

MARKUS TEPNER, BURKHARD ROEDER UND INSA MELLE

## Effektivität des Gruppenpuzzles im Chemieunterricht der Sekundarstufe I

Effectiveness of jigsaw-classroom in lower secondary classes of chemistry

### Zusammenfassung

In diesem Beitrag wird eine Studie vorgestellt, durch die untersucht werden sollte, 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. 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. Unmittelbar vor bzw. nach Durchführung der Unterrichtsreihe wurden Wissens- und anonyme Einstellungstests eingesetzt. Insgesamt nahmen 413 Schüler von 10. Gymnasialklassen an der Studie teil. Es zeigt sich u. a., dass das Gruppenpuzzle zu einem signifikant höheren Wissenszuwachs als vergleichbarer Frontalunterricht führt.

Schlüsselwörter: Gruppenpuzzle, Chemieunterricht, Kooperatives Lernen, Einstellung, Randomisierung, Matched-Pairs-Prinzip

### Abstract

This article presents a study which was aimed at the question, whether the use of jigsaw-classroom in lower secondary classes of chemistry can increase the students' learning outcome and alter their attitude towards the subject. In a real classroom situation the experimental group was compared to a conventionally taught learning group. In order to provide maximum comparability, the classes involved in this study were divided into two halves according to the principle of matched pairs. Their regular teacher instructed the two groups consecutively on the topic of soaps. The control group was taught in a "conventional" way, i. e. through direct teaching, while the treatment group was instructed via the jigsaw-classroom. The contents instructed were identical in both cases. Immediately before and after the teaching unit, knowledge tests and anonymous opinion tests have been conducted. All in all, 413 students from tenth-grade grammar-school classes participated in this study. Inter alia, it could be shown that group-puzzles lead to a significantly higher increase in knowledge than comparable teacher-centered learning.

Keywords: Jigsaw-classroom, chemistry lesson, cooperative learning, attitude, randomization, matched pairs.

---

## 1 Einleitung

Kooperative Lernformen wie z. B. Gruppenrallye, Kugellager oder Gruppenpuzzle gehören mittlerweile zum festen Bestandteil aktueller Richtlinien und Lehrpläne. Mit ihrem Einsatz wächst auch die Frage nach ihrem Erfolg bzw. ihrer Effektivität.

Dabei geht es einmal z. B. um die Stärkung des Selbstvertrauens, das Training sozialer Kompetenzen oder die Förderung der Lernzufriedenheit. Andererseits muss aber auch der Leistungsaspekt und damit auch die Effektivität in Bezug auf den Wissenszuwachs in den Blick genommen werden.

Zur Effektivität kooperativer Unterrichtsformen gibt es vor allem außerhalb Europas bereits zahlreiche Studien bzw. Meta-Analysen (vgl. Johnson & Johnson, 1989, Slavin, 1995 oder Lazarowitz & Hertz-Lazarowitz, 1998), in denen dieses Unterrichtsprinzip überwiegend positiv abschneidet.

Kooperatives Lernen in Deutschland – speziell im Chemieunterricht – ist bisher nur wenig empirisch untersucht (z. B. Sumfleth, Rumann & Nicolai, 2004); es existieren einige qualitative Untersuchungen (z. B. Eilks, 2003a und 2003b).

Da Ergebnisse internationaler Studien nicht ohne weiteres auf das deutsche Schulsystem übertragen werden können, sollte in der hier beschriebenen 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 an dieser Stelle der im weitesten Sinne darlegend-informative bzw. fragend-entwickelnde, lehrerzentrierte Unterricht verstanden werden.

## 2 Forschungsstand

### 2.1 Untersuchungen zum kooperativen Lernen

Kooperative Lernformen sind besonders in den USA und in Kanada vielfältig untersucht worden; Johnson, Johnson & Stanne (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. Die wichtigsten Ergebnisse bezüglich des kooperativen Lernens lassen sich folgendermaßen zusammenfassen, wobei die Studien allerdings stellenweise uneinheitlich sind (vgl. Eilks, Witteck, Rumann & Sumfleth, 2005, Lazarowitz, Hertz-Lazarowitz & Baird, 1994 und

Lazarowitz & Hertz-Lazarowitz, 1998, Green & Green (2005, 33-37):

#### 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.

Slavin (1995) prüft mehrere Dutzend Studien, die vier Wochen oder länger dauerten und unterschiedliche kooperative Methoden untersuchten. Insgesamt schneiden kooperativ unterrichtete Schüler in Wissenstests besser ab. Dabei hängt der positive Effekt maßgeblich von der verwendeten Methode ab (Mifflin, 2008).

Eine aktuelle Meta-Studie stammt von Johnson und Mitarbeitern aus dem Jahr 2000 (Johnson & Johnson, 2008). Dort werden 164 Studien aus den Jahren 1970 bis 1999 analysiert, in denen acht verschiedene kooperative Lernformen mit wettbewerblichem bzw. individualistischem Lernen hinsichtlich des Lernerfolges verglichen werden. Es zeigt sich eine deutliche Überlegenheit des kooperativen Arbeitens im Vergleich zu wettbewerblichem Lernen (z. B. Frontalunterricht, bei dem die Schüler als Konkurrenten um die Gunst des Lehrers – "Drankommen" – buhlen), jeweils mit hohen Effektgrößen. Gegenüber individuellem Lernen, z. B. selbstständigem Erarbeiten, können jedoch kaum Unterschiede (vgl. auch Johnson & Johnson, 1999, 1994) nach-

gewiesen werden. Anzumerken ist jedoch, dass Johnson et al. (2000) den Frontalunterricht („traditional instruction“) entweder mit „wettbewerblich“ oder „individualistisch“ kodieren, wobei eine eindeutige Zuordnung allerdings nicht immer möglich ist. Zudem gehen durch die Kodierung weitere Informationen über den erfolgten Unterricht verloren.

Untersuchungen im Fach Chemie sind unter den internationalen Studien kaum zu finden. Anzuführen ist Bowen (2000), der 15 Untersuchungen im Fach Chemie aus den Jahren 1990 bis 1998 in einer Meta-Studie zusammenfasst. Insgesamt nahmen 437 High School-Schüler und fast 1100 College-Studenten teil. Im Durchschnitt liegen die Leistungen der kooperativ unterrichteten Probanden über denen der traditionell unterrichteten Schülern/Studenten. Die Effektgröße beträgt im Mittel 0.37.

Bis 2003 existierten für den Chemieunterricht in Deutschland kaum systematische Studien über die Auswirkungen von kooperativen Lernformen (Eilks, 2003a). Walpuski und Sumfleth (2007) untersuchen kooperative Lernformen mit dem Fokus der strukturierten Instruktion. Hierbei geht es um die Optimierung von experimentellem Arbeiten in Kleingruppen. Diese Studie basiert auf einer Untersuchung von Rumann (2005), der das Konzept der Gruppenrecherche mit Interaktionsboxen an insgesamt acht Klassen zweier Gymnasien erprobt hatte. Dabei erzielen kooperativ unterrichtete Schüler einen signifikanten Leistungsvorteil (Rumann, 2005) gegenüber der Kontrollgruppe, wobei die Effektgröße des Leistungstests relativ klein ist (vgl. Sumfleth et al., 2004).

Beschreibungen über Lernerfolge in kooperativen Lernformen im deutschen Chemieunterricht werden auch auf Basis von Erfahrungen einzelner Lehrer bzw. im Rahmen partizipativer Aktionsforschung (Eilks & Ralle, 2002) erstellt (z. B. Markic & Eilks, 2005). Auch in anderen Fächern findet empirische Unterrichtsforschung in Deutschland bisher nur in geringem Umfang statt (vgl. Helmke, 2006).

## 2.2 Gruppenpuzzle

Von den 164 Studien, die Johnson et al. (2000) in ihre Meta-Studie einbeziehen, befassen sich 14 mit dem Gruppenpuzzle und davon fünf mit dem Vergleich zwischen Gruppenpuzzle und individualistischem Unterricht. Demnach kann das Gruppenpuzzle Schülern einen Lernvorteil verschaffen. Weiterhin kann das Gruppenpuzzle die Aggressionsbereitschaft in der Klasse verringern und die Wertschätzung der Schüler untereinander erhöhen, tendenziell sind größere Lernerfolge gegenüber dem lehrerzentrierten Unterricht möglich (vgl. Frey-Eiling & Frey, 1999).

Auch in Deutschland ist das Gruppenpuzzle mittlerweile Inhalt evaluativer Studien. Eine der ersten führten Eppler, Winter & Huber bereits 1985 im Deutschunterricht einer Haupt- und Realschule durch, allerdings ohne Kontrollgruppe (vgl. Huber, 1985). Borsch, Jürgen-Lohmann & Giesen (2002) evaluieren die Effektivität des Gruppenpuzzles im Sachunterricht der Primarstufe; Jürgen-Lohmann, Borsch & Giesen (2001) untersuchen die Auswirkungen an Teilnehmern von zwei Hochschulseminaren. Für den Sekundarstufen-II-Bereich liegen Erkenntnisse von Berger & Hänze (2004a) aus der Jahrgangsstufe 12 im Fach Physik vor. Hier zeigt sich kein Unterschied zwischen dem Gruppenpuzzle und dem Frontalunterricht. Borsch et al. (2002) stellen in der Grundschule einen Vorteil des Gruppenpuzzles bezüglich des deklarativen Wissens fest. An der Hochschule und im Physikunterricht ist der Wissenszuwachs in etwa vergleichbar. Berger & Hänze (2005) vergleichen das Gruppenpuzzle zudem mit einem Lernzirkel. Dabei schneiden Schüler der Lernzirkelgruppe signifikant besser im Wissenstest ab als die Gruppenpuzzle-Gruppe.

Vorrangig qualitative Studien zum Gruppenpuzzle im Chemieunterricht haben Eilks und Mitarbeiter durchgeführt. Auch sie berichten von überwiegend positiven Erfahrungen mit dem Gruppenpuzzle (Eilks, 2003a und b, Leerhoff & Eilks, 2002).

### 3 Hypothesen

Durch die hier beschriebene Studie sollte anhand von Referenzgruppen untersucht werden, ob das Gruppenpuzzle die Lernleistung und die Einstellung der Schüler zum Unterricht im direkten Vergleich mit inhaltlich vergleichbarem Frontalunterricht erhöhen kann (Näheres dazu siehe unter 5.).

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). Da diese so genannten Soft-Skills nur schwer quantitativ erfassbar sind, sollten sie von der Untersuchung ausgenommen werden.

Folgende Hypothesen lagen der Studie zugrunde:

1. Schüler, die nach der Gruppenpuzzle-Methode unterrichtet werden, erzielen einen höheren Wissenszuwachs als inhaltsgleich frontal unterrichtete Schüler.
2. Schüler haben eine positivere Einstellung zum Unterricht nach der Methode des Gruppenpuzzles als zu inhaltlich vergleichbarem Frontalunterricht.

Außerdem wurden folgende Nebenhypothesen formuliert:

1. Das Gruppenpuzzle eignet sich hinsichtlich der Lernleistung für Schüler mit besseren und schlechteren Chemieensuren 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.

### 4 Das eingesetzte Gruppenpuzzle

Die Arbeit der Schülerinnen und Schüler im Gruppenpuzzle (vgl. Tepner & Melle, 2005) gliederte sich in drei Phasen:

- Erarbeitungsphase: Jeder Schüler erarbeitet selbstständig sein Thema in Form vorstrukturierter Lernmaterialien in sog. Stammgruppen á 3 bis 5 Schüler.
- Expertenphase: Alle „Experten“ für das jeweilige Thema setzen sich zusammen und klären offene Fragen. Außerdem soll jeder Schüler einen strukturierten Vortrag ausarbeiten, um damit sein Wissen in der nächsten Phase seinen Mitschülern zu vermitteln.
- 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.

Für die Untersuchung wurde ein Gruppenpuzzle zum Thema Seifen für die Jahrgangsstufe 10 an Gymnasien konzipiert, welches drei Unterthemen hatte (vgl. Tepner & Melle, 2005, Tepner, Melle & Roeder, 2005b und Abb. 1, Abb. 2). 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, da es dadurch leichter war, ein für alle beteiligten Lerngruppen geeignetes Gruppenpuzzle zu konzipieren. Lediglich Wasserstoffbrückenbindungen sowie das Lösen von Stoffen in Wasser sollten bekannt sein. Ein weiterer Vorteil bestand im Einsatz ungefährlicher und zugleich preiswerter Chemikalien, so dass Schülerexperimente problemlos möglich waren. Die durchgeführten Schülerexperimente wurden vorgegeben, da eine eigenständige Entwicklung durch die Schüler zwei Nachteile mit sich bringt: Erstens resultiert daraus ein größerer Zeitbedarf, zweitens birgt die selbstständige Entwicklung ein höheres Fehlerpotential, z. B. durch falsche Ansätze, fehlerhafte Durchführung oder inkorrekte Interpretationen, so dass ein korrektes Er-

gebnis nicht immer sichergestellt werden kann (vgl. Sumfleth, Rumann & Nikolai, 2004). Somit wäre die Vergleichbarkeit der Bedingungen zwischen der Gruppenpuzzle-Gruppe und der Frontalunterricht-Gruppe beeinflusst gewesen. Hinzu kommt, dass die Entwicklung von Experimenten bei dieser Untersuchung nicht im Vordergrund stand, sondern die Experimente als Einstiegs- bzw. Bestätigungsexperimente für die Experten und als Anschauungsexperimente für die Novizen fungierten.

Die Aufteilung in nur drei Unterthemen erfolgte, weil erstens die Gruppengröße aufgrund des Untersuchungsdesigns teilweise recht gering war (11 – 19 Schüler, zur Anzahl der Schüler, vgl. Kap. 5) und ansonsten die Gruppenanzahl für die Untersuchung zu klein geworden wäre. Zweitens sollte das Gruppenpuzzle in 3 x 45 Minuten durchzuführen sein.

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 (Abb. 2).

Die drei Unterthemen des Gruppenpuzzles lauten:

- Aufbau und Löslichkeit von Seife/Waschwirkung
- Herabsetzung der Oberflächenspannung durch Seife
- Was ist Seife und die Nachteile von Seife

Bei der Konzeption der Materialien wurde darauf geachtet, dass – basierend auf einer Experteneinschätzung – alle Unterthemen etwa gleich schwer waren, und pro Thema mindestens ein Versuch durchgeführt werden musste. Die Lernmaterialien wiesen eine klare Strukturierung auf. Besonders wichtige

Datum:	<b>Thema 1</b>	Vorname:
Klasse:	<b>Expertenmaterial</b>	Erster Buchstabe des Nachnamens:

**Aufbau und Löseverhalten von Seife/Waschwirkung**

**Experiment:**  
Bedecke eine Wasseroberfläche mit Hilfe eines Pfefferstreuers leicht mit gemahlenem 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:

hydrophob / unpolare

hydrophil / polare

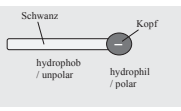


Abb. 1: Streichholzmodell

**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.

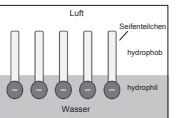


Abb. 2: Seifenteilchen an der Wasseroberfläche

**Erklärung der Versuchsbeobachtungen:**  
Beim Ausbreiten der Seifenteilchen werden die Pfefferkörnerchen „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.

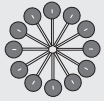


Abb. 3: Vereinfachte Darstellung einer Kugelmizelle

**Was macht die Seife mit dem Schmutz?**  
Um dreieckige 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 unpolare, 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. Schmutzteilen niederlassen, überziehen sie es schließlich mit einer neuen polaren Oberfläche, so dass sich auch unpolare Stoffe wie Schmutz in Wasser lösen (siehe Abb. 4).

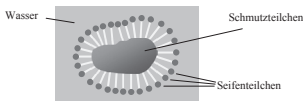


Abb. 4: Gelöster Schmutz in Wasser

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

Abb. 1: Expertenmaterial – Beispiel für Thema 1 (vgl. Tepner & Melle, 2005)

Datum:	<b>Thema 1</b>	Vorname:
Klasse:	<b>Arbeitsblatt Vermittlungsphase</b>	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, Pfefferfällchen 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 leere Reagenzglas gibst du soviel Seifenlösung hinzu, bis fast keine Luft mehr drin ist. Verschiebe mit einem Stopfen, schüttle 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?

Abb. 2: Arbeitsblatt Vermittlungsphase – Beispiel für Thema 1 (vgl. Tepner & Melle, 2005)

Sachverhalte wurden durch Unterstreichung hervorgehoben bzw. als Merksatz formuliert. Die Texte sollten in maximal 20 Minuten zu bearbeiten sein, Bilder und Zeichnungen veranschaulichten die Sachverhalte. Weiterhin befanden sich die Lernziele sowie eine Auflistung der wichtigen Begriffe auf dem Expertenmaterial, damit für die Schüler transparent ist, welche Kompetenzen von ihnen

erwartet werden (Frey-Eiling & Frey, 1999). Mit Hilfe der wichtigen Begriffe sollten die Schüler sich auf ihren Vortrag in der Vermittlungsphase vorbereiten, die Reihenfolge konnte dabei als Strukturierungshilfe dienen. Insgesamt wurde auf ein funktionelles Layout sowie eine angemessene, verständliche Sprache geachtet (Ballstaedt, 1993).

Im Zentrum der durch dieses Gruppenpuzzle angestrebten Ziele steht der Kompetenzbereich Fachwissen. Ebenfalls relevant ist der Kompetenzbereich Kommunikation, also die Beschreibung, Veranschaulichung oder Erklärung chemischer Sachverhalte unter Verwendung der Fachsprache und/oder mit Hilfe von Modellen und Darstellungen sowie die Planung, Strukturierung, Reflexion und Präsentation der Arbeit als Team. Die Kompetenzbereiche Erkenntnisgewinnung und Bewertung hatten eine untergeordnete Bedeutung. (vgl. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005)

## 5 Design und Stichprobe

Die Untersuchung wurde in zehnten Klassen nordrhein-westfälischer Gymnasien durchgeführt und gliederte sich in eine Vor- und zwei Hauptuntersuchungen (VU, HU I und HU II, Tab. 1), wobei die zweite Hauptuntersuchung ein Jahr nach der ersten stattfand, um den Befund zu replizieren.

Tab. 1: Umfang der Stichproben. Abkürzungen: m = männlich, w = weiblich, VU = Voruntersuchung, HU = Hauptuntersuchung

	VU	HU I	HU II
Lehrer	2	4	5
	m: 1, w: 1	m: 1, w: 3	m: 2, w: 3
Klassen	3	7	7
Schüler	74	158	181
	m: 35 w: 39	m: 86 w: 72	m: 81 w: 100
durchschnittliches Alter der Schüler (Jahre)	15.8	16.1	15.9

Die beteiligten Lerngruppen wurden in zwei Hälften geteilt, dieses geschah randomisiert nach dem Matched-Pairs-Prinzip. Dabei wurde auf der Grundlage von individuellen Schülerbeurteilungen (vgl. 6.1), 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 beiden Klassenhälften erhielten nacheinander von der gewohnten Lehrkraft Unterricht zum Thema Seifen, wobei dieses inhaltsgleich bei der ersten Gruppe „herkömmlich“ im Frontalunterricht („FU-Gruppe“ und bei der Experimentalgruppe nach der Gruppenpuzzle-Methode geschah („GP-

Gruppe“, siehe Abb. 3). Dabei durfte der Frontalunterricht durchaus Gruppenarbeits-elemente enthalten, jedoch nicht überwiegend. Insgesamt dominierte im Frontalunterricht das Unterrichtsgespräch.

Die Lehrer konzipierten den Frontalunterricht auf der Basis der Materialien des Gruppenpuzzles, so dass die behandelten Inhalte in beiden Gruppen identisch waren. Ferner wurden sie aufgefordert, die Experimente einzusetzen, die auch im Gruppenpuzzle enthalten sind. In jeder Stunde war der Untersuchungsleiter im Sinne einer teilnehmenden Beobachtung anwesend. Im Frontalunterricht kontrollierte er, dass alle Lerninhalte dargeboten wurden (siehe hierzu auch Tepner, Melle & Roeder, 2005a). Im abweichenden Fall wurde dieses in der Auswertung berücksichtigt (siehe Kapitel 7.1). Somit wurde die Kontrollgruppe inhaltlich soweit beeinflusst, dass die Vergleichbarkeit mit der Interventionsgruppe gegeben war. Hinsichtlich der eingesetzten Unterrichtsmethodik war die Kontrollgruppe – bis auf die Vorgabe, nicht überwiegend Gruppenunterricht durchzuführen, s. o. – prinzipiell unbeeinflusst. Auf Nachfrage teilten die beteiligten Lehrer dem Untersuchungsleiter mit, dass sie den Unterricht in der KG auch ohne Mitwirkung an dem Projekt methodisch in nahezu gleicher Form durchgeführt hätten.

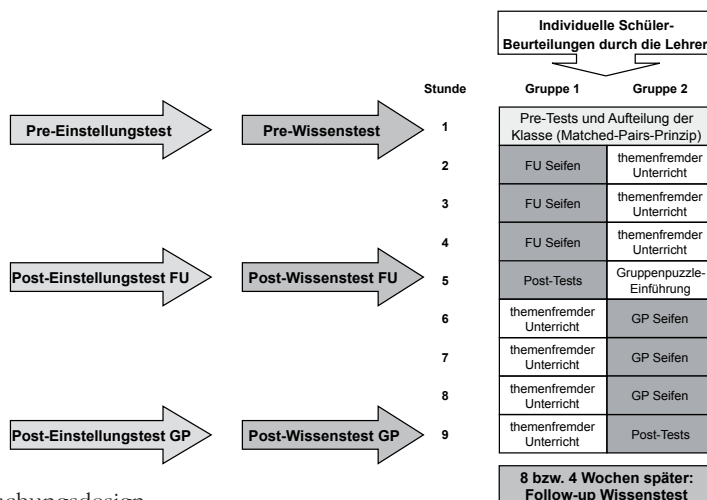


Abb. 3: Untersuchungsdesign

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 Einstellungstests eingesetzt. Mit letzteren sollte die Attraktivität der Unterrichtsmethode erfasst werden. Die Dauer eines Untersuchungsabschnitts pro Klasse betrug neun Schulstunden. Ca. acht bzw. ca. vier Wochen<sup>1</sup> nach dem jeweiligen Post-Test wurde die Nachhaltigkeit des erworbenen Wissens mit Follow-Up-Tests überprüft.

## 6 Erhebungsinstrumente

### 6.1 Individuelle Schülerbeurteilungen

Im Vorfeld der Untersuchung wurden von den Lehrern individuelle Schülerbeurteilungen angefertigt, durch die die Leistungsfähigkeit eingeschätzt und das Ranking erstellt wurde, das die Grundlage der Randomisierung war. Dabei kamen Polaritätenprofile (oder „semantische Differenziale“, vgl. auch Roeder, 2006) mit einer siebenstufigen bipolaren Skala zum Einsatz. Die Items der Fragebögen umfassten leistungsbezogene Eigenschaften (z. B. „meldet sich von alleine – muss aufgefordert werden“, „fachlich kompetent – inkompetent“, „selbstständig – unselbstständig“). Ziel der individuellen Schülerbeurteilungen war es, die Leistungsfähigkeit und das Arbeitsverhalten eines jeden Schülers einzuschätzen und mit den Ergebnissen ein Ranking zu erstellen, mit dem die Klasse in Interventions- und Kontrollgruppe aufgeteilt werden konnte. Dieses recht aufwändige Verfahren wurde gewählt, da eine möglichst exakte Bewertung der Kompetenzen der Schüler angestrebt wur-

de, deshalb wurde ihm auch gegenüber der Verwendung der Zeugniszensuren den Vorzug gegeben. Auch eine Gruppenbildung aufgrund von CFT-Werten erschien weniger geeignet, da dadurch Aspekte der Mitarbeit im Unterricht nicht erfasst werden.

### 6.2 Wissenstest

Der Wissenstest im Multiple-Choice-Format wurde als Pre- und Post-Test eingesetzt und umfasste 34 Fragen zum Thema Seifen (fünf Antwortalternativen, jeweils genau eine richtig); abgefragt wurden sowohl deklaratives Wissen als auch zu einem geringen Teil Transferwissen. Die Dauer des Tests betrug ca. 15 Minuten. Das Maß für die individuelle Leistung im Wissenstest war der Anteil der richtig gelösten Items. Dieser Score kann somit zwischen 0 und 1 liegen.

Als niveaubereinigtes Veränderungsmaß zwischen Pre- und Post-Test wurden im Rahmen der weiteren Auswertung Residuen berechnet. Als Residuum galt die Differenz zwischen dem tatsächlichen Post-Test-Score und dem anhand des Pre-Test-Scores und der Abhängigkeit zwischen Pre- und Post-Test geschätzten Post-Test-Score. Hierbei wurden Pre- und Post-Test zuvor z-standardisiert, so dass die Residuen ihrerseits als z-Werte resultieren.

Die residuenbasierte Auswertung wurde gewählt, da dieses Verfahren das größte Maß an Objektivität und Sicherheit der Auswertung bietet, auch wenn es aufgrund der gelungenen Aufteilung der Schüler in IG und KG (erkennbar an nahezu identischen Ergebnissen im Pre-Wissenstest) nicht zwingend erforderlich gewesen wäre.

---

1 Bedingt durch äußere Einflüsse, z. B. Ferien, Praktika, Auslandsaufenthalte, variierte der Abstand, blieb jedoch für Interventions- und Kontrollgruppe immer gleich.



### 6.3 Einstellungstests

Für die Untersuchung wurde ein Test zur allgemeinen Einstellung zum Fach Chemie (pre/post) und einer zur Einstellung über die Unterrichtseinheit (nur post) entwickelt. Beide Tests umfassten jeweils 24 Items, die ungefähr zur Hälfte positiv und negativ formuliert waren. Bei der Formulierung der Items wurden die von Roeder (2006, 641) formulierten Regeln zur Konstruktion von Einstellungstests berücksichtigt.

Der Einstellungstest zum Fach Chemie misst allgemein die individuelle Bewertung dieses Faches; es wurden Items zu den Aspekten Anwendungsbezug, Inhalt, Sinn, Interesse und persönliches Befinden formuliert. Beispielfhaft seien einige Items aufgeführt: „Ich glaube, dass ich im Fach Chemie etwas fürs Leben lernen kann.“, „Chemie finde ich zu theoretisch.“, „Der Sinn des Faches Chemie ist mir unklar.“, „Wenn ich könnte, würde ich Chemie abwählen.“, „Ich kann im Chemieunterricht etwas leisten“.

Der Einstellungstest zur Unterrichtseinheit wurde derart konzipiert, dass die Schüler durch die Items aufgefordert wurden, eine Einschätzung der letzten drei Stunden, also der Unterrichtseinheit ‚Seifen‘ vorzunehmen. Dieser Test misst allgemein die Unterrichtsqualität mit ihren verschiedenen Aspekten, wie Inhalte, Methodik, persönliches Befinden und Motivation. Da sich der Unterricht in IG und KG nur hinsichtlich der verwendeten Unterrichtsmethode unterscheidet, wird durch diesen Test ermittelt, ob die Schüler die Methode des Gruppenpuzzles positiver als vergleichbaren Frontalunterricht einschätzen. Damit der Fragebogen gleichermaßen in der Gruppenpuzzle- und Frontalunterrichtsgruppe eingesetzt werden konnte, tauchte weder der Begriff „Gruppenpuzzle“ noch „Frontalunterricht“ auf. Auch hier seien beispielhaft einige Items aufgeführt: „Das Thema Seifen fand ich interessant.“ „So wie in den letzten 3 Stunden sollten auch andere Themen unterrichtet werden.“, „Ich fand den Unterricht der letzten 3 Stunden

stressig.“. „In den letzten 3 Stunden war ich unaufmerksam.“.

Jedes Item der Einstellungstests konnte durch Ankreuzen auf einer 5-stufigen Likert-Skala von „sehr zutreffend“ bis „sehr unzutreffend“ beantwortet werden. Für die Auswertung wurden die Items mediandichotomisiert, anschließend erfolgte eine Scoreberechnung, und zwar nach folgenden Schritten (vgl. Roeder, 1995):

1. Umkodieren der Werte der negativ formulierten Items (spiegeln um die Mitte der Skala)
2. Berechnung des exakten Medians der Antwortverteilung bei jedem Einzel-Item (zum Verfahren siehe Bortz, 2005, 37)
3. Umkodierung der individuellen Antworten: 0 = über dem Median, 1 = unter dem Median; 1 = Antwort in positive Richtung, 0 = Antwort in negative Richtung
4. Berechnung des individuellen Scores jedes Schülers als Anteil der Antworten in positive Richtung an allen Antworten. Dieser Score liegt zwischen 0 und 1 und repräsentiert den Anteil an Antworten in Dimensionsrichtung.

Der Einstellungstest zum Fach Chemie wurde als Pre- und Post-Test durchgeführt, hier wurden bei der Mediandichotomisierung diese beiden Tests gepoolt. Für die Einstellung zum Fach Chemie (Pre- und Post-Test) erfolgte analog der Auswertung der Wissenstests eine Residuenberechnung.

Die mediandichotomisierte Auswertung hat gegenüber der Mittelwertberechnung Vorteile. So wird einer schiefgipfligen Verteilung der Antworten bei bestimmten Items (Beispiel: sehr viele Probanden kreuzen „sehr zutreffend“ oder „zutreffend“ an) besser begegnet, da jede Antwort gemessen an ihrem Median als Zustimmung bzw. Ablehnung gewertet wird. Hinzu kommt, dass sich Mittelwertberechnungen streng genommen nur bei kardinal skalierten (metrisch skalierten) Daten rechtfertigen lassen (vgl. Bortz 2005, S. 26, 37). Gerade bei Meinungstests jedoch kann nicht unbedingt Äquidistanz z. B. zwi-

schen „sehr zutreffend“ und „zutreffend“ und zwischen „mittel“ und „unzutreffend“ angenommen werden, dafür ist die Sprache zu stark interpretierbar. Die Mediandichotomisierung nivelliert unterschiedliche Abstände der Skalenstufen durch Aufteilung in „größer“ bzw. „kleiner“ Median (vgl. Bortz 2005, S. 568). Ein weiteres Argument für die mediandichotomisierte Auswertung ist, dass man dabei eine Skala mit einer „Mitte“, also hier eine 5-stufige Skala, einsetzen kann und die Probanden nicht durch eine Skala mit geradzahligem Antwortmöglichkeiten „zwingen“ muss, sich zwischen Zustimmung und Ablehnung zu entscheiden; dieses erfolgt bei der Auswertung durch die Mediandichotomisierung. Probanden, die eine mittlere Einschätzung geben wollen, können dieses somit zum Ausdruck bringen und ziehen nicht die Konsequenz, bei den betreffenden Items gar kein Kreuz zu setzen, was sich negativ auf die Auswertbarkeit des Fragebogens auswirken würde.

## 7 Ergebnisse des Wissenstests

### 7.1 Leistungen in der Interventions- und Kontrollgruppe

In beiden Hauptuntersuchungen zeichnet sich das gleiche Bild ab: Schüler, die nach der Gruppenpuzzle-Methode unterrichtet wurden, erreichen signifikant höhere Lernzuwächse als die entsprechende FU-Gruppe, wobei die Auswertung residuenbasiert erfolgte (vgl. 6.2).

Die Unterschiede sind signifikant (Tab. 2 und Abb. 4). Die Effektgröße  $d$  liegt zwischen 0.46 (mittlerer Effekt) und 0.78 (starker Effekt) und ist damit beachtlich. Cronbachs Alpha des Wissenstests beträgt mindestens 0.73.

Zur besseren Vorstellbarkeit sind in Tab. 2 ergänzend zu den Residuen die Scores der Wissenstests aller Untersuchungen mit aufgeführt. Vor der Intervention lagen beide Gruppen in den Hauptuntersuchungen mit ca. 25 bis 30 % korrekter Antworten ungefähr gleich auf; damit waren die erzielten Ergeb-

nisse etwas höher als die Ratewahrscheinlichkeit von 20 %. Lediglich in der Voruntersuchung werden höhere Werte erreicht. Letzteres liegt daran, dass der Test nach der Voruntersuchung dahingehend überarbeitet wurde, dass das Anspruchsniveau angehoben wurde. Nach der Intervention konnten GP-Schüler in den Hauptuntersuchungen mit 65 bis 73 % zutreffender Antworten im Schnitt acht Prozentpunkte mehr erzielen als die Kontrollgruppe mit 60 bis 62 %. Die Testschwierigkeit ist demnach als mittelmäßig einzustufen. Auch hier liegen die Werte der Voruntersuchung aus den genannten Gründen deutlich darüber.

Insgesamt weisen die Untersuchungen gute Reliabilitäten auf (vgl. Cronbachs  $\alpha$ , Tab. 2). Der etwas niedrigere Wert für Cronbachs Alpha und die geringe Effektgröße  $d$  der HU I kommen durch die Verschärfung der Bedingungen in dieser Untersuchung 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 (in der HU II war dieses nur bei einer Lehrkraft und einem Item erforderlich). Diese Maßnahme macht die Untersuchung zwar konservativer, führt jedoch auch zu einer schlechteren inneren Konsistenz des Wissenstests. Nimmt man die entsprechenden Items aus der Analyse heraus, so ergibt sich für die HU I ein Cronbachs Alpha von 0.76.

Tab. 2: Von den Schülern der Gruppenpuzzle- und Frontalunterricht-Gruppe (IG und KG) erzielte Ergebnisse im Wissenstest. Cronbachs  $\alpha$  wurde für den Post-Test bestimmt. Die Stichprobengröße, Effektgröße und die Signifikanztests beziehen sich jeweils auf die Residuen (diese wurden für jede Untersuchung separat berechnet). Beim t-Test wurde ggf. die Verletzung der Varianzhomogenität berücksichtigt<sup>2</sup>. Soweit nicht anders angegeben beziehen sich die Irrtumswahrscheinlichkeiten auf die einseitige Testdurchführung. MW = Mittelwert. Signifikanzen sind mit entsprechenden Sternchen markiert.

	VU	HU I	HU II
Stichprobenumfang N	FU: 33 GP: 35	FU: 78 GP: 80	FU: 89 GP: 91
MW Scores, Pre-Test	FU: 0.38 GP: 0.41	FU: 0.26 GP: 0.25	FU: 0.30 GP: 0.28
MW Scores, Post-Test	FU: 0.73 GP: 0.82	FU: 0.60 GP: 0.65	FU: 0.62 GP: 0.73
MW Residuen <sub>pre-post</sub>	FU: -0.313 GP: 0.295	FU: -0.234 GP: 0.228	FU: -0.395 GP: 0.386
Standardabweichung S der Residuen	FU: 1.116 GP: 0.765	FU: 1.005 GP: 0.940	FU: 0.937 GP: 0.902
Prüfgröße	$t(56.231) = -2.607$	$t(156) = -2.982$	$t(178) = -5.698$
t-Test Residuen p	0.012* (2-seitig)	0.002**	0.000***
Effektgröße d	0.61	0.46	0.78
Cronbachs $\alpha$	0.79	0.73	0.80

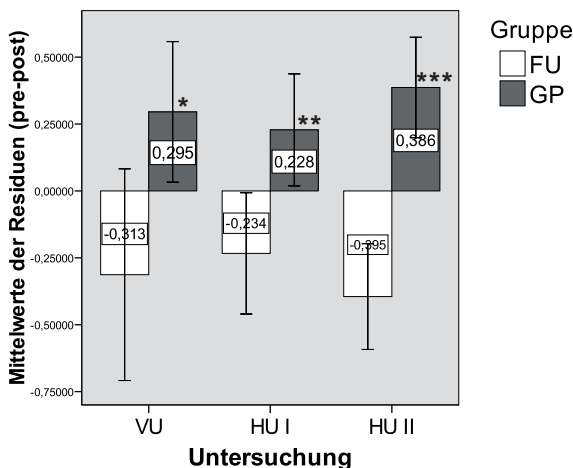


Abb. 4: Von den Schülern der Gruppenpuzzle- (GP) und Frontalunterricht-Gruppe (FU) erzielte Ergebnisse im Wissenstest. Dargestellt sind für die VU, HU I und HU II die Mittelwerte der Residuen (diese wurden für jede Untersuchung separat berechnet) in Abhängigkeit von der verwendeten Methode. Fehlerbalken kennzeichnen das 95 % Konfidenzintervall, schwarze Sternchen signifikante Unterschiede.

2 Für die HU I kann eine Normalverteilung der Residuen mit  $p = 0.066$  nicht uneingeschränkt angenommen werden. Allerdings führen sowohl der t-Test als auch die Varianzanalyse (vgl. Kap. 7.4) bei moderaten Verletzungen der Anwendungsvoraussetzungen zu angemessenen Schlussfolgerungen (z. B. Box, 1954). Wichtig ist, dass die Stichprobengrößen hinreichend groß und annähernd gleich sind. Das ist bei der vorliegenden Untersuchung gegeben. Für die beiden anderen Untersuchungsabschnitte VU und HU II kann eine Normalverteilung angenommen werden.

## 7.2 Vergleich der Ergebnisse mit anderen Studien

Zu abweichenden Ergebnissen kamen Berger & Hänze 2004a im Fach Physik (siehe auch Kapitel 2.2). Im Fokus ihrer Untersuchung waren Unterrichtseinheiten zum Rasterelektronenmikroskop und Mikrowellenofen in der Jahrgangsstufe 12. Hier konnte die Interventionsgruppe keinen Leistungsvorteil gegenüber der Frontalunterrichtsgruppe erzielen, wengleich sie auch nicht signifikant schlechter abschnitt. Die Ergebnisse sind jedoch nur bedingt mit der hier beschriebenen Untersuchung vergleichbar, da dort keine Auswertung über Residuen erfolgte und in Pre- und Post-Wissenstest unterschiedliche Aufgaben gestellt wurden. Hinzu kam außerdem, dass vor dem Gruppenpuzzle eine Frontalunterrichtssequenz stattfand und dass beide Gruppen in ihren Vorleistungen deutlich voneinander abwichen (Berger & Hänze, 2004a).

## 7.3 Abhängigkeit der Lernleistung von der Art der Wissensvermittlung

Unabhängig davon, welches Themengebiet des Gruppenpuzzles man betrachtet, erreichen die Experten (des GP) für das jeweilige Thema höhere Wissenszuwächse als Schüler, denen ihre Mitschüler (im GP) oder der Lehrer (im FU) das Thema vermittelt haben (Abb. 5). Der Vergleich zwischen der Vermittlung von Wissen durch die Mitschüler (im GP) bzw. durch die Lehrer (im FU) fällt nicht eindeutig aus, bei einigen Themen bzw. Untersuchungen ist die Vermittlung durch Mitschüler effektiver, manchmal die durch die Lehrer. In den meisten Fällen sind diese Unterschiede aber zufällig, so dass die Vermittlung durch Mitschüler und Lehrer als gleich effektiv angesehen werden kann. Dort, wo die Unterschiede signifikant sind, ist die Wissensvermittlung durch die Mitschüler effektiver (Tab. 3). In Abb. 6 sind, zusammen für die HU I und HU II (in beiden Untersuchungen wurde der gleiche Test eingesetzt), aus Gründen der Anschaulichkeit ergänzend die Scores dargestellt.

Tab. 3: Vergleich der Wissensvermittlung durch Mitschüler (in der GP-Gruppe) und Lehrer (in der FU-Gruppe). Dargestellt sind die Ergebnisse des U-Tests (hier wurde dieser Test verwendet, weil die Stichprobengrößen unterschiedlich sind), mit dem die mittleren Residuen (diese wurden für jedes Thema und jede Untersuchung separat berechnet) auf Unterschiede geprüft wurden. Angegeben ist die Irrtumswahrscheinlichkeit  $p$ , in Klammern sind die Prüfgrößen des U-Tests angegeben. Signifikanzen sind mit entsprechenden Sternchen markiert.

	Thema	VU	HU I	HU II
Stichprobenumfang N		56	130	150
U-Test p (2-seitig)	1: Aufbau und Löseverhalten	0.220 (U = 306.00)	0.831 (U = 2021.50)	0.018 (U = 2095.50)
	2: Oberflächenspannung	0.559 (U = 360.00)	0.525 (U = 1894.50)	0.118 (U = 2306.50)
	3: Nachteile von Seife	0.275 (U = 314.00)	0.929 (U = 2125.50)	0.000*** (U = 1474.50)

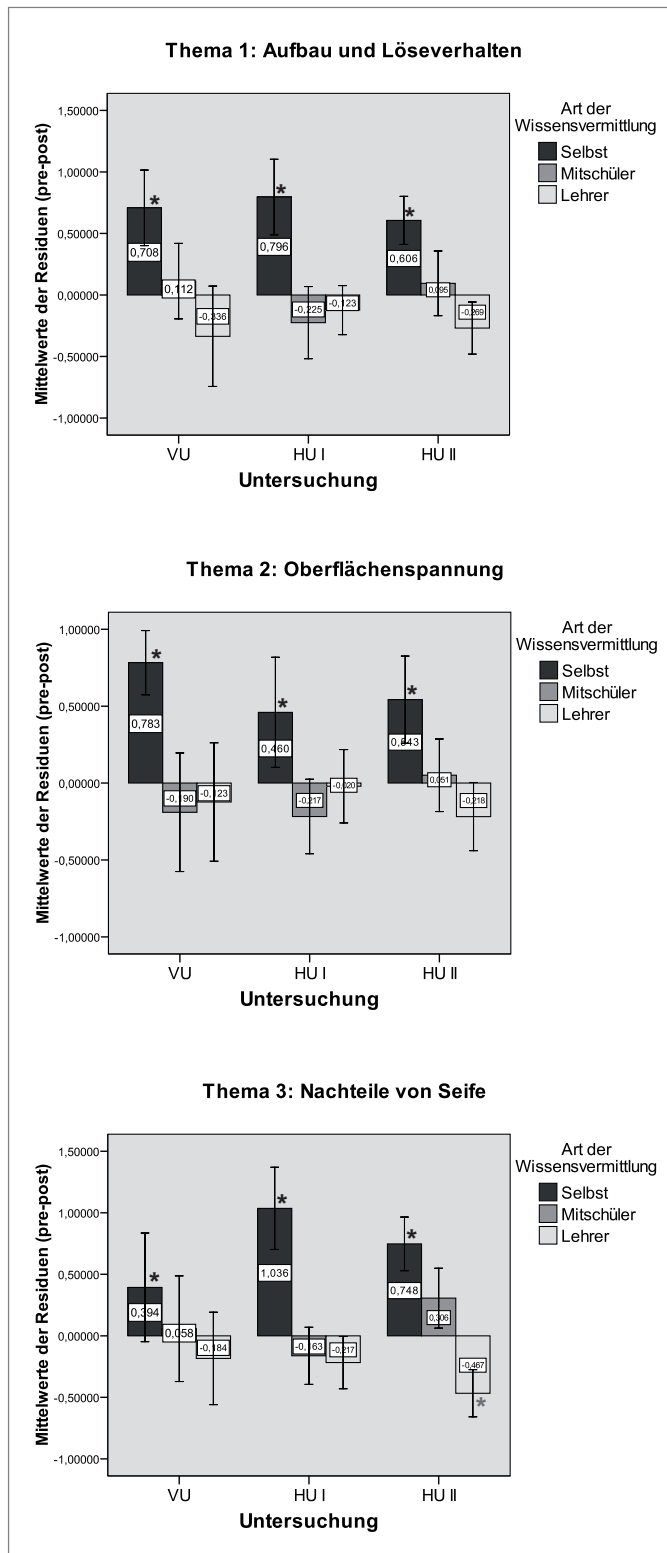


Abb. 5: Vergleich der drei Vermittlungsarten: „selbst“ beigebracht (als Experte in der GP-Gruppe), vermittelt durch die „Mitschüler“ (in der GP-Gruppe) bzw. durch den „Lehrer“ (in der FU-Gruppe). Dargestellt sind für die VU, HU I und HU II die Ergebnisse des Wissenstests, also die Mittelwerte der Residuen (diese wurden für jedes Thema und jede Untersuchung separat berechnet) in Abhängigkeit von der Vermittlungsart. Fehlerbalken kennzeichnen das 95 % Konfidenzintervall, schwarze Sternchen signifikante Unterschiede zwischen Selbst- und Fremdvermittlung (durch Mitschüler oder Lehrer), das graue Sternchen signifikante Unterschiede zwischen Lehrern und Mitschülern. Hier wurden U-Tests durchgeführt, weil die Stichprobengrößen unterschiedlich groß sind: Jedes Thema wird nur von einem Drittel der GP-Schüler „selbst“ beigebracht, bei zwei Dritteln erfolgt die Vermittlung durch die Mitschüler, im FU vermittelt ausschließlich der Lehrer.

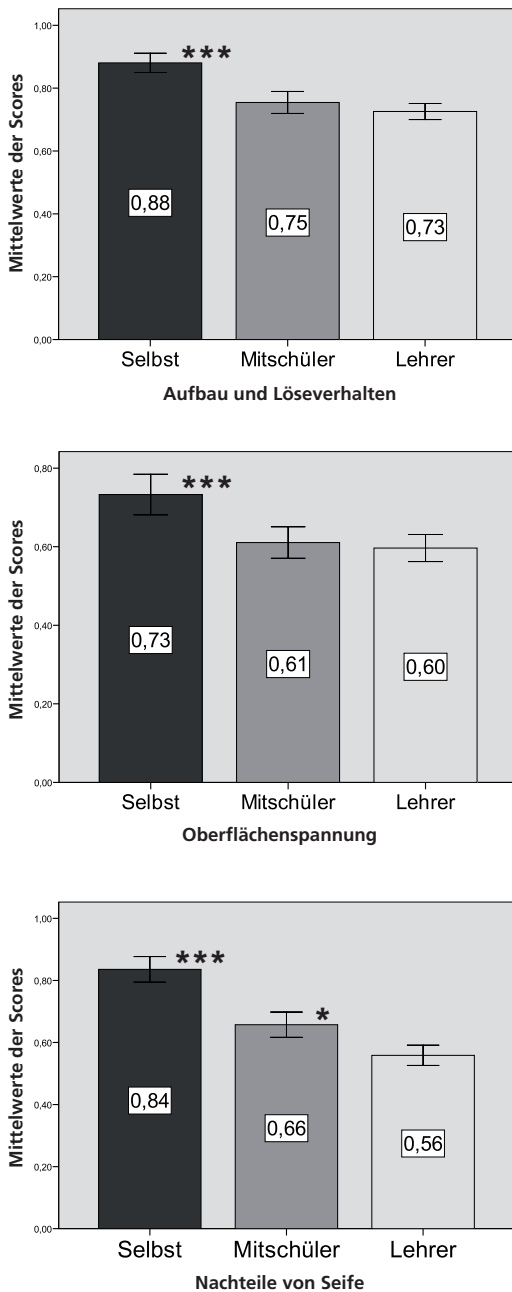


Abb. 6: Vergleich der drei Vermittlungsarten: „selbst“ beigebracht (als Experte in der GP-Gruppe), vermittelt durch die „Mitschüler“ (in der GP-Gruppe) bzw. durch den „Lehrer“ (in der FU-Gruppe). Dargestellt sind für die HU I und HU II die im Wissenstest im Mittel erreichten Scores in Abhängigkeit von der Vermittlungsart, getrennt für die drei Themen.

Die Analyse der Vermittlungsarten zeigt die hohe Überlegenheit von selbst beigebrachtem Wissen als Experte gegenüber rezeptiv erworbenem Wissen. Diese Erkenntnis deckt sich mit Untersuchungen von Berger & Hänze (2004a, 2005), auch wenn diese aus o. g. Gründen nur bedingt mit der vorliegenden Studie vergleichbar sind. Auch Borsch et al. (2002) 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.

Zwei Faktoren dürften in der hier beschriebenen Untersuchung für das bessere Abschneiden der Experten und damit des gesamten Gruppenpuzzles verantwortlich sein: Der erste Faktor ist der höhere Anteil an Selbsttätigkeit durch das eigenständige Aneignen von Wissen. 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 Meta-Studie von Johnson & Johnson (1994) hin. In dieser wurden 59 Einzelstudien auf die Effektivität von kooperativem Lernen und individuellem Lernen untersucht. Slavin (1993) nennt weiterhin die Bedeutung von Gruppenzielen und die individuelle Verantwortlichkeit. Der zweite Faktor ist die Vermittlung an die Mitschüler durch die Experten (Green & Green, 2005, 12). Das Gruppenpuzzle beinhaltet diese als Lernen durch Lehren bekannte Variante kooperativen Lernens, deren Vorteil sowohl Renkl (2000) als auch Huber, Konrad & Wahl (2001) bestätigen. Gleichzeitig erfolgt durch die Vermittlung an die Mitschüler auch eine Wiederholung für die Experten.

Renkl (1997) erklärt Effekte des Lernens 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).

#### 7.4 Wechselwirkungsanalysen

Um zu überprüfen, ob es Zusammenhänge zwischen der Wirksamkeit der Unterrichtsmethode (geschlussfolgert aus den Ergebnissen des Wissenstests) und den Chemiezensuren der Schüler gibt, wurde eine Varianzanalyse durchgeführt. Hinsichtlich der Chemiezensuren wurden die Schüler in zwei Gruppen eingeteilt (Mediansplit):

Eine Hälfte bildeten die Schüler mit der besseren Chemiezensur, die andere die mit der schlechteren Chemiezensur. Hieraus resul-

tierte, dass alle Schüler mit den Zensuren 1 und 2 der besseren Hälfte, alle Schüler mit den Zensuren 4, 5 und 6 der schlechteren Hälfte zuzuordnen waren. Die Schüler mit der Zensur 3 wurden nach einem Zufallsprinzip so auf die bessere und schlechtere Hälfte aufgeteilt, dass beide Gruppen gleich groß waren.

Betrachtet man die 2-Weg-Wechselwirkungen von Methode (Gruppenpuzzle bzw. Frontalunterricht) und Chemiezensur (bessere bzw. schlechtere Chemiezensur), so sind keine Wechselwirkungen nachweisbar (Tab. 4). Der höhere Lernzuwachs der GP-Schüler im Vergleich zu den FU-Schülern scheint demnach unabhängig von ihrer Chemiezensur zu sein; Schüler aller Zensurenstufen profitieren also gleichermaßen vom Gruppenpuzzle. Auch zwischen Geschlecht und Methode besteht keine Wechselwirkung (Tab. 4).

Tab. 4: Untersuchung auf Wechselwirkungen zwischen der Chemiezensur, dem Geschlecht und der Unterrichtsmethode (GP oder FU) hinsichtlich der Leistungen der Schüler im Wissenstest (residuenbasierte Auswertung): Dargestellt sind die Ergebnisse der zweifaktoriellen Varianzanalyse (2-Weg-Wechselwirkungen). Angegeben sind die Prüfgrößen sowie die Irrtumswahrscheinlichkeiten  $p$  und Effektgrößen  $\eta^2$ .

		VU	HU I	HU II
Stichprobenumfang N		67	158	180
Haupteffekte	Methode	F(1,59) = 6.361 $p = 0.014^*$ $\eta^2 = 0.097$	F(1,150) = 8.899 $p = 0.002^{**}$ $\eta^2 = 0.056$	F(1,172) = 33.424 $p = 0.000^{***}$ $\eta^2 = 0.163$
	Chemienote (besser – schlechter)	F(1,59) = 2.533 $p = 0.117$ $\eta^2 = 0.041$	F(1,150) = 0.993 $p = 0.321$ $\eta^2 = 0.007$	F(1,172) = 6.349 $p = 0.013^*$ $\eta^2 = 0.036$
	Geschlecht	F(1,59) = 0.015 $p = 0.904$ $\eta^2 = 0.000$	F(1,150) = 0.419 $p = 0.519$ $\eta^2 = 0.003$	F(1,172) = 0.155 $p = 0.694$ $\eta^2 = 0.001$
2-Weg-Wechselwirkungen	Methode x Chemienote (besser – schlechter)	F(1,59) = 1.591 $p = 0.212$ $\eta^2 = 0.026$	F(1,150) = 0.951 $p = 0.331$ $\eta^2 = 0.006$	F(1,172) = 1.288 $p = 0.258$ $\eta^2 = 0.007$
	Methode x Geschlecht	F(1,59) = 2.222 $p = 0.639$ $\eta^2 = 0.004$	F(1,150) = 0.076 $p = 0.783$ $\eta^2 = 0.001$	F(1,172) = 1.406 $p = 0.237$ $\eta^2 = 0.008$

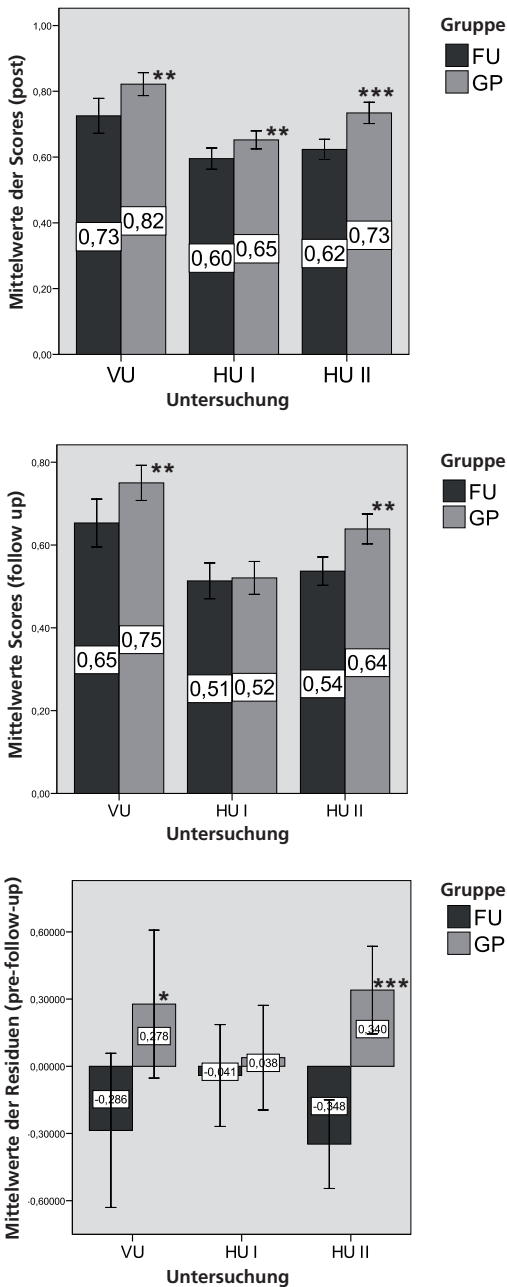


Abb. 7: Ergebnisse der Follow-Up-Wissenstests im Vergleich zum Post-Wissenstest. Dargestellt sind für die VU, HU I und HU II die Scores (obere beiden Grafiken) sowie Residuen (diese wurden für jede Untersuchung separat berechnet, untere Grafik), in Abhängigkeit von der verwendeten Methode (GP bzw. FU). Fehlerbalken kennzeichnen das 95 % Konfidenzintervall, Sternchen signifikante Unterschiede zwischen den Residuen (t-Test).

Der Einfluss des Geschlechts wurde auch von Lazarowitz (1991), bezogen auf das Fach Biologie, untersucht. Er konnte zwar leichte Vorteile für weibliche Schüler diagnostizieren, jedoch waren die Unterschiede nicht signifikant. Eine umfangreiche andere Untersuchung zum Thema Photosynthese in neunzehn 10. Klassen ( $n = 708$ ), zeigte die Unabhängigkeit des Gruppenpuzzles vom Geschlecht (vgl. Lazarowitz & Karsenty, 1990).

## 7.5 Nachhaltigkeit des erworbenen Wissens

Wie schon beim Post-Test schneidet die Interventionsgruppe beim Follow-up-Test durchweg besser ab als die Kontrollgruppe (Abb. 7). Dabei liegen die Post-Test-Scores der GP-Gruppe in den Hauptuntersuchungen bei 0.65 bzw. 0.73, die der FU-Gruppe bei 0.60 bzw. 0.62. Die Unterschiede zwischen der GP- und der FU-Gruppe im Post-Test betragen also durchschnittlich 8 %-Punkte. Die Follow-Up-Test-Scores der GP-Gruppe liegen in den Hauptuntersuchungen bei 0.52 und 0.64, die der FU-Gruppe bei 0.51 und 0.54. Die Unterschiede zwischen GP- und FU-Gruppe im Follow-Up-Test betragen damit etwa 5,5 %-Punkte und sind damit geringer als beim Post-Test. Wiederum liegen die Werte der Voruntersuchung etwas höher, da die erste (leichtere) Testversion verwendet wurde.

Betrachtet man die Residuen des Follow-up-Tests, so ergibt sich das tendenziell gleiche Bild: Die kooperative Lernform liegt vor der FU-Gruppe. Im Vergleich zu den Residuen des Post-Tests sind die Unterschiede generell weniger stark ausgeprägt (Abb. 4).

Sowohl bei den Scores als auch bei den Residuen erreichen die Schüler der Interventionsgruppe durchweg bessere Werte, der Vorsprung bleibt also erhalten. Damit konnten Ergebnisse von Borsch et al. (2002) bestätigt werden. Für die Erscheinung, dass es



in der ersten Hauptuntersuchung bei den Follow-up-Tests keine Unterschiede mehr zwischen den beiden Stichproben gibt, sind verschiedene Erklärungen denkbar:

Einerseits mag es daran liegen, dass ein Lehrer die in der Frontalunterricht-Gruppe nicht bearbeiteten Inhalte 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 auch das Wissen, so dass auch hier Unterschiede verschwimmen. Auch mag zu Unwägbarkeiten geführt haben, dass einige Lehrer den Follow-up-Test angekündigt haben, mit der Folge, dass deren Schüler mit evtl. eigenen Methoden dafür geübt und somit das Ergebnis beeinflusst haben.

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 die Follow-up-Tests jedoch nur einen kleinen Teil der Untersuchung ausmachen, soll ihre Aussagekraft nicht überbewertet werden. Insofern soll das Fazit vorsichtig formuliert werden: Auch nach 4-8 Wochen scheinen Gruppenpuzzle-Schüler einen Wissensvorteil gegenüber frontal unterrichteten Schülern zu haben.

Auch im Follow-up-Test erreichen die Experten für das jeweilige Thema die höchsten Wissenszuwächse (Auswertung über Pre-/Follow-up-Residuen). Die Unterschiede zwischen den Experten des Gruppenpuzzles und den Schülern, denen das Wissen im Gruppenpuzzle durch Mitschüler vermittelt wurde, sind jedoch hier nicht bei allen Themen und in allen Untersuchungsabschnitten signifikant.

## 8 Ergebnisse der Einstellungstests

### 8.1 Einstellungstest zum Fach Chemie

Die Ergebnisse des Einstellungstests zum Fach Chemie fallen uneinheitlich aus: Einmal beurteilt die Kontrollgruppe das Fach etwas besser (HU I + HU II, vgl. Residuen, Tab. 6), ein anderes Mal die Interventionsgruppe (VU, vgl. Residuen, Tab. 6). Der Unterschied ist nur in der HU II signifikant.

Zusammenfassend lässt sich festhalten, dass die Methode keinen eindeutigen Einfluss auf die Einstellung zum Fach Chemie hat. Dieses Ergebnis verwundert nicht, da die Interventionsmaßnahme lediglich über drei Stunden erfolgte und ein Effekt ggf. erst nach längerer Dauer nachzuweisen wäre. Offensichtlich scheinen bereits vorhandene und evtl. manifestierte Schülereinstellungen durch kurze Interventionsmaßnahmen nicht beeinflusst zu werden.

### 8.2 Einstellungstest zur Unterrichtseinheit

Die Schüler, die nach der Methode des Gruppenpuzzles unterrichtet wurden, schätzen die Unterrichtseinheit, also die letzten drei Stunden, durchweg positiver ein (d. h. höherer Score) als die Schüler, die Frontalunterricht erhielten (Abb. 8). Die Unterschiede sind bis auf die HU II signifikant (Tab. 6). Im Prinzip sprechen die Ergebnisse dafür, dass die Schüler das GP positiver einschätzen als den Frontalunterricht. Dies konnte in zwei unabhängigen Teil-Untersuchungen

Tab. 5: Ergebnisse des Einstellungstests zum Fach Chemie. Cronbachs  $\alpha$  wurde für den Post-Test auf Basis der umkodierten, noch nicht mediandichotomisierten Rohdaten bestimmt; die Stichprobengröße, die Signifikanztests und die Effektgröße beziehen sich jeweils auf die Residuen. Beim t-Test wurde ggf. die Verletzung der Varianzhomogenität berücksichtigt, aufgrund der Daten konnte eine Normalverteilung angenommen werden. Die Irrtumswahrscheinlichkeiten beziehen sich auf die zweiseitige Testdurchführung. Grundsätzlich gilt: je höher der Score, desto positiver die Einschätzung. MW = Mittelwert.

	VU	HU I	HU II
Stichprobenumfang N	FU: 33 GP: 35	FU: 78 GP: 80	FU: 89 GP: 92
MW Scores, Pre-Test	FU: 0.46 GP: 0.50	FU: 0.50 GP: 0.57	FU: 0.46 GP: 0.44
MW Scores, Post-Test	FU: 0.52 GP: 0.55	FU: 0.56 GP: 0.59	FU: 0.49 GP: 0.42
MW Residuen <sub>pre-post</sub>	FU: -0.013 GP: 0.013	FU: 0.050 GP: -0.048	FU: 0.180 GP: -0.174
Standardabweichung S der Residuen	FU: 0.790 GP: 1.163	FU: 1.085 GP: 0.907	FU: 1.033 GP: 0.934
Prüfgröße	$t(60.130) = -0.108$	$t(156) = 0.616$	$t(179) = 2.423$
t-Test Residuen p	0.914	0.539	0.016*
Effektgröße d	0.03	0.10	0.36
Cronbachs $\alpha$	0.91	0.93	0.92

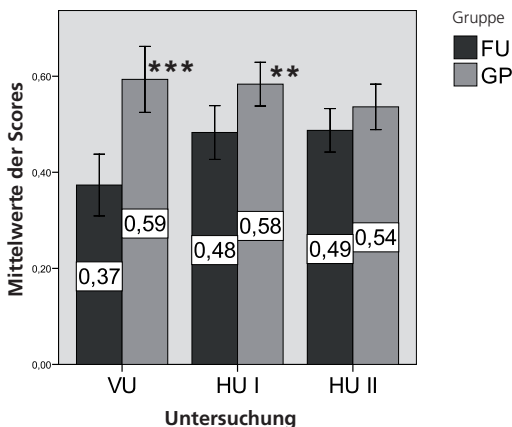


Abb. 8: Ergebnisse des Einstellungstests zur Unterrichtseinheit. Dargestellt sind die mittleren Scores in Abhängigkeit von der verwendeten Methode (GP bzw. FU). Je höher der Score, desto positiver die Einschätzung. Sternchen kennzeichnen signifikante Unterschiede.

Gruppe  
 ■ FU  
 ■ GP

(VU und HU I) bestätigt werden. Die Hauptuntersuchung II jedoch weist knapp keine signifikanten Unterschiede mehr auf, wenngleich das GP auch hier positiver bewertet wurde. Auch Berger & Hänze (2004b) 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).

Die insgesamt deutlich bessere Beurteilung der Gruppenpuzzle-Methode ist insofern wichtig, als dass Zufriedenheit und Wohlbefinden im Unterricht eine entscheidende Rolle für den Lernerfolg spielen (vgl. z. B. Bolte, 2004). Nur eine von den Schülern akzeptierte und gemochte Methode wird

Tab. 6: Ergebnisse des Einstellungstests zur Unterrichtseinheit (nur Post-Test). Cronbachs  $\alpha$  wurde auf Basis der umkodierten, noch nicht mediandichotomisierten Rohdaten bestimmt; die Stichprobengröße, die Signifikanztests und die Effektgröße beziehen sich jeweils auf Scores. Beim t-Test wurde ggf. die Verletzung der Varianzhomogenität berücksichtigt, aufgrund der Daten konnte eine Normalverteilung angenommen werden. Die Irrtumswahrscheinlichkeiten beziehen sich auf die einseitige Testdurchführung. Grundsätzlich gilt: je höher der Score, desto positiver die Einschätzung. MW = Mittelwert.

	VU	HU I	HU II
Stichprobenumfang N	FU: 35 GP: 36	FU: 78 GP: 81	FU: 89 GP: 92
MW Scores, Post-Test	FU: 0.37 GP: 0.59	FU: 0.48 GP: 0.58	FU: 0.49 GP: 0.54
Standardabweichung S der Scores	FU: 0.0.373 GP: 0.594	FU: 0.248 GP: 0.206	FU: 0.215 GP: 0.229
Prüfgröße	t(69) = - 4.744	t(157) = - 2.785	t(179) = - 1.480
t-Test Scores p	0.000**	0.003**	0.070
Effektgröße d	0.98	0.43	0.22
Cronbachs $\alpha$	0.91	0.91	0.90

auch erfolgreich eingesetzt werden können. Auch Detailauswertungen 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, Baumert, Julius-McElvany & Peschar, 2004, 3).

## 9 Zusammenfassung und Schlussbemerkungen

Durch diese Studie konnten folgende Hypothesen bestätigt werden:

Hypothese 1: *Schüler, die nach der Gruppenpuzzle-Methode 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 Befunde internationaler Studien auf den deutschen Chemieunterricht am Gymnasium übertragen.

Hypothese 2: *Schüler haben eine positivere Einstellung zu Unterricht nach der Methode des Gruppenpuzzles als zu inhaltlich vergleichbarem Frontalunterricht.*

Diese Hypothese konnte in zwei von drei Teiluntersuchungen bestätigt werden.

Nebenhypothese 1: *Das Gruppenpuzzle eignet sich hinsichtlich der Lernleistung für Schüler mit besseren und schlechteren Chemiezensuren gleichermaßen.*

Diese Hypothese wird in der vorliegenden Untersuchung durch die Daten unterstützt. Es konnten keinerlei Abhängigkeiten von Methode und besseren bzw. schlechteren Chemie-Zeugnisnoten festgestellt werden.

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.

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

Diese Hypothese konnte nicht bestätigt werden. Das Ergebnis wird erklärbar, wenn man die äußerst kurze Interventionszeit (drei Unterrichtsstunden) bedenkt. Um derartige Hypothesen zu überprüfen, müssten längerfristige Untersuchungen durchgeführt werden.

Als Nebenergebnis konnte festgestellt werden, dass die Methode des Gruppenpuzzles auf beide Geschlechter gleichermaßen positiv wirkt.

Bisher galten kooperative Unterrichtsmethoden trotz diverser positiver Erfahrungsberichte bei vielen Lehrern als unproduktiv und eher als Abwechslung denn als eine effektive Unterrichtsmethode, die sie aus Überzeugung im Unterricht einsetzen würden. Mit dieser Studie wurde für das Unterrichtsfach Chemie der Nachweis über die Lernwirksamkeit des Gruppenpuzzles, als einer spezifischen kooperativen Unterrichtsform, erbracht. Hierbei stand der Kompetenzbereich Fachwissen im Mittelpunkt. Andere Ziele des Einsatzes kooperativer Unterrichtsformen wurden in der durchgeführten Studie aus untersuchungsmethodischen Gründen ausgeklammert.

Untersuchungsdesign und Auswertung sind durchgängig konsequent konservativ angelegt, d. h. wo immer möglich, wurden die Rahmenbedingungen so gewählt, dass keinerlei Vorteile für die Gruppenpuzzle-Gruppe resultieren sollten. Somit liegen nun sehr fundierte und belastbare Erkenntnisse über die Effektivität des Gruppenpuzzles vor, die in den bisherigen Studien im deutschen Chemieunterricht, welche mehr den Charakter von Erfahrungsberichten hatten, nicht erzielt werden konnten.

Der Vorsprung der Schüler, die das Gruppenpuzzle bearbeitet haben, ist hauptsächlich darauf zurückzuführen, dass sie bei den Fragen des Wissenstests bessere Ergebnisse erzielen, die sich auf das Thema beziehen, für das sie Experten waren und das sie einerseits selbst erarbeitet bzw. sich selbst beige-

bracht haben und das sie andererseits anderen Schülern vermittelt haben. Hier erzielen sie signifikant höhere Lernzuwächse als die Schüler, denen die Lerninhalte entweder von anderen Experten oder im Frontalunterricht vom Lehrer vermittelt wurden. Besonders fruchtbar für das Lernen sollte sich demzufolge die Selbstständigkeit der Erarbeitung gekoppelt mit einer Verantwortung für den Lernerfolg – sowohl den eigenen als auch den der Mitschüler, denen das Fachwissen vermittelt wird – auswirken. In diesem Zusammenhang ist auch zu beachten, dass in der Expertenphase ein Austausch mit anderen Experten über das eigene Thema erfolgt und dass in der Vermittlungsphase auch die Fragen der Mitschüler, denen das Thema vermittelt wird, zu einer Wiederholung und einem tieferen Verständnis des eigenen Themas führen.

Der Vorsprung der Gruppenpuzzle-Schüler bleibt auch langfristig erhalten. Berücksichtigt man außerdem, dass das Gruppenpuzzle für unterschiedlich leistungsstarke Schüler sowie für Jungen und Mädchen gleichermaßen geeignet ist, so kann die Schlussfolgerung gezogen werden, dass im Fach Chemie das Gruppenpuzzle als Alternative zum Frontalunterricht durchaus empfohlen werden kann.

In der durchgeführten Untersuchung waren die Lerngruppen aufgrund des Untersuchungsdesigns (Teilung der Klassen) sowohl beim Gruppenpuzzle als auch im Frontalunterricht meistens kleiner als im realen Unterricht, jedoch dürfte sich die geringere Größe der Klasse eher bei der Frontalunterricht-Gruppe positiv auswirken. Allerdings scheint die Gruppengröße wenig Einfluss auf die Ergebnisse zu besitzen, da zum einen in dieser Studie teilweise zwei Klassen zusammengelegt und dann gemeinsam randomisiert in Interventions- und Kontrollgruppe aufgeteilt wurden, so dass übliche Lerngruppengrößen resultierten.

Die Untersuchung liefert darüber hinaus 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 hoch, so dass es sich vorwiegend für kurze Unterrichtseinheiten eignet.

Der große Gewinn dieses Designs liegt darin, sehr vergleichbare Gruppen zu erhalten, die von derselben Lehrkraft unterrichtet werden. Als Nachteil muss man lediglich in Kauf nehmen, dass die unterrichteten Gruppen nur eine halbe Klassenstärke haben, sofern nicht zwei Klassen zusammengelegt werden können. Allerdings wird sich dieses in den meisten Studien in der Interventions- und Kontrollgruppe gleichermaßen auswirken.

Außerdem bietet dieses Design den Vorteil, dass die Untersuchung in der Kontrollgruppe komplett abgeschlossen ist, bevor sie in der Interventionsgruppe beginnt, so dass eine „Verschmierung“ der Effekte verringert wird. So kann z. B. bei einem Parallelklassendesign nicht ausgeschlossen werden, dass Schüler der Interventionsgruppe von der Intervention berichten und die Schüler der Kontrollgruppe damit in deren Einschätzung der Unterrichtseinheit beeinflussen. Diese könnten aufgrund möglicherweise sehr positiver Berichte über die Intervention ihren eigenen Unterricht negativer beurteilen als sie es ohne Kenntnis der Intervention getan hätten, was das Ergebnis des Einstellungstests beeinflussen würde.

## Literatur

- Artelt, C., Baumert, J., Julius-McElvany N. & Peschar, J. (2004). *Das Lernen lernen: Voraussetzungen für lebensbegleitendes Lernen. Ergebnisse von PISA 2000*. Paris: OECD.
- Ballstaedt, S.-P. (1993). Schriftliche Unterlagen für Schülerhand. *Pädagogik*, 93(5), 24-27.
- Berger, R. & Hänze, M. (2004a). Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II – Einfluss auf Motivation, Lernen und Leistung. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 205-219.
- Berger, R. & Hänze, M. (2004b). Förderung intrinsischer Motivation durch das Gruppenpuzzle. *Naturwissenschaften im Unterricht – Physik*, 15(84), 33.
- Berger, R. & Hänze, M. (2005). Gruppenpuzzle im Physikunterricht der Sekundarstufe II – welche Unterschiede gibt es zum Lernzirkel? In A. Pitton (Hrsg.), *Relevanz fachdidaktischer Forschungsergebnisse für die Lehrerbildung*, (152-154). Münster: LIT.
- Bolte, C. (2004). Motivation und Lernerfolg im Chemieunterricht der Sekundarstufe I. *PdN-Ch – Chemie in der Schule*, 53(2), 2-5.
- 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.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- 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.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. In *Review of Educational Research*, 64, 1-35.
- Box, G. E. P (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Math. Statistics*, 25, 484-498.
- Eilks, I. (2003a). Kooperatives Lernen im Chemieunterricht (Teil 1). *Der mathematische und naturwissenschaftliche Unterricht*, 56(1), 51-55.
- Eilks, I. (2003b). Kooperatives Lernen im Chemieunterricht (Teil 2). *Der mathematische und naturwissenschaftliche Unterricht*, 56(2), 111-115.

- Eilks, I. & Ralle, B. (2002). Partizipative Fachdidaktische Aktionsforschung – Ein Modell für eine begründete und praxisnahe curriculare Entwicklungsforschung in der Chemiedidaktik. *Chemie konkret*, 1(9), 13-18.
- Eilks, I., Witteck, T., Rumann, S. & Sumfleth, E. (2005). Kooperatives Lernen. *Naturwissenschaften im Unterricht Chemie*, 16(88/89), 6-11.
- Eppler, R., Winter, M. & Huber, G. L. (1985). Kooperatives Lernen als "Gruppenpuzzle". Tuebingen.
- Frey-Eiling, A. & Frey, K. (1999). Das Gruppenpuzzle. In J. Wiechmann (Hrsg.), *Zwölf Unterrichtsmethoden. Vielfalt für die Praxis* (50-57). Weinheim: Beltz.
- Fürstenau, B. (2002). Das kooperative Lehr-Lern-Arrangement Gruppenpuzzle. *Berufsbildung*, 56(76), 41-42.
- Green, N. & Green, K. (2005). Kooperatives Lernen im Klassenraum und im Kollegium. *Das Trainingsbuch*. Seelze-Velber: Kallmeyer.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? *Pädagogik*, 58(2), 42-45.
- Huber, A., Konrad, K. & Wahl, D. (2001). Lernen durch wechselseitiges Lehren. *Pädagogisches Handeln* (5), 33-46.
- Huber, G. L. (1985). *Pädagogisch-psychologische Grundlagen für das Lernen in Gruppen, Studienbrief 1: Lernen in Schülergruppen*. Tübingen: Deutsches Institut für Fernstudien.
- Johnson, D. W. & Johnson, R.T. (1989). *Cooperation and competition: Theory and research*. Interaction Book Company. MN: Edina.
- Johnson, D. W. & Johnson, R. T. (1994). An Overview of Cooperative Learning. In J. S. Thousand et al. (Eds.) *Creativity and Collaborative Learning: A Practical Guide to Empowering Students and Teachers*, (31-44). Baltimore, MD: Brooks Publishing Co.
- Johnson, D. W. & Johnson, R.T. (1999). *Learning Together and Alone. Cooperative, Competitive and Individualistic Learning*. Boston: Allyn & Bacon.
- Johnson, D. W. & Johnson, R.T. *The Cooperative Learning Center at The University of Minnesota*. <http://www.co-operation.org/> (Abrufdatum: 28.05.2009).
- 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> (Abrufdatum: 28.05.2009).
- 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(2), 74-84.
- Lazarowitz, R. (1991). Learning biology cooperatively. *Cooperative Learning*, 11(3), 19-21.
- Lazarowitz, R., Hertz-Lazarowitz, R. & Baird, H. (1994). Learning science in a cooperative setting: Academic achievement and affective outcomes. *Journal of Research in Science Teaching*, 31(10), 1121-1131.
- Lazarowitz, R. & Hertz-Lazarowitz, R. (1998). Cooperative Learning in the Science Curriculum. In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education* (449-469). Netherlands: Kluwer Academic Publishers.
- 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* (123-149). New York: Praeger.
- Leerhoff, G. & Eilks, I. (2002). Schülerinnen und Schüler erarbeiten sich den Atombau. *Praxis Schule* 5 – 10, 2002(5), 48-54.
- 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?* (115-125). Berlin: Logos.
- Mifflin, H.. *Project-based Learning Space*. <http://college.cengage.com/education/pbl/tc/coop.html>. (Abrufdatum: 28.05.2009).
- Renkl, A. (1997). *Lernen durch Lehren. Zentrale Wirkmechanismen beim kooperativen Lernen*. Wiesbaden: Deutscher Universitätsverlag.
- Renkl, A. (2000). Automatisierung allein reicht nicht aus: Üben aus kognitionspsychologischer Perspektive. In R. Meier, U. Rampilon, U. Sandfuchs & L. Stäudel (Hrsg.), *Üben und Wiederholen* (Jahresheft 2000) (16-19). Seelze: Friedrich Verlag.
- Roeder, B. (1995). Einführung in die empirischen Forschungsmethoden. In K. Bräuer (Hrsg.), *Psychologische Potentiale für eine interdisziplinäre Lehrerbildung: Motivation – Kognition – Entwicklung* (198-210). Essen: Die Blaue Eule.
- Roeder, B. (2006). Schülerbefragung. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Auflage) (637-642). Weinheim: Beltz PVU.
- Rumann, Stefan (2005). *Kooperatives Arbeiten im Chemieunterricht. Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik*. Studien zum Physik- und Chemielernen. (45). Berlin: Logos-Verlag.

- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005), Hrsg. *Beschlüsse der Kultusministerkonferenz, Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. München: Luchterhand 2005.
- Sharan, S. (1980). Cooperative learning in teams: Recent methods and effects on achievement, attitudes and ethnic relations. *Review of Educational Research*, 50, 241-272.
- Slavin, R. E. (1977). Classroom reward structure: An analytical and practical review. *Review of Educational Research*, 47, 633-650.
- 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) (151-170). Baltmannsweiler: Schneider-Verlag, Hohengehren.
- Slavin, R. E. (1995). *Cooperative learning: Theory, research and practice. 2nd ed.* Boston: Allyn and Bacon.
- 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* (284-302). Münster: Waxmann.
- Tepner, M. & Melle, I. (2005). *Saubere Sache! Ein Gruppenpuzzle zum Thema Seifen*. Stuttgart: RAABits Chemie Sekundarstufe I. Raabe.
- 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* (214-216). Münster: LIT.
- Tepner, M., Melle, I. & Roeder, B. (2005b). Gruppenpuzzle und Frontalunterricht im Vergleich. *Naturwissenschaften im Unterricht Chemie* 16(88/89), 82-85.
- Walpuski, M. & Sumfleth, E. (2007). Strukturierungshilfen und Feedback zur Unterstützung experimenteller Kleingruppenarbeit im Chemieunterricht. *ZfDN* 13, 181-198.

## Kontakt

Prof. Dr. Insa Melle  
 TU Dortmund  
 Lehrstuhl für Didaktik der Chemie II  
 Otto-Hahn-Str. 6  
 44227 Dortmund  
[insa.melle@tu-dortmund.de](mailto:insa.melle@tu-dortmund.de)

## Autoreninformation

Dr. Markus Tepner hat an der RWTH Aachen Chemie und Biologie für das Lehramt an Gymnasien studiert und anschließend im Arbeitskreis von Prof. Dr. Insa Melle an der TU Dortmund seine Dissertation angefertigt. Er hat sein Referendariat am Studienseminar Kleve II absolviert und ist jetzt als Studienrat am Julius-Stursberg-Gymnasium in Neukirchen-Vluyn tätig.

Prof. Dr. Burkhard Roeder hatte bis zu seiner Emeritierung den Lehrstuhl für Methodenlehre der Psychologie an der TU Dortmund im Institut für Psychologie inne. Seine Forschungsschwerpunkte sind Lehr-Lernforschung, Geschlechtsspezifität und Einstellungsmessung.

Prof. Dr. Insa Melle ist Inhaberin des Lehrstuhls für Didaktik der Chemie II an der TU Dortmund. Ihre Forschungsschwerpunkte sind die empirische Lehr-Lernforschung und die Entwicklung von Lernmaterialien.