	81	-0,1658002
	Gültige Werte (Listenweise)	81	
Lehrer	Residuen Thema 3	115	-0,0618216
	Gültige Werte (Listenweise)	115	

Fälle vorliegen.

12.3.7 Wechselwirkung Unterrichtsmethode – Ausgangsleistung

Univariate Varianzanalyse Chemienote				
Levene-Test auf Gleichheit der Fehlervarianzen(a)				
Abhängige Variable: Residuen prae post				
Untersuchung	F	df1	df2	Signifikanz
Voruntersuchung	1,392	9	58	0,213
Hauptuntersuchung I	2,190	9	148	0,026
Hauptuntersuchung II intern	1,307	9	170	0,237
Hauptuntersuchung II extern	1,006	9	223	0,436

Prüft die Nullhypothese, daß die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.
a. Design: Intercept+gruppe+chemnote+gruppe * chemnote

Tests der Zwischensubjekteffekte							
Abhängige Variable: Residuen prae post							
Untersuchung	Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
Voruntersuchung	gruppe	6,284	1	6,284	7,059	0,010	0,109
	chemnote	2,679	4	0,670	0,752	0,560	0,049
	gruppe * chemnote	5,403	4	1,351	1,517	0,209	0,095
Hauptuntersuchung I	gruppe	8,410	1	8,410	9,160	0,003	0,058
	chemnote	9,952	4	2,488	2,710	0,032	0,068
	gruppe * chemnote	1,754	4	0,439	0,478	0,752	0,013
Hauptuntersuchung II intern	gruppe	27,455	1	27,455	35,762	0,000	0,174
	chemnote	16,905	4	4,226	5,505	0,000	0,115
	gruppe * chemnote	3,130	4	0,782	1,019	0,399	0,023
Hauptuntersuchung II extern	gruppe	15,967	1	15,967	17,769	0,000	0,074
	chemnote	10,250	4	2,563	2,852	0,025	0,049
	gruppe * chemnote	1,600	4	0,400	0,445	0,776	0,008

a. R-Quadrat = ,218 (korrigiertes R-Quadrat = ,096)
b. R-Quadrat = ,129 (korrigiertes R-Quadrat = ,076)
c. R-Quadrat = ,267 (korrigiertes R-Quadrat = ,228)
d. R-Quadrat = ,122 (korrigiertes R-Quadrat = ,086)

Nichtparametrische Korrelationen Residuen - Chemienote

Korrelationen			Residuen prae post	Chemienote
Voruntersuchung	Spearman-Rho	Korrelationskoeffizient	1,000	-0,151
		Residuen prae post		0,219
		N	68	68
	Spearman-Rho	Korrelationskoeffizient	-0,151	1,000
		Chemienote	0,219	
		N	68	72
Hauptuntersuchung I	Spearman-Rho	Korrelationskoeffizient	1,000	-,222(**)
		Residuen prae post		0,005
		N	158	158
	Spearman-Rho	Korrelationskoeffizient	-,222(**)	1,000
		Chemienote	0,005	
		N	158	159
Hauptuntersuchung II intern	Spearman-Rho	Korrelationskoeffizient	1,000	-,246(**)
		Residuen prae post		0,001
		N	180	180
	Spearman-Rho	Korrelationskoeffizient	-,246(**)	1,000
		Chemienote	0,001	
		N	180	181
Hauptuntersuchung II extern	Spearman-Rho	Korrelationskoeffizient	1,000	-,168(*)
		Residuen prae post		0,010
		N	235	233
	Spearman-Rho	Korrelationskoeffizient	-,168(*)	1,000
		Chemienote	0,010	
		N	233	233

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

* Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

12.3.8 Abhängigkeit des Lernerfolgs vom Geschlecht

Univariate Varianzanalyse Geschlecht				
Levene-Test auf Gleichheit der Fehlervarianzen(a)				
Abhängige Variable: Residuen prae post				
Untersuchung	F	df1	df2	Signifikanz
Voruntersuchung	2,218	3	64	0,095
Hauptuntersuchung I	0,277	3	154	0,842
Hauptuntersuchung II intern	0,684	3	176	0,563
Hauptuntersuchung II extern	0,310	3	231	0,818

ist.
a. Design: Intercept+gruppe+sex+gruppe * sex

Tests der Zwischensubjekteffekte							
Abhängige Variable: Residuen prae post							
Untersuchung	Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
Voruntersuchung	gruppe	6,284	1	6,284	6,740	0,012	0,095
	sex	0,003	1	0,003	0,003	0,958	0,000
	gruppe * sex	0,041	1	0,041	0,044	0,834	0,001
Hauptuntersuchung I	gruppe	8,410	1	8,410	8,806	0,003	0,054
	sex	0,465	1	0,465	0,487	0,486	0,003
	gruppe * sex	0,036	1	0,036	0,038	0,847	0,000
Hauptuntersuchung II intern	gruppe	27,455	1	27,455	32,438	0,000	0,156
	sex	0,199	1	0,199	0,235	0,629	0,001
	gruppe * sex	1,380	1	1,380	1,631	0,203	0,009
Hauptuntersuchung II extern	gruppe	17,247	1	17,247	19,022	0,000	0,076
	sex	6,296	1	6,296	6,944	0,009	0,029
	gruppe * sex	0,008	1	0,008	0,008	0,927	0,000

a. R-Quadrat = ,096 (korrigiertes R-Quadrat = ,053)
b. R-Quadrat = ,057 (korrigiertes R-Quadrat = ,039)
c. R-Quadrat = ,163 (korrigiertes R-Quadrat = ,149)
d. R-Quadrat = ,101 (korrigiertes R-Quadrat = ,089)

12.3.9 Abhängigkeit des Lernerfolgs von der Schulform (nur HU IIextern)

Univariate Varianzanalyse Schulform und Geschlecht				
Levene-Test auf Gleichheit der Fehlervarianzen(a)				
Abhängige Variable: Residuen prae post				
F	df1	df2	Signifikanz	
1,367	15	219	0,166	

Gruppen hinweg gleich ist.
* Schulform+gruppe * sex * Schulform

Tests der Zwischensubjekteffekte							
Abhängige Variable: Residuen prae post							
Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat	
gruppe	17,247	1	17,247	24,501	0,000	0,101	
sex	6,296	1	6,296	8,944	0,003	0,039	
Schulform	48,052	3	16,017	22,754	0,000	0,238	
gruppe * sex	0,050	1	0,050	0,071	0,790	0,000	
gruppe * Schulform	5,592	3	1,864	2,648	0,050	0,035	
sex * Schulform	1,017	3	0,339	0,482	0,695	0,007	
gruppe * sex * Schulform	0,582	3	0,194	0,276	0,843	0,004	

a. R-Quadrat = ,338 (korrigiertes R-Quadrat = ,293)

12.3.10 Einfluss der Übungsdauer auf den Lernerfolg

Univariate Varianzanalyse HU II intern ohne Hagen

Levene-Test auf Gleichheit der Fehlervarianzen(a)

Abhängige Variable: Residuen prae post

F	df1	df2	Signifikanz
1,358	26	95	0,145

Gruppen hinweg gleich ist.

a. Design: Intercept+gruppe+übedauer+gruppe * übedauer

Tests der Zwischensubjekteffekte

Abhängige Variable: Residuen prae post

Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
gruppe	17,485	1	17,485	18,190	0,000	0,161
übedauer	7,642	16	0,478	0,497	0,943	0,077
gruppe * übedauer	4,670	9	0,519	0,540	0,842	0,049

a. R-Quadrat = ,246 (korrigiertes R-Quadrat = ,040)

Univariate Varianzanalyse HU II extern Gymnasien

Levene-Test auf Gleichheit der Fehlervarianzen(a)

Abhängige Variable: Residuen prae post

F	df1	df2	Signifikanz
1,123	15	31	0,378

Gruppen hinweg gleich ist.

a. Design: Intercept+gruppe+übedauer+gruppe * übedauer

Tests der Zwischensubjekteffekte

Abhängige Variable: Residuen prae post

Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
gruppe	13,484	1	13,484	17,498	0,000	0,361
übedauer	2,403	9	0,267	0,347	0,951	0,091
gruppe * übedauer	5,224	5	1,045	1,356	0,268	0,179

a. R-Quadrat = ,469 (korrigiertes R-Quadrat = ,212)

Univariate Varianzanalyse HU andere Schulformen

Levene-Test auf Gleichheit der Fehlervarianzen(a)

Abhängige Variable: Residuen prae post

F	df1	df2	Signifikanz
1,447	31	146	0,077

Gruppen hinweg gleich ist.

a. Design: Intercept+gruppe+übedauer+gruppe * übedauer

Tests der Zwischensubjekteffekte

Abhängige Variable: Residuen prae post

Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
gruppe	9,677	1	9,677	11,607	0,001	0,074
übedauer	28,555	19	1,503	1,803	0,027	0,190
gruppe * übedauer	10,688	11	0,972	1,165	0,316	0,081

a. R-Quadrat = ,287 (korrigiertes R-Quadrat = ,135)

Einfluss der Übungsdauer auf den Lernerfolg – Korrelationen

HU II intern ohne Hagen

Korrelationen				Übedauer	Residuen prae post
FU	Spearman-Rho	Übedauer	Korrelationskoeffizient	1,000	0,001
			Sig. (2-seitig)		0,992
			N	60	60
		Residuen prae post	Korrelationskoeffizient	0,001	1,000
			Sig. (2-seitig)	0,992	
			N	60	61
GP	Spearman-Rho	Übedauer	Korrelationskoeffizient	1,000	0,039
			Sig. (2-seitig)		0,764
			N	62	62
		Residuen prae post	Korrelationskoeffizient	0,039	1,000
			Sig. (2-seitig)	0,764	
			N	62	63

Deskriptive Statistik			
Gruppe		N	Mittelwert
FU	Übedauer	60	10,60
	Residuen prae post	61	-0,3771583
	Gültige Werte (Listenweise)	60	
GP	Übedauer	62	10,23
	Residuen prae post	63	0,3651851
	Gültige Werte (Listenweise)	62	

HU II extern Gymnasien

Korrelationen				Übedauer	Residuen prae post
FU	Spearman-Rho	Übedauer	Korrelationskoeffizient	1,000	-0,102
			Sig. (2-seitig)		0,644
			N	23	23
		Residuen prae post	Korrelationskoeffizient	-0,102	1,000
			Sig. (2-seitig)	0,644	
			N	23	23
GP	Spearman-Rho	Übedauer	Korrelationskoeffizient	1,000	0,096
			Sig. (2-seitig)		0,655
			N	24	24
		Residuen prae post	Korrelationskoeffizient	0,096	1,000
			Sig. (2-seitig)	0,655	
			N	24	24

Deskriptive Statistik			
Gruppe		N	Mittelwert
FU	Übedauer	23	11,09
	Residuen prae post	23	-0,5471445
	Gültige Werte (Listenweise)	23	
GP	Übedauer	24	10,79
	Residuen prae post	24	0,5243468
	Gültige Werte (Listenweise)	24	

HU II extern andere Schulformen

Korrelationen				Übedauer	Residuen prae post
FU	Spearman-Rho	Übedauer	Korrelationskoeffizient	1,000	-,219(*)
			Sig. (2-seitig)		0,046
			N	84	84
		Residuen prae post	Korrelationskoeffizient	-,219(*)	1,000
			Sig. (2-seitig)	0,046	
			N	84	92
GP	Spearman-Rho	Übedauer	Korrelationskoeffizient	1,000	-0,174
			Sig. (2-seitig)		0,093
			N	94	94
		Residuen prae post	Korrelationskoeffizient	-0,174	1,000
			Sig. (2-seitig)	0,093	
			N	94	96

*. Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Deskriptive Statistik			
Gruppe		N	Mittelwert
FU	Übedauer	84	20,50
	Residuen prae post	92	-0,2581931
	Gültige Werte (Listenweise)	84	
GP	Übedauer	94	37,52
	Residuen prae post	96	0,2474351
	Gültige Werte (Listenweise)	94	

12.3.11 Nachhaltigkeit des erworbenen Wissens

Wilcoxon Signed Ranks Test

		Ranks		N	Mean Rank	Sum of Ranks
Voruntersuchung	FU	Residuen prae follow up - Residuen prae post	Negative Ranks	16(a)	18,75	300,00
			Positive Ranks	17(b)	15,35	261,00
			Ties	0(c)		
		Total	33			
		Score WT follow up - Score WT post	Negative Ranks	23(d)	16,50	379,50
			Positive Ranks	7(e)	12,21	85,50
	Ties		5(f)			
	GP	Residuen prae follow up - Residuen prae post	Negative Ranks	16(a)	17,31	277,00
			Positive Ranks	18(b)	17,67	318,00
			Ties	0(c)		
		Total	34			
		Score WT follow up - Score WT post	Negative Ranks	24(d)	14,42	346,00
Positive Ranks			3(e)	10,67	32,00	
Ties	8(f)					
Hauptuntersuchung I	FU	Residuen prae follow up - Residuen prae post	Negative Ranks	32(a)	37,84	1.211,00
			Positive Ranks	40(b)	35,43	1.417,00
			Ties	0(c)		
		Total	72			
		Score WT follow up - Score WT post	Negative Ranks	52(d)	35,60	1.851,00
			Positive Ranks	15(e)	28,47	427,00
	Ties		5(f)			
	GP	Residuen prae follow up - Residuen prae post	Negative Ranks	41(a)	42,41	1.739,00
			Positive Ranks	36(b)	35,11	1.264,00
			Ties	0(c)		
		Total	77			
		Score WT follow up - Score WT post	Negative Ranks	59(d)	41,42	2.443,50
Positive Ranks			14(e)	18,39	257,50	
Ties	4(f)					
Hauptuntersuchung II intern	FU	Residuen prae follow up - Residuen prae post	Negative Ranks	43(a)	44,56	1.916,00
			Positive Ranks	46(b)	45,41	2.089,00
			Ties	0(c)		
		Total	89			
		Score WT follow up - Score WT post	Negative Ranks	64(d)	44,77	2.865,50
			Positive Ranks	19(e)	32,66	620,50
	Ties		6(f)			
	GP	Residuen prae follow up - Residuen prae post	Negative Ranks	54(a)	43,80	2.365,00
			Positive Ranks	37(b)	49,22	1.821,00
			Ties	0(c)		
		Total	91			
		Score WT follow up - Score WT post	Negative Ranks	63(d)	46,38	2.922,00
Positive Ranks			20(e)	28,20	564,00	
Ties	8(f)					
Hauptuntersuchung II extern	FU	Residuen prae follow up - Residuen prae post	Negative Ranks	48(a)	59,90	2.875,00
			Positive Ranks	63(b)	53,03	3.341,00
			Ties	0(c)		
		Total	111			
		Score WT follow up - Score WT post	Negative Ranks	80(d)	56,00	4.480,00
			Positive Ranks	21(e)	31,95	671,00
	Ties		10(f)			
	GP	Residuen prae follow up - Residuen prae post	Negative Ranks	59(a)	59,51	3.511,00
			Positive Ranks	55(b)	55,35	3.044,00
			Ties	0(c)		
		Total	114			
		Score WT follow up - Score WT post	Negative Ranks	85(d)	55,45	4.713,50
Positive Ranks			19(e)	39,29	746,50	
Ties	10(f)					
Total	114					

- a. Residuen prae follow up < Residuen prae post
- b. Residuen prae follow up > Residuen prae post
- c. Residuen prae follow up = Residuen prae post
- d. Score WT follow up < Score WT post
- e. Score WT follow up > Score WT post
- f. Score WT follow up = Score WT post

		Test Statistics(c)	
Untersuchung	Gruppe	Residuen prae follow up - Residuen prae post	Score WT follow up - Score WT post
Voruntersuchung	FU	Z	-0,348(a)
		Asymp. Sig. (2-tailed)	0,728
Hauptuntersuchung I	GP	Z	-3,350(b)
		Asymp. Sig. (2-tailed)	0,726
Hauptuntersuchung II intern	FU	Z	-5,578(b)
		Asymp. Sig. (2-tailed)	0,563
Hauptuntersuchung II extern	GP	Z	-1,206(a)
		Asymp. Sig. (2-tailed)	0,228
Hauptuntersuchung II intern	FU	Z	-3,354(b)
		Asymp. Sig. (2-tailed)	0,723
Hauptuntersuchung II extern	GP	Z	-1,077(a)
		Asymp. Sig. (2-tailed)	0,282
Hauptuntersuchung II extern	FU	Z	-6,886(b)
		Asymp. Sig. (2-tailed)	0,493
Hauptuntersuchung II extern	GP	Z	-6,660(a)
		Asymp. Sig. (2-tailed)	0,509

- a. Based on positive ranks.
- b. Based on negative ranks.
- c. Wilcoxon Signed Ranks Test

Nachhaltigkeit des erworbenen Wissens

Univariate Varianzanalyse Abstand					
Levene-Test auf Gleichheit der Fehlervarianzen(a)					
Abhängige Variable: Residuen prae follow up					
Untersuchung	F	df1	df2	Signifikanz	
Voruntersuchung	1,027	3	63	0,387	
Hauptuntersuchung I	3,114	3	145	0,028	
Hauptuntersuchung II intern	0,794	3	176	0,499	
Hauptuntersuchung II extern	1,247	3	221	0,294	

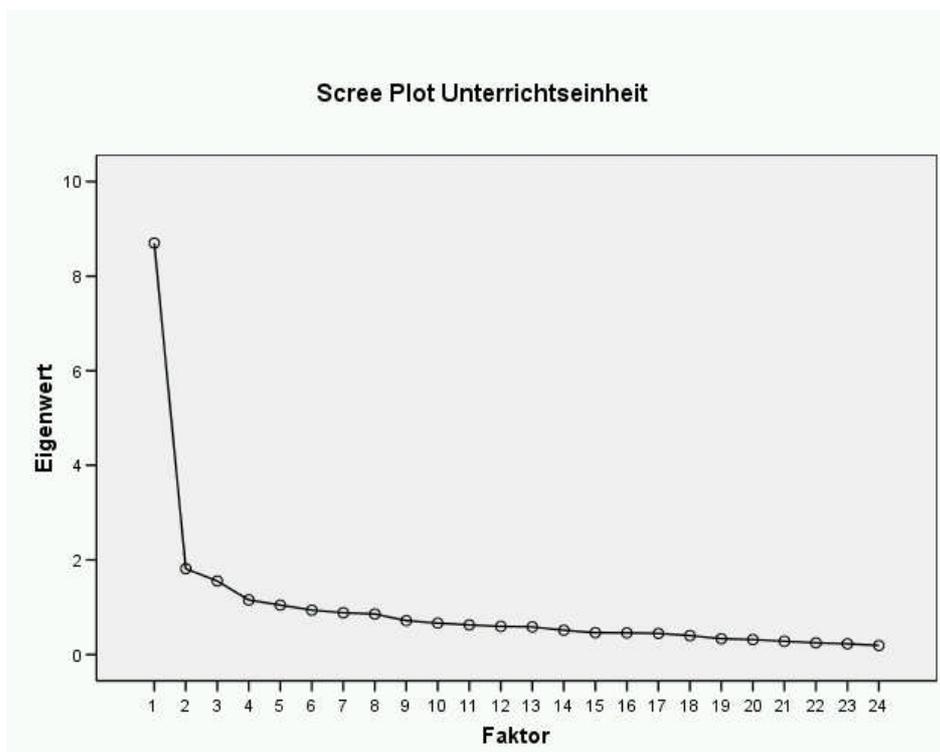
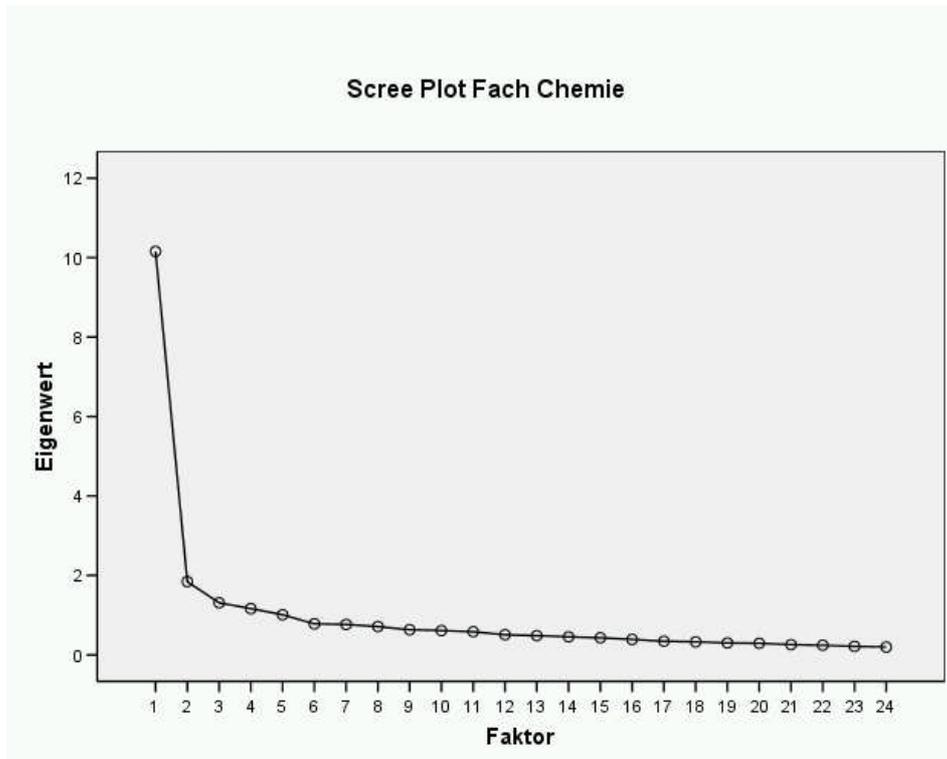
Prüft die Nullhypothese, daß die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.
a. Design: Intercept+gruppe+Abstand_md+gruppe * Abstand_md

Tests der Zwischensubjekteffekte							
Abhängige Variable: Residuen prae follow up							
Untersuchung	Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
Voruntersuchung	gruppe	5,320	1	5,320	5,712	0,020	0,083
	Abstand_md	0,159	1	0,159	0,171	0,680	0,003
	gruppe * Abstand_md	0,854	1	0,854	0,917	0,342	0,014
Hauptuntersuchung I	gruppe	0,236	1	0,236	0,269	0,605	0,002
	Abstand_md	19,144	1	19,144	21,884	0,000	0,131
	gruppe * Abstand_md	0,779	1	0,779	0,891	0,347	0,006
Hauptuntersuchung II intern	gruppe	21,279	1	21,279	24,210	0,000	0,121
	Abstand_md	1,397	1	1,397	1,589	0,209	0,009
	gruppe * Abstand_md	0,630	1	0,630	0,716	0,399	0,004
Hauptuntersuchung II extern	gruppe	12,971	1	12,971	13,660	0,000	0,058
	Abstand_md	0,132	1	0,132	0,139	0,710	0,001
	gruppe * Abstand_md	0,033	1	0,033	0,035	0,852	0,000

a. R-Quadrat = ,097 (korrigiertes R-Quadrat = ,054)
b. R-Quadrat = ,137 (korrigiertes R-Quadrat = ,119)
c. R-Quadrat = ,131 (korrigiertes R-Quadrat = ,116)
d. R-Quadrat = ,059 (korrigiertes R-Quadrat = ,046)

12.4 Statistische Berechnungen Meinungstest

12.4.1 Faktorenanalyse Hauptuntersuchung post (umkod. Rohdaten)



12.4.2 Einstellung zum Fach Chemie – t-Test, U-Test

T-Test Residuen Fach Chemie											
Test bei unabhängigen Stichproben											
Untersuchung		Levene-Test der Varianzgleichheit	T-Test für die Mittelwertgleichheit					Differenz			
			F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
Voruntersuchung	Residuen Fach Chemie	Varianzen sind gleich	4,194	0,045	-0,107	66	0,915	-0,02597577	0,24261953	-0,51038129	0,45842975
		Varianzen sind nicht gleich			-0,108	60,130	0,914	-0,02597577	0,23996657	-0,50595895	0,45400741
Hauptuntersuchung I	Residuen Fach Chemie	Varianzen sind gleich	2,459	0,119	0,616	156	0,539	0,09786202	0,15893119	-0,21607276	0,41179681
		Varianzen sind nicht gleich			0,614	149,830	0,540	0,09786202	0,15929171	-0,21688622	0,41281026
Hauptuntersuchung II intern	Residuen Fach Chemie	Varianzen sind gleich	0,206	0,651	2,423	179	0,016	0,35442897	0,14630015	0,06573409	0,64312384
		Varianzen sind nicht gleich			2,419	175,898	0,017	0,35442897	0,14654332	0,06521952	0,64363841
Hauptuntersuchung II extern	Residuen Fach Chemie	Varianzen sind gleich	3,014	0,084	-0,462	231	0,645	-0,06050108	0,13099407	-0,31859694	0,19759478
		Varianzen sind nicht gleich			-0,460	221,907	0,646	-0,06050108	0,13144628	-0,31954383	0,19854167

U-Test Residuen + Scores Fach Chemie			
Statistik für Test(a)			
Untersuchung		Score Fach Chemie post	Residuen Fach Chemie
Voruntersuchung	Mann-Whitney-U	571,000	571,500
	Wilcoxon-W	1.201,000	1.201,500
	Z	-0,679	-0,074
	Asymptotische Signifikanz (2-seitig)	0,497	0,941
Hauptuntersuchung I	Mann-Whitney-U	2.916,000	3.001,500
	Wilcoxon-W	5.997,000	6.241,500
	Z	-0,838	-0,412
	Asymptotische Signifikanz (2-seitig)	0,402	0,680
Hauptuntersuchung II intern	Mann-Whitney-U	3.516,000	3.217,000
	Wilcoxon-W	7.794,000	7.495,000
	Z	-1,642	-2,489
	Asymptotische Signifikanz (2-seitig)	0,101	0,013
Hauptuntersuchung II extern	Mann-Whitney-U	6.432,000	6.537,000
	Wilcoxon-W	12.987,000	13.092,000
	Z	-0,683	-0,478
	Asymptotische Signifikanz (2-seitig)	0,494	0,632

a. Gruppvariable: gruppe

12.4.3 Einstellung zum Fach Chemie – Effektstärken

Untersuchung	gruppe	Fach Chemie			Effektgröße d Scores	Effektgröße d Residuen
		Mittelwert	Score Fach Chemie post	Residuen Fach Chemie		
Voruntersuchung	FU	Mittelwert	0,519815668	-0,013369882	0,16	0,03
		Standardabweichung	0,224006301	0,790333446		
	GP	Mittelwert	0,554659498	0,012605889		
		Standardabweichung	0,213115303	1,163166024		
	Insgesamt	Mittelwert	0,537482962	-2,99661E-16		
		Standardabweichung	0,217691147	0,992509258		
Hauptuntersuchung I	FU	Mittelwert	0,58807135	0,049550391	0,11	0,10
		Standardabweichung	0,272826061	1,085213998		
	GP	Mittelwert	0,589148327	-0,048311631		
		Standardabweichung	0,281568266	0,906653916		
	Insgesamt	Mittelwert	0,574263969	7,99901E-16		
		Standardabweichung	0,276854598	0,996810199		
Hauptuntersuchung II intern	FU	Mittelwert	0,485018727	0,180151739	0,23	0,36
		Standardabweichung	0,260984018	1,032720915		
	GP	Mittelwert	0,423460145	-0,174277226		
		Standardabweichung	0,265346806	0,93446913		
	Insgesamt	Mittelwert	0,453729282	3,41595E-15		
		Standardabweichung	0,264286799	0,997218353		
Hauptuntersuchung II extern	FU	Mittelwert	0,533260234	-0,030899693	0,10	0,06
		Standardabweichung	0,279820527	1,07835685		
	GP	Mittelwert	0,561046158	0,029601387		
		Standardabweichung	0,272344221	0,917736881		
	Insgesamt	Mittelwert	0,547451328	-6,51184E-16		
		Standardabweichung	0,275782745	0,9978425		

12.4.4 Einstellung zum Fach Chemie – Reliabilität

Reliabilität Fach Chemie post VU

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	71	95,9
	Ausgeschlossen(a)	3	4,1
	Insgesamt	74	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,913	31

HU I

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	156	96,9
	Ausgeschlossen(a)	5	3,1
	Insgesamt	161	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,934	24

HU II intern

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	181	99,5
	Ausgeschlossen(a)	1	0,5
	Insgesamt	182	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,924	24

HU II extern

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	230	92,0
	Ausgeschlossen(a)	20	8,0
	Insgesamt	250	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,938	24

12.4.5 Attraktivität der Unterrichtseinheit – t-Test, U-Test

T-Test Unterrichtseinheit											
Test bei unabhängigen Stichproben											
Untersuchung	Score 3 Stunden post	Varianzen sind gleich	Varianzgleichheit				T-Test für die Mittelwertgleichheit				
			F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Konfidenzintervall	
Voruntersuchung	Score 3 Stunden post	Varianzen sind gleich	0,365	0,548	-4,744	69	0,000	-0,22011	0,04640	-0,31267	-0,12755
		Varianzen sind nicht gleich			-4,750	68,805	0,000	-0,22011	0,04634	-0,31257	-0,12765
Hauptuntersuchung I	Score 3 Stunden post	Varianzen sind gleich	3,134	0,079	-2,785	157	0,006	-0,10060	0,03612	-0,17194	-0,02926
		Varianzen sind nicht gleich			-2,775	149,590	0,006	-0,10060	0,03625	-0,17222	-0,02898
Hauptuntersuchung II intern	Score 3 Stunden post	Varianzen sind gleich	0,151	0,698	-1,480	179	0,141	-0,04885	0,03299	-0,11395	0,01626
		Varianzen sind nicht gleich			-1,482	178,844	0,140	-0,04885	0,03296	-0,11388	0,01619
Hauptuntersuchung II extern	Score 3 Stunden post	Varianzen sind gleich	2,627	0,106	-5,556	229	0,000	-0,17795	0,03203	-0,24106	-0,11484
		Varianzen sind nicht gleich			-5,541	223,098	0,000	-0,17795	0,03211	-0,24124	-0,11467

U-Test Unterrichtseinheit		
Statistik für Test(a)		
Untersuchung		Score 3 Stunden post
Voruntersuchung	Mann-Whitney-U	275,500
	Wilcoxon-W	905,500
	Z	-4,082
	Asymptotische Signifikanz (2-seitig)	0,000
Hauptuntersuchung I	Mann-Whitney-U	2.388,000
	Wilcoxon-W	5.469,000
	Z	-2,660
	Asymptotische Signifikanz (2-seitig)	0,008
Hauptuntersuchung II intern	Mann-Whitney-U	3.583,500
	Wilcoxon-W	7.588,500
	Z	-1,452
	Asymptotische Signifikanz (2-seitig)	0,147
Hauptuntersuchung II extern	Mann-Whitney-U	4.073,000
	Wilcoxon-W	10.514,000
	Z	-5,116
	Asymptotische Signifikanz (2-seitig)	0,000

a. Gruppvariable: gruppe

12.4.6 Attraktivität der Unterrichtseinheit – Effektstärken

Untersuchung	gruppe	Unterrichtseinheit		Effektgröße d Scores	
		Score 3 Stunden post	MW Likert 3 Stunden post		
Voruntersuchung	FU	Mittelwert	0,3734447	2,494210829	
		N	35	35	
		Standardabweichung	0,187186885	0,391783768	
	GP	Mittelwert	0,593553987	2,047407034	
		N	36	36	
		Standardabweichung	0,203165511	0,388219056	
	Insgesamt	Mittelwert	0,485049409	2,267662426	
		N	71	71	
		Standardabweichung	0,223471783	0,447797769	
	Hauptuntersuchung I	FU	Mittelwert	0,482905983	2,357371795
			N	78	78
			Standardabweichung	0,248320104	0,553932844
GP		Mittelwert	0,583506157	2,127549651	
		N	81	81	
		Standardabweichung	0,205879864	0,416277068	
Insgesamt		Mittelwert	0,534155128	2,24029259	
		N	159	159	
		Standardabweichung	0,232502964	0,50055986	
Hauptuntersuchung II intern		FU	Mittelwert	0,487386691	2,253616403
			N	89	89
			Standardabweichung	0,214711115	0,453115565
	GP	Mittelwert	0,536231884	2,149456522	
		N	92	92	
		Standardabweichung	0,228653471	0,503541797	
	Insgesamt	Mittelwert	0,512214082	2,200673259	
		N	181	181	
		Standardabweichung	0,222642032	0,480924151	
	Hauptuntersuchung II extern	FU	Mittelwert	0,472072592	2,554884163
			N	113	113
			Standardabweichung	0,257819576	0,544482113
GP		Mittelwert	0,650024564	2,213967698	
		N	118	118	
		Standardabweichung	0,228649591	0,512703005	
Insgesamt		Mittelwert	0,562974465	2,380736359	
		N	231	231	
		Standardabweichung	0,258671236	0,554301485	

12.4.7 Attraktivität der Unterrichtseinheit – Reliabilität

Reliabilität 3 Stunden post

VU

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	69	93,2
	Ausgeschlossen(a)	5	6,8
	Insgesamt	74	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,914	32

HU I

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	156	96,9
	Ausgeschlossen(a)	5	3,1
	Insgesamt	161	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,914	24

HU II intern

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	179	98,4
	Ausgeschlossen(a)	3	1,6
	Insgesamt	182	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,900	24

HU II extern

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	227	90,8
	Ausgeschlossen(a)	23	9,2
	Insgesamt	250	100,0

in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,923	24

12.4.8 Attraktivität des Gruppenpuzzles – Reliabilität

Reliabilität GP - negative umkodiert

VU

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	35	47,3
	Ausgeschlossen(a)	39	52,7
	Insgesamt	74	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,834	31

HU I

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	79	49,1
	Ausgeschlossen(a)	82	50,9
	Insgesamt	161	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,851	19

HU II intern

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	92	50,5
	Ausgeschlossen(a)	90	49,5
	Insgesamt	182	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,886	23

HU II extern

Zusammenfassung der Fallverarbeitung

		Anzahl	%
Fälle	Gültig	115	46,0
	Ausgeschlossen(a)	135	54,0
	Insgesamt	250	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen in der Prozedur.

Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
0,909	23

12.5 Statistische Berechnungen Individuelle Schülerbeurteilungen

U-Test auf Unterschiede der beiden Stichproben, pre

Deskriptive Statistik

Deskriptive Statistik		N	Mittelwert	Standardabweichung
Voruntersuchung	PP Score Roeder Mitarbeit pre	74	0,5222	0,33751
	Gültige Werte (Listenweise)	74		
Hauptuntersuchung I	PP Score Roeder Mitarbeit pre	161	0,5708	0,33401
	Gültige Werte (Listenweise)	161		
Hauptuntersuchung II intern	PP Score Roeder Mitarbeit pre	181	0,5387	0,33181
	Gültige Werte (Listenweise)	181		
Hauptuntersuchung II extern	PP Score Roeder Mitarbeit pre	244	0,5499	0,33885
	Gültige Werte (Listenweise)	244		

Gruppenstatistiken

Untersuchung	gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	
Voruntersuchung	PP Score Roeder Mitarbeit pre	FU	37	0,5135	0,34291	0,05637
		GP	37	0,5309	0,33652	0,05532
Hauptuntersuchung I	PP Score Roeder Mitarbeit pre	FU	78	0,5621	0,33816	0,03829
		GP	83	0,5790	0,33190	0,03643
		Gesamt	161			
Hauptuntersuchung II intern	PP Score Roeder Mitarbeit pre	FU	89	0,5325	0,32831	0,03480
		GP	92	0,5446	0,33684	0,03512
		Gesamt	181			
Hauptuntersuchung II extern	PP Score Roeder Mitarbeit pre	FU	123	0,5655	0,33432	0,03014
		GP	121	0,5340	0,34405	0,03128
		Gesamt	244			

Mann-Whitney-Test

Mann-Whitney-Test		Ränge	N	Mittlerer Rang	Rangsumme
Voruntersuchung	PP Score Roeder Mitarbeit pre	FU	37	37,16	1.375,00
		GP	37	37,84	1.400,00
		Gesamt	74		
Hauptuntersuchung I	PP Score Roeder Mitarbeit pre	FU	78	79,79	6.223,50
		GP	83	82,14	6.817,50
		Gesamt	161		
Hauptuntersuchung II intern	PP Score Roeder Mitarbeit pre	FU	89	89,88	7.999,00
		GP	92	92,09	8.472,00
		Gesamt	181		
Hauptuntersuchung II extern	PP Score Roeder Mitarbeit pre	FU	123	126,04	15.502,50
		GP	121	118,90	14.387,50
		Gesamt	244		

Statistik für Test(a)

Untersuchung	Statistik für Test(a)	PP Score Roeder Mitarbeit pre
Voruntersuchung	Mann-Whitney-U	672,000
	Wilcoxon-W	1.375,000
	Z	-0,136
	Asymptotische Signifikanz (2-seitig)	0,892
Hauptuntersuchung I	Mann-Whitney-U	3.142,500
	Wilcoxon-W	6.223,500
	Z	-0,320
	Asymptotische Signifikanz (2-seitig)	0,749
Hauptuntersuchung II intern	Mann-Whitney-U	3.994,000
	Wilcoxon-W	7.999,000
	Z	-0,284
	Asymptotische Signifikanz (2-seitig)	0,776
Hauptuntersuchung II extern	Mann-Whitney-U	7.006,500
	Wilcoxon-W	14.387,500
	Z	-0,791
	Asymptotische Signifikanz (2-seitig)	0,429

a. Gruppenvariable: gruppe

12.5.1 Reliabilität

Reliabilität Mitarbeit pre Zusammenfassung der Fallverarbeitung

Untersuchung			Anzahl	%
Voruntersuchung	Fälle	Gültig	71	95,9
		Ausgeschlossen(a)	3	4,1
		Insgesamt	74	100,0
Hauptuntersuchung I	Fälle	Gültig	150	93,2
		Ausgeschlossen(a)	11	6,8
		Insgesamt	161	100,0
Hauptuntersuchung II intern	Fälle	Gültig	145	79,7
		Ausgeschlossen(a)	37	20,3
		Insgesamt	182	100,0
Hauptuntersuchung II extern	Fälle	Gültig	240	96,0
		Ausgeschlossen(a)	10	4,0
		Insgesamt	250	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen in der Prozedur.

Reliabilitätsstatistiken

Untersuchung	Cronbachs Alpha	Anzahl der Items
Voruntersuchung	0,945	14
Hauptuntersuchung I	0,969	18
Hauptuntersuchung II intern	0,957	18
Hauptuntersuchung II extern	0,957	18

12.5.2 Effektstärken

Individuelle S-Beurteilungen pre					Effektgröße d
Deskriptive Statistik					
	N	Mittelwert	Standardabweichung		
Voruntersuchung	74	0,522200772	0,337508359	$d = \frac{M_{IG} - M_{KG}}{S_{gesamt}}$	0,05
Hauptuntersuchung I	161	0,570835059	0,334008068		0,05
Hauptuntersuchung II intern	181	0,538692449	0,331805473		0,04
Hauptuntersuchung II extern	244	0,54989071	0,338851007		0,06
Gruppenstatistiken					
	Gruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Voruntersuchung	FU	37	0,51	0,342905849	0,017513871
	GP	37	0,53	0,336524074	0,016966374
Hauptuntersuchung I	FU	78	0,56	0,338162137	0,026095274
	GP	83	0,58	0,331901992	0,017169998
Hauptuntersuchung II intern	FU	89	0,53	0,328310301	0,028488826
	GP	92	0,54	0,336842535	0,020882298
Hauptuntersuchung II extern	FU	123	0,57	0,334322863	0,194238165
	GP	121	0,53	0,344048331	0,129232929

12.5.3 Korrelationen Scores – Chemienote

Korrelation Individuelle S-Beurteilungen Score Mitarbeit pre - Chemienote

Korrelationen						
Untersuchung	Gruppe					Chemienote
Voruntersuchung	FU	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,689(**)
					Sig. (2-seitig)	0,000
					N	36
	GP	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,637(**)
					Sig. (2-seitig)	0,000
					N	36
Hauptuntersuchung I	FU	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,730(**)
					Sig. (2-seitig)	0,000
					N	77
	GP	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,728(**)
					Sig. (2-seitig)	0,000
					N	81
Hauptuntersuchung II intern	FU	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,725(**)
					Sig. (2-seitig)	0,000
					N	88
	GP	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,799(**)
					Sig. (2-seitig)	0,000
					N	92
Hauptuntersuchung II extern	FU	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,748(**)
					Sig. (2-seitig)	0,000
					N	114
	GP	Spearman-Rho	PP Score	Mitarbeit pre	Korrelationskoeffizient	-,660(**)
					Sig. (2-seitig)	0,000
					N	119

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

12.5.4 Test auf Normalverteilung

Test auf Normalverteilung Individuelle S-Burteilungen pre Scores Mitarbeit					
Kolmogorov-Smirnov-Anpassungstest					
Untersuchung	Gruppe			PP Score Mitarbeit pre	
Voruntersuchung	FU	N		37	
		Parameter der Normalverteilung(a,b)	Mittelwert	0,5135	
			Standardabweichung	0,34291	
		Extremste Differenzen	Absolut	0,193	
			Positiv	0,152	
			Negativ	-0,193	
			Kolmogorov-Smirnov-Z	1,175	
			Asymptotische Signifikanz (2-seitig)	0,126	
		GP	N		37
			Parameter der Normalverteilung(a,b)	Mittelwert	0,5309
			Standardabweichung	0,33652	
	Extremste Differenzen		Absolut	0,158	
			Positiv	0,118	
		Negativ	-0,158		
		Kolmogorov-Smirnov-Z	0,962		
		Asymptotische Signifikanz (2-seitig)	0,313		
Hauptuntersuchung I	FU	N		78	
		Parameter der Normalverteilung(a,b)	Mittelwert	0,5621	
			Standardabweichung	0,33816	
		Extremste Differenzen	Absolut	0,161	
			Positiv	0,110	
			Negativ	-0,161	
			Kolmogorov-Smirnov-Z	1,424	
			Asymptotische Signifikanz (2-seitig)	0,035	
		GP	N		83
			Parameter der Normalverteilung(a,b)	Mittelwert	0,5790
			Standardabweichung	0,33190	
	Extremste Differenzen		Absolut	0,159	
			Positiv	0,112	
		Negativ	-0,159		
		Kolmogorov-Smirnov-Z	1,449		
		Asymptotische Signifikanz (2-seitig)	0,030		
Hauptuntersuchung II intern	FU	N		89	
		Parameter der Normalverteilung(a,b)	Mittelwert	0,5325	
			Standardabweichung	0,32831	
		Extremste Differenzen	Absolut	0,120	
			Positiv	0,101	
			Negativ	-0,120	
			Kolmogorov-Smirnov-Z	1,128	
			Asymptotische Signifikanz (2-seitig)	0,157	
		GP	N		92
			Parameter der Normalverteilung(a,b)	Mittelwert	0,5446
			Standardabweichung	0,33684	
	Extremste Differenzen		Absolut	0,114	
			Positiv	0,097	
		Negativ	-0,114		
		Kolmogorov-Smirnov-Z	1,096		
		Asymptotische Signifikanz (2-seitig)	0,181		
Hauptuntersuchung II extern	FU	N		123	
		Parameter der Normalverteilung(a,b)	Mittelwert	0,5655	
			Standardabweichung	0,33432	
		Extremste Differenzen	Absolut	0,127	
			Positiv	0,124	
			Negativ	-0,127	
			Kolmogorov-Smirnov-Z	1,414	
			Asymptotische Signifikanz (2-seitig)	0,037	
		GP	N		121
			Parameter der Normalverteilung(a,b)	Mittelwert	0,5340
			Standardabweichung	0,34405	
	Extremste Differenzen		Absolut	0,141	
			Positiv	0,099	
		Negativ	-0,141		
		Kolmogorov-Smirnov-Z	1,550		
		Asymptotische Signifikanz (2-seitig)	0,016		

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

12.5.5 U-Test auf Unterschiede in der Mitarbeit

Ergebnisse U-Test Individuelle Schülerbeurteilungen Leistung

Descriptive Statistics

Untersuchung		N	Mean
Hauptuntersuchung I	MWLei_pr	161	3,2501
	MWLei_po	134	3,1236
	Standardized Residual	134	0,0000000
	Valid N (listwise)	134	
Hauptuntersuchung II intern	MWLei_pr	181	3,3215
	MWLei_po	170	3,2098
	Standardized Residual	170	0,0000000
	Valid N (listwise)	170	
Hauptuntersuchung II extern	MWLei_pr	244	3,5996
	MWLei_po	104	3,8674
	Standardized Residual	104	0,0000000
	Valid N (listwise)	104	

Mann-Whitney Test

Ranks

Untersuchung	Gruppe	N	Mean Rank	Sum of Ranks	
Hauptuntersuchung I	MWLei_pr	FU	78	82,68	6.449,00
		GP	83	79,42	6.592,00
		Total	161		
	MWLei_po	FU	65	69,22	4.499,50
		GP	69	65,88	4.545,50
		Total	134		
Standardized Residual	FU	FU	65	68,30	4.439,50
		GP	69	66,75	4.605,50
		Total	134		
	MWLei_pr	FU	89	92,13	8.199,50
		GP	92	89,91	8.271,50
		Total	181		
MWLei_po	FU	FU	81	86,15	6.978,50
		GP	89	84,90	7.556,50
		Total	170		
	Standardized Residual	FU	81	87,01	7.048,00
		GP	89	84,12	7.487,00
		Total	170		
Hauptuntersuchung II extern	MWLei_pr	FU	123	119,74	14.728,50
		GP	121	125,30	15.161,50
		Total	244		
	MWLei_po	FU	51	54,05	2.756,50
		GP	53	51,01	2.703,50
		Total	104		
Standardized Residual	FU	51	53,30	2.718,50	
	GP	53	51,73	2.741,50	
	Total	104			

Test Statistics(a)

Untersuchung		MWLei_pr	MWLei_po	Standardized Residual
Hauptuntersuchung I	Mann-Whitney U	3.106,000	2.130,500	2.190,500
	Wilcoxon W	6.592,000	4.545,500	4.605,500
	Z	-0,443	-0,499	-0,231
	Asymp. Sig. (2-tailed)	0,658	0,618	0,817
Hauptuntersuchung II intern	Mann-Whitney U	3.993,500	3.551,500	3.482,000
	Wilcoxon W	8.271,500	7.556,500	7.487,000
	Z	-0,285	-0,165	-0,382
	Asymp. Sig. (2-tailed)	0,775	0,869	0,702
Hauptuntersuchung II extern	Mann-Whitney U	7.102,500	1.272,500	1.310,500
	Wilcoxon W	14.728,500	2.703,500	2.741,500
	Z	-0,615	-0,514	-0,267
	Asymp. Sig. (2-tailed)	0,539	0,607	0,790

a. Grouping Variable: Gruppe

12.5.6 U-Test auf Unterschiede im Verhalten

U-Test Individuelle Schülerbeurteilungen Verhalten

Descriptive Statistics

Untersuchung		N	Mean
Hauptuntersuchung I	MWVer_pr	161	3,0587
	MWVer_po	134	3,0321
	Standardized Residual	134	0,0000000
	Valid N (listwise)	134	
Hauptuntersuchung II intern	MWVer_pr	181	3,2200
	MWVer_po	170	3,1048
	Standardized Residual	170	0,0000000
	Valid N (listwise)	170	
Hauptuntersuchung II extern	MWVer_pr	244	3,1663
	MWVer_po	104	3,4532
	Standardized Residual	104	0,0000000
	Valid N (listwise)	104	

Mann-Whitney Test

Ranks

Untersuchung		Gruppe	N	Mean Rank	Sum of Ranks	
Hauptuntersuchung I	MWVer_pr	FU	78	81,57	6.362,50	
		GP	83	80,46	6.678,50	
		Total	161			
	MWVer_po	FU	65	66,45	4.319,50	
		GP	69	68,49	4.725,50	
		Total	134			
Standardized Residual	FU	FU	65	63,35	4.117,50	
		GP	69	71,41	4.927,50	
		Total	134			
	Hauptuntersuchung II intern	MWVer_pr	FU	89	95,04	8.458,50
			GP	92	87,09	8.012,50
			Total	181		
MWVer_po		FU	81	89,53	7.252,00	
		GP	89	81,83	7.283,00	
		Total	170			
Standardized Residual	FU	FU	81	86,40	6.998,50	
		GP	89	84,68	7.536,50	
		Total	170			
	Hauptuntersuchung II extern	MWVer_pr	FU	123	121,48	14.942,50
			GP	121	123,53	14.947,50
			Total	244		
MWVer_po		FU	51	53,61	2.734,00	
		GP	53	51,43	2.726,00	
		Total	104			
Standardized Residual	FU	51	51,80	2.642,00		
	GP	53	53,17	2.818,00		
	Total	104				

Test Statistics(a)

Untersuchung		MWVer_pr	MWVer_po	Standardized Residual
Hauptuntersuchung I	Mann-Whitney U	3.192,500	2.174,500	1.972,500
	Wilcoxon W	6.678,500	4.319,500	4.117,500
	Z	-0,151	-0,303	-1,202
	Asymp. Sig. (2-tailed)	0,880	0,762	0,229
Hauptuntersuchung II intern	Mann-Whitney U	3.734,500	3.278,000	3.531,500
	Wilcoxon W	8.012,500	7.283,000	7.536,500
	Z	-1,020	-1,019	-0,228
	Asymp. Sig. (2-tailed)	0,308	0,308	0,820
Hauptuntersuchung II extern	Mann-Whitney U	7.316,500	1.295,000	1.316,000
	Wilcoxon W	14.942,500	2.726,000	2.642,000
	Z	-0,227	-0,368	-0,231
	Asymp. Sig. (2-tailed)	0,820	0,713	0,817

a. Grouping Variable: Gruppe

12.6 Audioaufzeichnungen

Rededauern in min pro Unterrichtsstunde

Unterrichtsstunde	Methode	N	Rededauer in ms pro Unterrichtsstunde		Rededauer in min pro Unterrichtsstunde
			Insgesamt	Mittelwert	Insgesamt
1	FU	183	609.806	3.332,27	10
2	FU	166	604.486	3.641,48	10
3	FU	175	755.723	4.318,42	13
4	GP	58	263.550	4.543,97	4
5	GP	79	341.924	4.328,15	6
6	GP	55	206.172	3.748,58	3
7	FU	125	627.870	5.022,96	10
8	FU	162	814.028	5.024,86	14
9	FU	182	1.032.251	5.671,71	17
10	GP	43	292.518	6.802,74	5
11	GP	78	509.792	6.535,79	8
12	GP	75	237.880	3.171,73	4
13	FU	198	995.318	5.026,86	17
14	FU	168	853.360	5.079,52	14
15	FU	155	843.034	5.438,93	14
16	GP	41	327.287	7.982,61	5
17	GP	77	267.656	3.476,05	4
18	GP	102	911.018	8.931,55	15
19	FU	104	979.476	9.418,04	16
20	FU	162	911.036	5.623,68	15
21	FU	148	945.921	6.391,36	16
22	GP	35	388.624	11.103,54	6
23	GP	16	221.951	13.871,94	4
24	GP	40	115.117	2.877,93	2
25	FU	182	983.909	5.406,09	16
26	FU	181	1.000.149	5.525,69	17
27	FU	193	986.270	5.110,21	16
28	GP	47	338.515	7.202,45	6
29	GP	43	286.361	6.659,56	5
30	GP	21	120.807	5.752,71	2

MW FU	14
MW GP	5