

Knobloch, Rebecca; Sumfleth, Elke; Walpuski, Maik

Förderung der Qualität fachinhaltlicher Schüleräußerungen in experimenteller Kleingruppenarbeit im Chemieunterricht

Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 19 (2013), S. 347-373



Quellenangabe/ Reference:

Knobloch, Rebecca; Sumfleth, Elke; Walpuski, Maik: Förderung der Qualität fachinhaltlicher Schüleräußerungen in experimenteller Kleingruppenarbeit im Chemieunterricht - In: Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 19 (2013), S. 347-373 - URN: urn:nbn:de:0111-pedocs-317105 - DOI: 10.25656/01:31710

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-317105>

<https://doi.org/10.25656/01:31710>

in Kooperation mit / in cooperation with:



IPN

Leibniz-Institut für die Pädagogik der
Naturwissenschaften und Mathematik

<https://www.leibniz-ipn.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

REBECCA KNOBLOCH, ELKE SUMFLETH UND MAIK WALPUSKI

Förderung der Qualität fachinhaltlicher Schüleräußerungen in experimenteller Kleingruppenarbeit im Chemieunterricht

Fostering the quality of content-related students' statements in experimental group-work in chemistry classes

ZUSAMMENFASSUNG

In dem vorliegenden Beitrag wird eine Untersuchung zur Förderung der Qualität fachinhaltlicher Schüleräußerungen in schülerexperimentbasierten Kleingruppenarbeitsphasen im Chemieunterricht vorgestellt. Hierzu wurden zunächst mittels einer Reanalyse bestehender Videodaten die Merkmale erfolgreicher und weniger erfolgreicher Kleingruppen beschrieben. Anschließend wurde eine Interventionsstudie durchgeführt, um aufzuklären, ob Schülerinnen und Schüler mit hohem Lernerfolg eine hohe Kommunikationsqualität aufweisen oder ob eine hohe Kommunikationsqualität zu höherem Lernerfolg führt. Hierzu wurde ein experimentelles Kontrollgruppendesign verwendet. Die Interventionsmaßnahmen basierten auf den Ergebnissen der Reanalyse und hatten das Ziel, qualitativ hochwertigere Redebeiträge zu fördern. An der Untersuchung nahmen 192 Probandinnen und Probanden der 7. Klasse von zwölf Gymnasien in NRW teil. In fünf Unterrichtsstunden bearbeiteten sie experimentelle Problemlöseaufgaben. Kognitive Fähigkeiten und Vorwissen wurden kontrolliert und zum Balancieren von Kontroll- und Interventionsgruppe verwendet. Mittels Videoanalyse von 240 Videos der Kleingruppenarbeit sowie der Fachwissenstests konnte gezeigt werden, dass die Förderung der Fachkommunikation erfolgreich war. Es konnte ebenfalls gezeigt werden, dass eine Steigerung der Qualität fachinhaltlicher Aussagen auch mit einer Steigerung des Lernerfolgs einhergeht.

Schlüsselwörter: Fachkommunikation, Kleingruppenarbeit, Videoanalyse, Experimente

ABSTRACT

The goal of this study is to foster the quality of content-related students' statements in chemistry classes using experimental inquiry-tasks. First of all, the characteristics of successful and less successful small-groups concerning their quality of content-related communication were described by a reanalysis of existing video data. On basis of this reanalysis the instructions for the group work were revised in order to achieve a higher rate of high quality content-related statements. These revised instructions were used in an intervention study to investigate if a higher rate of high quality content-related statements also leads to a higher learning achievement or vice versa. For this reason an experimental pre-test post-test control group design was used. Data was collected over five lessons on acids and bases at 12 secondary schools in North Rhine-Westphalia making

a sample of 192 students in total. Cognitive skills and previous knowledge were controlled and used for balancing control and intervention group. By means of the analysis of 240 videos of the group work and tests on content knowledge it could be shown that the revised instructions lead to a higher rate of high quality content-related statements. It could be shown that an increase of the content-related quality of students' statements also leads to a higher learning achievement.

Keywords: communication, small groups, video analysis, experiments

1 Einleitung

Die Forschungsliteratur zeigt einen engen Zusammenhang zwischen Verbalisierungen und dem Wissenserwerb der Schülerinnen und Schüler auf (z. B. Chi & Bassok, 1989; Chi, de Leeuw, Chiu & LaVancher, 1994; Ferguson-Hessler & de Jong, 1990; Renkl, 1997). Insbesondere die Kommunikation der Schülerinnen und Schüler untereinander kann zu einer tieferen Auseinandersetzung mit dem Inhalt und zu einem tieferen Verständnis des Inhalts führen (Rotering-Steinberg, 1995). Zudem konnte Schulz (2011) einen engen Zusammenhang zwischen den Redeanteilen der Schülerinnen und Schüler im Chemieunterricht und ihrem Wissenserwerb aufzeigen. Auch andere Arbeiten stützen die Annahme, dass ein Zusammenhang zwischen dem Lernerfolg und der Kommunikation besteht (z. B. Solomon, 1988). Die Wirkrichtung zwischen den beiden Aspekten ist bisher jedoch noch nicht geklärt. Es kann also nicht endgültig gesagt werden, ob sich erfolgreiche Schülerinnen und Schüler häufiger am Unterrichtsgespräch beteiligen, oder ob diese erfolg-

reicher sind, weil sie sich häufiger am Unterrichtsgespräch beteiligen. Durch die Benennung als gleichberechtigter Kompetenzbereich in den Nationalen Bildungsstandards wird die wesentliche Rolle der fachlichen Kommunikation für das Lernen auch seitens der Bildungsadministration verdeutlicht. Es muss jedoch festgestellt werden, dass im regulären Unterricht die Lehrpersonen das kommunikative Unterrichtsgeschehen unverändert dominieren – so liegt der Satzanteil der Lehrkraft in der Regel zwischen 60 und 85 % (Schulz, 2011; Sumfleth, 1998).

1.1 Kommunikationsanalysen

Eine Analyse der umfangreichen Forschungsliteratur im Bereich der Kommunikation im Unterricht zeigt, dass sich diese im Wesentlichen auf Argumentationen, Entscheidungsfindungen, soziale Aspekte von Kommunikation und Verbalisierungen bezieht, jedoch beschäftigen sich nur wenige Arbeiten mit der Qualität fachinhaltlicher Aussagen. Dennoch können aus diesen Forschungsrichtungen wichtige Hinweise für

die Beurteilung der inhaltlichen Qualität von Schüleräußerungen gewonnen werden. Daher soll an dieser Stelle ein Überblick über die wichtigsten Erkenntnisse gegeben werden, der aufgrund der Vielzahl von Publikationen jedoch keinen Anspruch auf Vollständigkeit erhebt.

Am interessantesten erscheinen im Hinblick auf die hier vorgestellte Studie die Arbeiten zur Argumentationsforschung. Diese thematisieren, wie Personen Behauptungen und Schlussfolgerungen produzieren und begründen (Driver, Newton & Osborne, 2000; Sadler, 2004; Zohar & Nemet, 2002). Der Fokus der Forschung liegt zumeist auf der Beschreibung und Optimierung des Argumentationsverhaltens (von Aufschnaiter, Erduran, Osborne & Simon, 2008; Driver et al., 2000; Means & Voss, 1991; Zohar & Nemet, 2002). Im Bereich der Argumentationsanalyse liegen zudem Arbeiten vor, welche die Qualität von Argumenten beurteilen (z. B. Kuhn & Reiser, 2005; Lizotte, McNeill & Krajcik, 2004; Osborne, Erduran & Simon, 2004). Die meisten der Studien legen hierbei das Argumentstrukturmodell nach Toulmin (1958) zugrunde, wobei die Qualität von Argumenten anhand struktureller Kriterien definiert wird, d. h., es wird die Häufigkeit bestimmter Argument-Strukturelemente bestimmt (z. B. Erduran, Simon & Osborne, 2004). Studien, deren Schwerpunkt auf Entscheidungsfindungen liegt, fokussieren meist auf die Beschreibung und Optimierung von Entscheidungsstrategien (z. B. Abelson & Levi, 1985). Fachinhaltliche Gründe für eine Entscheidung werden hierbei zwar erfasst, jedoch wird die Qualität der

fachinhaltlichen Begründungen der Entscheidungen nicht untersucht. Sowohl im Bereich der Argumentationsforschung als auch im Bereich der Entscheidungsfindungen wird auf das Vorhandensein und die Vernetzung von Begründungen fokussiert, jedoch nicht auf die Qualität der Begründungen selbst. Bei der Betrachtung von Kommunikation spielen auch immer soziale Aspekte eine Rolle. Studien, die sich auf diesen Aspekt der Kommunikation fokussieren, untersuchen häufig Rollenverteilungen und soziale Strukturen innerhalb von Gruppen. Hierbei konnte beispielsweise gezeigt werden, dass bei aufgabenbezogenen Gruppenarbeiten häufig zentralisierte Kommunikationsmuster vorherrschen. Dabei nehmen eher aufgabenorientierte Gruppenmitglieder die zentrale Rolle ein.

Instrumente zur Analyse der Qualität fachinhaltlicher Aussagen liegen bisher kaum vor. Die existierenden Instrumente unterscheiden zudem oftmals nur zwischen aufgaben- und nicht-aufgabenbezogenen Aussagen (Gijlers & de Jong, 2005) oder auch zwischen inhalts- und nicht-inhaltsbezogenen Aussagen (Rumann, 2005; Sangin, Molinari, Dillenbourg, Rebetez & Bétrancourt, 2006), differenzieren aber die Qualität nicht weiter. Nur wenige Arbeiten unterscheiden inhaltliche Aussagen weiter, beispielsweise hinsichtlich ihrer fachlichen Richtigkeit (z. B. Franke-Braun, 2008; Hänze & Berger, 2007). Durch die kategoriale Zuordnung in „richtig“ und „nicht richtig“ wird allerdings lediglich eine quantitative Analyse der inhaltlichen Aussagen möglich, keine qualitative. Ein Beispiel für ein Kategoriensystem

zur Kommunikationsanalyse stammt von Franke-Braun (2008). Sie untersuchte in ihrer Arbeit den Einfluss gestufter Lernhilfen auf die sachbezogene Kommunikation. Hierbei unterscheidet sie aufgaben- und inhaltsbezogene Aussagen von nicht-aufgabenbezogenen Aussagen. Die aufgaben- und inhaltsbezogenen Aussagen wurden anschließend noch hinsichtlich ihrer Richtigkeit untersucht. Sie konnte zeigen, dass sich der Einsatz der gestuften Hilfen positiv auf den Anteil der sachbezogenen Kommunikation auswirkt.

1.2 Analyse der Komplexität von Aussagen

Hänze und Berger (2007) unterscheiden in ihrem Kategoriensystem neben der Richtigkeit der Aussagen auch zwischen den Kategorien *physikalischer Begriff* und *Verknüpfung von physikalischen Begriffen*. Durch die Kategorie *Verknüpfung von physikalischen Begriffen* werden Äußerungen beschrieben, bei denen mindestens zwei

physikalische Begriffe in Beziehung gesetzt werden.

Diese Kategorien weisen Ähnlichkeiten zu den Komplexitätsstufen „1 Fakt“ und „1 Zusammenhang“ der Dimension der *Komplexität* des Kompetenzstrukturmodells zur Evaluation der Nationalen Bildungsstandards in den Naturwissenschaften auf (Walpuski et al., 2010) (siehe Abbildung 1). In dem Modell wird die *Komplexität* systematisch variiert, um Testaufgaben einer definierten Schwierigkeit konstruieren zu können. Die *Komplexität* wird hierbei sowohl auf qualitativer Ebene (Fakt, Zusammenhang, Konzept) als auch auf quantitativer Ebene (1 Fakt, 2 Fakten, 1 Zusammenhang, 2 Zusammenhänge) unterschieden (Kauertz; Fischer; Mayer; Sumfleth, Walpuski 2010). Hänze und Berger (2007) beschreiben im Gegensatz zum Kompetenzstrukturmodell allerdings kein Konzeptniveau. Zudem werden bei dem Modell zur Evaluation der Nationalen Bildungsstandards in den Naturwissenschaften auch die dahinterstehenden kognitiven Prozesse betrachtet.

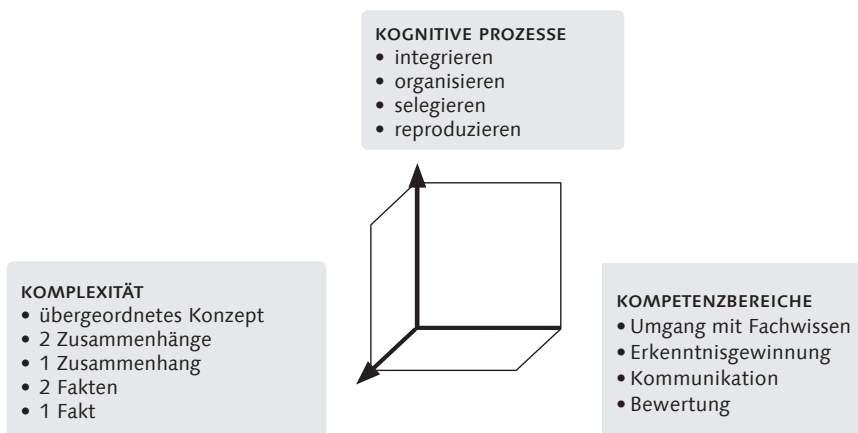


Abb. 1: Kompetenzstrukturmodell aus dem Projekt ESNaS (Walpuski et al., 2010, S. 177).

Die Dimension der kognitiven Prozesse (reproduzieren, selektieren, organisieren, integrieren) definiert, welcher kognitive Anspruch an die Bearbeitung der Aufgabe gestellt wird (Walpuski et al., 2010). Sowohl die Komplexität als auch die kognitiven Prozesse gelten bei Testaufgaben als schwierigkeitsbestimmend (Kauertz, 2008; Ropohl, 2010). Mit Hilfe des Kompetenzstrukturmodells zur Evaluation der Nationalen Bildungsstandards in den Naturwissenschaften kann dementsprechend die Leistung von Schülerinnen und Schülern beschrieben werden.

In der vorliegenden Studie soll die Schülerleistung aufgrund von Sprachhandlungen beschrieben werden. Aufgrund dessen wird für die vorliegende Studie die Grundidee des Modells adaptiert und für eine Videokodierung optimiert. Analog zu dem Kompetenzmodell wurde zuvor bereits ein Videokodierschema entwickelt, welches der Beschreibung des Vernetzungsgrads von Unterrichtsinhalten dient (Knobloch, Sumfleth & Walpuski, 2011). Hierbei wird die Komplexität von Aussagen analysiert, indem Aussagen auf Faktenniveau, Zusammenhangsniveau und Konzeptniveau unterschieden werden.

Insgesamt lässt sich festhalten, dass viele der bereits vorliegenden Instrumente zur inhaltlichen Qualitätsanalyse nur ausgewählte Teilbereiche beschreiben. Für eine detaillierte Analyse der fachlichen Qualität von Schüleräußerungen wird in dieser Studie das ESNaS (Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I) Kompetenzstrukturmodell (Walpuski et al., 2010) mit den Dimensionen *Komplexität* und *kognitive*

Prozesse als mögliche Qualitätsmerkmale inhaltlicher Schüleraussagen verwendet.

2 Ziele der Untersuchung

Das Ziel der Studie besteht darin, den Einfluss qualitativ hochwertiger Fachkommunikation (fachlich richtige Aussagen auf einem höheren Komplexitätsniveau und einem höheren kognitiven Prozess) auf den Lernerfolg zu untersuchen. Zu diesem Zweck soll ein Instrument entwickelt werden, mit dem eine detaillierte Analyse der Qualität fachinhaltlicher Schüleräußerungen videobasiert durchgeführt werden kann. Ausgehend von dieser Zielsetzung ergeben sich die folgenden Forschungsfragen:

- (1) Können in der Videoanalyse die Komplexität und die kognitiven Prozesse als Qualitätsmerkmale fachinhaltlicher Schüleräußerungen nachgewiesen werden?
- (2) Kann durch die gezielte Anregung fachlicher Äußerungen deren Qualität und Umfang erhöht werden?
- (3) Kann durch optimierte fachinhaltliche Kommunikationsanteile das fachliche Lernen gesteigert werden?
- (4) Kann eine Steigerung der fachinhaltlichen Qualität der Schüleräußerungen aufrechterhalten werden, wenn die Anregung dazu weggelassen wird?

Zu den Forschungsfragen wurden folgende Annahmen formuliert:

Es wird davon ausgegangen, dass die Qualität der fachinhaltlichen Aussagen einen

Einfluss auf den Lernerfolg hat. Um die Qualität zu bestimmen, wird basierend auf mehreren Vorarbeiten (Neumann et al., 2008; Kauertz et al., 2010; Sumfleth, Neuroth, & Leutner, 2010) davon ausgegangen, dass die Qualität fachlicher Schüleräußerungen bestimmt wird durch:

- die fachliche Richtigkeit
- die Komplexität (Fakten-, Zusammenhangs- oder Konzeptniveau)
- die kognitiven Prozesse (reproduzieren, selektieren, organisieren und integrieren)

Zudem wird angenommen, dass die Schülerinnen und Schüler, welche Aussagen einer hohen Qualität formulieren, d. h. fachlich richtige Äußerungen, die sich auf einem hohen Komplexitätsniveau befinden und durch einen hohen kognitiven Prozess gekennzeichnet sind, auch in Leistungstests besser abschneiden.

3 Design und Instrumente der Studie

3.1 Ablauf der Studie

Zur Beantwortung der Forschungsfragen wurde die Studie in mehreren Phasen durchgeführt (siehe Abbildung 2). Zunächst wurde ein Kategoriensystem entwickelt, mit dem die fachinhaltliche Qualität von Schüleräußerungen detailliert analysiert werden kann (siehe Kapitel 4). Mittels des entwickelten Kategoriensystems wurden bereits existierende Videodaten eines vorhergehenden Projekts (Walpuski, 2006) reanalysiert. Hierbei handelt es sich um Videos experimenteller Kleingruppenarbeit zum Thema Säuren und Basen. Das Ziel der Reanalyse bestand zum einen darin, das entwickelte Kategoriensystem zu überprüfen (Knobloch et al., 2011). Zum anderen

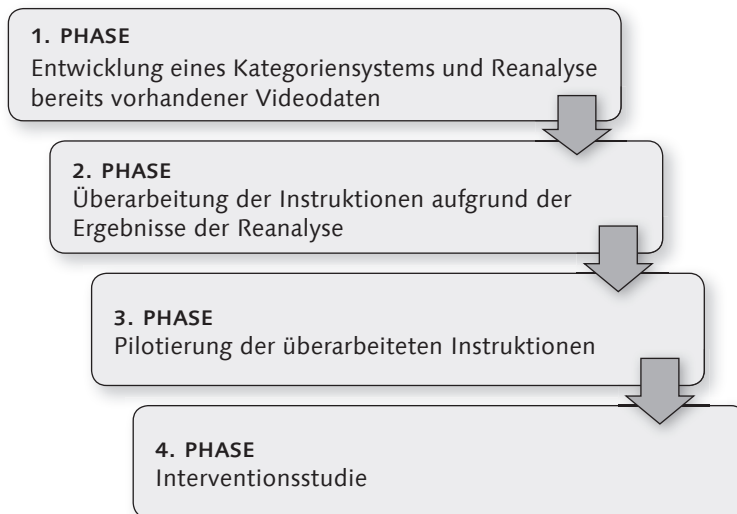


Abb. 2: Durchführung der Studie in mehreren Phasen.

sollte mittels der Reanalyse die erste Forschungsfrage beantwortet werden, indem die Merkmale besonders erfolgreicher und weniger erfolgreicher Kleingruppen – bezogen auf die fachinhaltliche Qualität ihrer Aussagen – beschrieben werden. Durch Korrelationsanalysen werden mögliche Qualitätsmerkmale als unabhängige Variablen generiert. Da durch diese Korrelationsanalysen jedoch lediglich Zusammenhänge aufgezeigt werden können und keine Wirkrichtung bestimmt werden kann, ist anschließend eine Interventionsstudie notwendig. In der durchgeführten Interventionsstudie wurden die Instruktionen der experimentellen Kleingruppenarbeit des vorherigen Projekts (Walpuski, 2006) auf Grundlage der Ergebnisse der Reanalyse dahingehend überarbeitet, dass qualitativ hochwertige Schüleräußerungen gefördert werden. In einer weiteren Phase wurden die überarbeiteten Instruktionen der experimentellen Kleingruppenarbeit zunächst mittels einer kleinen Stichprobe ($N=8$ Schülerinnen und Schüler) pilotiert. Anschließend wurde die Interventionsstudie durchgeführt. Um die Forschungsfragen beantworten zu können sollte die in der Interventionsstudie eingesetzte Unterrichtsreihe ein hohes Maß an Schüler-Schüler-Interaktion zulassen, um die fachinhaltliche Qualität der Äußerungen und ihren Bezug zum Lernerfolg optimal untersuchen zu können. Die konkrete methodische Konzeption der Unterrichtsreihe wird im Folgenden detailliert beschrieben.

3.2 Methodische Konzeption der Unterrichtsreihe

Für die vorliegende Studie wurde eine bereits existierende Unterrichtsreihe verwendet, bei der die Schülerinnen und Schüler selbstständig in Kleingruppen mit Interaktionsboxen arbeiten (Rumann, 2005; Walpuski, 2006). Es handelt sich hierbei um Interaktionsboxen mit experimentellen Problemlöseaufgaben zum Thema Säuren und Basen für das erste Lernjahr Chemie (in NRW: 7. Jahrgangsstufe). In den Interaktionsboxen sind die jeweilige Aufgabenstellung, Informationen und kleine Hilfestellungen sowie die Chemikalien und Geräte enthalten (Rumann, 2005). Die Thematik Säuren und Basen wurde gewählt, da es sich hierbei um kein obligatorisches Thema in der Jahrgangsstufe 7 handelt. Somit kann davon ausgegangen werden, dass das Vorwissen der Lernenden zu diesem Thema nicht sehr groß ist und daher der Lernerfolg angemessen untersucht werden kann. Die Schwierigkeit der experimentellen Problemlöseaufgaben steigt innerhalb der fünf Stunden stetig an (Walpuski & Sumfleth, 2007).

3.3 Instrumente

Bezüglich der eingesetzten Instrumente muss zwischen der Reanalyse (Knobloch et al., 2011) sowie der Interventionsstudie unterschieden werden. Im Rahmen der Studie von Walpuski (2006), aus der die Videodaten für die Reanalyse verwendet wurden, wurden begleitend zu der Klein-

gruppenarbeit sowie zu drei weiteren Zeitpunkten Leistungsdaten erhoben.

Der Interventionsstudie der vorliegenden Arbeit liegt ebenfalls ein experimentelles „Pretest-posttest control group design“ nach Campbell und Stanley (1963) zugrunde. Eine detaillierte Übersicht der Testzeitpunkte und eingesetzten Instrumente findet sich in Tabelle 1.

Ein Großteil der eingesetzten Erhebungsinstrumente wurde aus dem Projekt von Walpuski (2006) übernommen, sodass diese Instrumente als bereits überprüfte, reliable und valide Messinstrumente vorlagen. Da der fachspezifische Leistungstest von Walpuski (2006) hauptsächlich Faktenwissen erhebt, wurde zusätzlich im Post- und Follow-up Test der Triadentest (nach Neuroth, 2002) eingesetzt, welcher komplexeres vernetztes Zusammenhangswissen überprüft. Es handelt sich um einen Test mit offenem Aufgabenformat. Hierbei muss der Zusammenhang von jeweils drei Begriffen beschrieben werden (Neuroth, 2002). In der vorliegenden

Studie wird davon ausgegangen, dass unter anderem das Komplexitätsniveau von Aussagen (Faktenniveau, Zusammenhangsniveau, Konzeptniveau) eine Rolle für das Lernen spielt. Aufgrund dessen sollte der Leistungsunterschied nicht nur auf Faktenniveau erhoben werden. Entsprechend wurde der Triadentest von Neuroth (2002) überarbeitet und es wurden neue Triaden entwickelt und pilotiert. Zur Entwicklung des Kategoriensystems wurden zunächst theoriegeleitet (deduktiv) relevante Kategorien aus der Literatur identifiziert und anschließend empiriegeleitet (induktiv) weitere Kategorien ergänzt (siehe dazu auch Knobloch et al., 2011). Die Videos wurden in einer turnbasierten Kodierung analysiert, d.h. jeder zusammenhängende Sprechakt, der auf einen inhaltlichen Aspekt begrenzt ist, wurde als „Turn“ definiert. Ein Turn entspricht demnach in der Regel einer Schüleräußerung. Die Kodierung erfolgte in zwei Durchgängen. Zunächst wurde eine Basiskodierung durchgeführt, um eine spätere vertiefende Analyse zu ver-

Tab. 1: Übersicht über die eingesetzten Testinstrumente sowie Testzeitpunkte der Interventionsstudie

ZEITPUNKT	TESTINSTRUMENT
Pretest	• Kognitiver Fähigkeitstest (Heller & Perleth, 2000) • Fachtest (Multiple-Choice) (Walpuski, 2006) • Naturwissenschaftliche-Arbeitsweisen-Test (NAW) (Klos, Henke, Kieren, Walpuski, & Sumfleth, 2008)
Begleitend	• Stundenspezifischer Fachtest (Multiple-Choice) (Walpuski, 2006) • Videos (Kategoriensystem)
Posttest	• Fachtest (Multiple-Choice) (Walpuski, 2006) • Fachtest (Triadentest) (nach Neuroth, 2002) • Naturwissenschaftliche-Arbeitsweisen-Test (NAW) (Klos et al., 2008)
Follow-up Test	• Fachtest (Multiple-Choice) (Walpuski, 2006) • Fachtest (Triadentest) (nach Neuroth, 2002) • Naturwissenschaftliche-Arbeitsweisen-Test (NAW) (Klos et al., 2008)

einfachen. In dieser vertiefenden Analyse wurden die inhaltlichen Aussagen hinsichtlich der drei angenommenen Qualitätsmerkmale *fachliche Richtigkeit*, *Komplexität* und *kognitive Prozesse* detailliert analysiert. Da in authentischen Gesprächssituationen sehr unterschiedliche Anzahlen von Fakten oder Zusammenhängen pro Aussage zu erwarten sind, wurde bezogen auf die Komplexität nur eine Unterscheidung auf qualitativer Ebene (Fakt, Zusammenhang, Konzept) und keine zusätzliche quantitative Unterscheidung (1 Fakt, 2 Fakten, 1 Zusammenhang, 2 Zusammenhänge) gewählt. Daher werden im Weiteren die Begrifflichkeiten *Fakten-*, *Zusammenhangs-* und *Konzeptniveau* verwendet. Die kognitiven Prozesse *Organisieren* und *Integrieren* sind nach dem Kompetenzstrukturmodell aus dem Projekt ESNaS (Walpuski et al., 2010, S. 177) nur auf *Zusammenhangsniveau* und *Konzeptniveau* möglich.

Beispiele für Aussagen verschiedener kognitiver Prozesse:

Reproduzieren: „Jetzt ist die Lösung rot geworden.“

Selektieren: Sprecher A: „Welche Farbe hat Bromthymolblau nochmal bei einer neutralen Lösung?“

Sprecher B: „Nach Zugabe von Bromthymolblau zu einer neutralen Lösungen färbt sich die Lösung grün.“ (*Selektieren!*)

Organisieren: „Die Salzsäure mit dem pH-Wert 2 ist ätzender als die Salzsäure mit dem pH-Wert 3.“

Integrieren: „Wenn der Indikator bei Zugabe von Salzsäure rot wird, dann muss alles Saure rot werden.“

In Abbildung 3 werden das Kategoriensystem sowie die beiden Kodierdurchgänge dargestellt.

Zur Überprüfung des Kategoriensystems wurde die Beobachterübereinstimmung berechnet. Hierzu wurde Cohen's Kappa (κ) als Übereinstimmungsmaß gewählt, da bei diesem Maß eine Zufallskorrektur der getroffenen Einschätzungen erfolgt (Bortz & Döring, 2006; Wirtz & Caspar, 2007). Die Überprüfung des Gütekriteriums der Objektivität erfolgte aufgrund einer Doppelkodierung von 10 Unterrichtsvideos der Reanalyse. Die Analyse der Videos erfolgte mit Hilfe der Software Videograph® (Rimmele, 2004). Die Rater wurden zunächst auf Basis eines erstellten Kodiermanuals mit Hilfe von Videodaten geschult, welche nicht mit in die Auswertung eingingen. Die durch das Training erzielte Beobachterübereinstimmung zweier Rater liegt bei $.82 < \kappa \leq .99$ und ist somit sehr gut.

4 Reanalyse

4.1 Stichprobe der Reanalyse

Die Videodaten der Reanalyse stammen aus dem 2. Halbjahr des Schuljahres 2004/2005 (Walpuski, 2006). Walpuski untersuchte in einer Studie insgesamt 84 Kleingruppen mit je 4 Schülerinnen und Schülern ($N=336$). Hiervon wurden 28 Gruppen ($N=112$ Schülerinnen und Schüler) während der fünf Kleingruppenarbeitsphasen videographiert. Insgesamt lagen also 140 Videos vor. Zudem lagen von allen Schülerinnen und Schülern die Leistungsdaten aus den pre und post durchge-

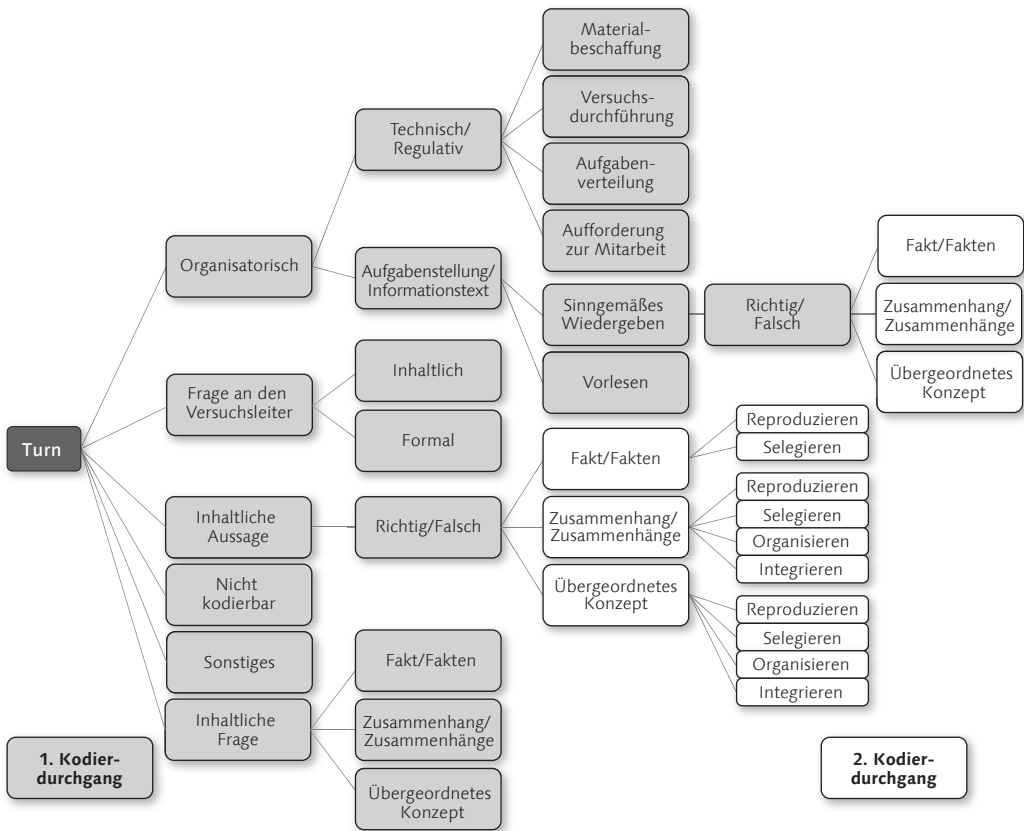


Abb. 3: Kategoriensystem zur Analyse der Qualität fachinhaltlicher Schüleräußerungen (1. Kodierdurchgang in grau dargestellt, 2. Kodierdurchgang in weiß dargestellt).

fürten Fachwissenstests vor. Da durch die Reanalyse unter anderem die Merkmale der fachinhaltlichen Qualität der Schüleräußerungen besonders erfolgreicher und weniger erfolgreicher Gruppen beschrieben werden sollten, wurden bezogen auf den Lernerfolg Extremgruppen gebildet. Um die Gruppenmerkmale beschreiben zu können, wurden die Gruppenaggregate des residualen Lernzuwachses im fachspezifischen Leistungstest verwendet. Eine Auswertung auf Ebene der Gruppenaggregate wurde bei den Leistungstests deshalb gewählt, weil sowohl der Experimentiererfolg als auch die Gruppen-

kommunikation per se ein Gruppenaggregat bzw. einen Gruppenprodukt darstellen. Im Weiteren wird nicht auf die Leistung einzelner Schülerinnen und Schüler in Gruppen fokussiert, sondern ausschließlich auf die Gruppenaggregate, sodass auf eine Analyse im Mehrebenenmodell verzichtet wurde, zumal die Größe der Stichprobe auf Ebene 2 in den Extremgruppen relativ klein wäre. Damit einerseits eine Aufteilung der Kleingruppen in erfolgreiche und weniger erfolgreiche Kleingruppen möglich war, aber dennoch eine ausreichend große Stichprobe gewährleistet war, wurde eine Grenze

von 20 % verwendet. Erfolgreiche Gruppen zählten demnach zu den besten 20 %; weniger erfolgreiche Gruppen zählten zu den schlechtesten 20 % bezogen auf den Lernzuwachs im fachspezifischen Leistungstest. Von jeder Extremgruppe wurden jeweils 15 Videos analysiert. Die Stichprobe der Reanalyse bestand insgesamt aus 64 Schülerinnen und Schülern der Extremgruppen, wobei 58,7 % weiblich waren.

4.2 Datenauswertung und Ergebnisse der Reanalyse

Die 30 Videos wurden mit Hilfe des oben dargestellten Kategoriensystems analysiert. Zur Beschreibung der Merkmale erfolgreicher und weniger erfolgreicher Kleingruppen wurden die so gewonnenen Daten mit der Software PASW 18 mit t-Tests verglichen. Die Voraussetzungen für den t-Test sind erfüllt. Um die Vergleichbarkeit der Videos trotz unterschiedlicher Vi-

deolänge gewährleisten zu können, musste die Anzahl der beobachteten Turns in Relation zur Minutenanzahl sowie in einem zweiten Auswertungsschritt in Relation zur gesamten Turnanzahl des entsprechenden Videos gesetzt werden. Im Folgenden werden die Ergebnisse der Videoauswertung dargestellt sowie die Ergebnisse der durchgeführten Korrelationsanalysen. Die Auswertung der Anzahl der Turns in Relation zur Gesamtzahl der Turns zeigt die gleichen Tendenzen wie im Verhältnis zur Arbeitszeit (die für alle Gruppen annähernd gleich war), sodass davon ausgegangen werden kann, dass es keine großen Unterschiede in der rein quantitativen Nutzung der Redezeit gab. Im Folgenden wird zur Vereinfachung nur die Auswertung der Turns pro Minute dargestellt.

Die Betrachtung der Oberkategorien zeigt keine statistisch bedeutsamen Unterschiede zwischen den erfolgreichen und weniger erfolgreichen Kleingruppen (siehe Abbildung 4). Lediglich ist die Ten-

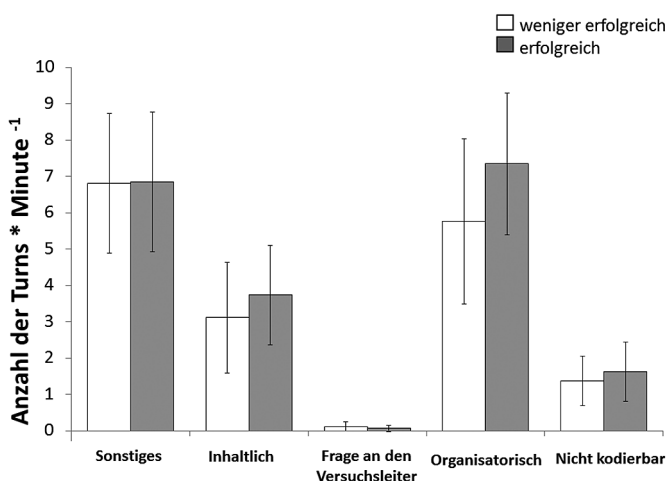


Abb. 4: Anzahl der Turns pro Minute der Oberkategorien (Mittelwert und Standardabweichung).

denz erkennbar, dass die erfolgreichen Gruppen insgesamt mehr Äußerungen tätigen. Anhand der Sichtstruktur ist demnach keine hinreichende Unterscheidung der beiden Extremgruppen möglich.

Aufgrund dessen wird im Folgenden die detaillierte Auswertung der drei angenommenen Qualitätsmerkmale fachinhaltlicher Schüleräußerungen – die *fachliche Richtigkeit*, die *Komplexität* und die *kognitiven Prozesse* – dargestellt. Die Auswertung hinsichtlich der *fachlichen Richtigkeit* der inhaltlichen Aussagen zeigt, dass die erfolgreichen Gruppen hoch signifikant ($p < .01$) mehr fachlich richtige Aussagen tätigen als die weniger erfolgreichen Kleingruppen. Tendenziell machen die weniger erfolgreichen Gruppen mehr fachlich falsche Aussagen, wobei dieser Unterschied jedoch nicht statistisch bedeutsam ist (siehe Abbildung 5).

Die Auswertung der Komplexität der inhaltlichen Aussagen zeigt zudem, dass die erfolgreichen Kleingruppen tendenziell weniger Aussagen auf Faktenniveau und mehr Aussagen auf der höheren Komplexitätsstufe, dem Zusammenhangsniveau, ($p < .01$) tätigen als die weniger erfolgreichen Gruppen (siehe Abbildung 6). Insgesamt konnte nur ein einziger Turn der Komplexitätsstufe Konzeptniveau gefunden werden. Da es sich bei dem Konzeptniveau jedoch um das schwierigste Komplexitätsniveau handelt, war es durchaus erwartungskonform wenige Aussagen dieses Niveaus zu finden, obwohl durchaus in der untersuchten Jahrgangsstufe (1. Lernjahr Chemie: 7. Klasse in NRW) auf Basis der verwendeten Materials Aussagen auf Konzeptniveau, beispielsweise zum Konzept der Neutralisation denkbar waren. Bei Betrachtung der kognitiven Prozesse

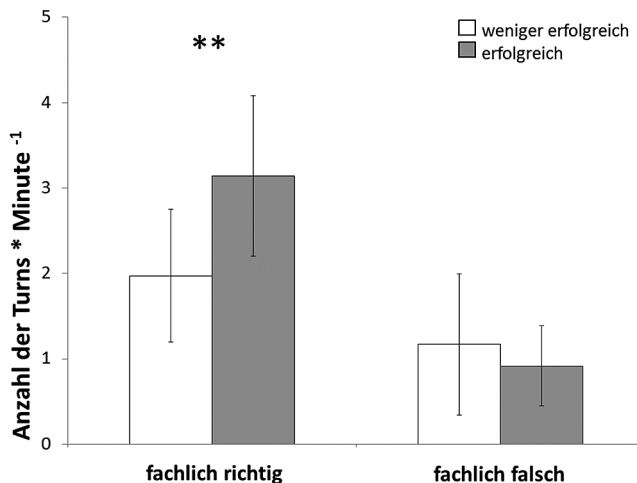


Abb. 5: Häufigkeiten der inhaltlichen Äußerungen hinsichtlich der fachlichen Richtigkeit (Mittelwert und Standardabweichung).

der inhaltlichen Aussagen zeigt sich, dass die erfolgreichen Kleingruppen höchst signifikant ($p < .001$) mehr Aussagen der höheren kognitiven Prozesse Selegieren und Organisieren machen und signifikant ($p < .05$) mehr Aussagen des kognitiven Prozesses Integrieren tätigen (siehe Abbildung 7).

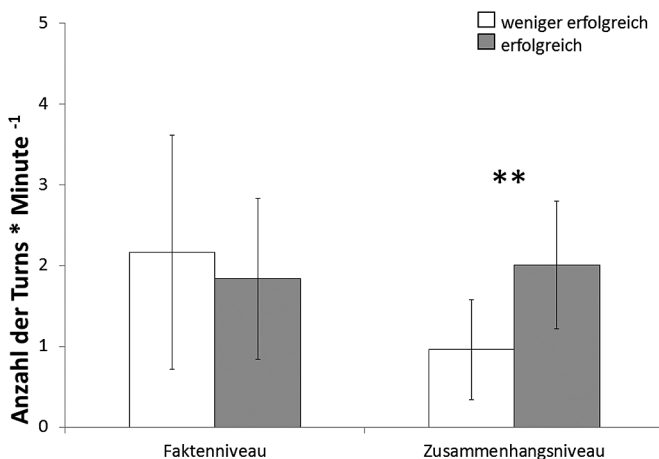


Abb. 6: Häufigkeiten der inhaltlichen Äußerungen hinsichtlich der Komplexität (Mittelwert, SA).

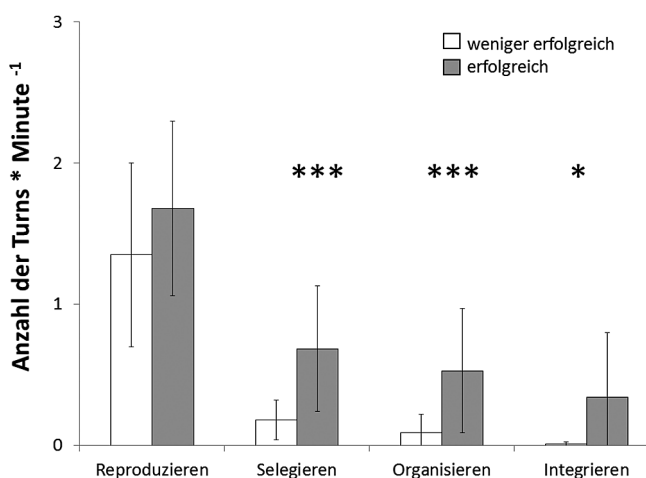


Abb. 7: Kognitive Prozesse bezogen auf inhaltliche Äußerungen (Mittelwert, SA).

4.3 Konsequenzen aus den Ergebnissen der Reanalyse

Die dargestellten Ergebnisse der Reanalyse zeigen, dass bei rein quantitativer Betrachtung der Redeanteile kaum Unterschiede zwischen den erfolgreichen und weniger erfolgreichen Gruppen bestehen. Die Analysen bezüglich der drei angenommenen Qualitätsmerkmale inhaltlicher Aussagen zeigen jedoch signifikante Unterschiede zwischen den beiden Gruppen. So konnte gezeigt werden, dass die erfolgreichen Gruppen sich durch eine größere Anzahl an fachlich richtigen Aussagen, Aussagen auf Zusammenhangsniveau und Aussagen, die den höheren kognitiven Prozessen selektieren, organisieren und integrieren zuzuordnen sind, auszeichnen. Die Ergebnisse zeigen einen engen korrelativen Zusammenhang zwischen der Qualität inhaltlicher Aussagen und dem Lernerfolg. Die dargestellten Ergebnisse zeigen allerdings nur korrelative Zusammenhänge und lassen keine Schlussfolgerung auf eine Wirkrichtung zu. Aufgrund dessen wurde, wie bereits erwähnt, in einer anschließenden Interventionsstudie untersucht, ob eine Steigerung der fachinhaltlichen Qualität der Schüleräußerungen zum einen möglich ist und ob diese zum anderen auch mit einem größeren Lernerfolg einhergeht.

5 Interventionsstudie

5.1 Design der Interventionsstudie

Zur Überprüfung der Lernwirksamkeit der überarbeiteten Instruktionen wurde

ein experimentelles pretest-posttest control group design analog zur Studie von Walpuski (2006) verwendet. Die Studie wurde jeweils an fünf aufeinanderfolgenden Tagen in einer 45-minütigen Unterrichtsstunde im Anschluss an den Regelunterricht durchgeführt. In der ersten Stunde erhalten sowohl die Kontroll- als auch die Interventionsgruppe die unveränderten Materialien aus der Studie von Walpuski (2006). Die erste Stunde dient dazu, zu kontrollieren, ob die fachinhaltliche Qualität der Schüleräußerungen in Kontroll- und Interventionsgruppe zu Beginn gleich ist. Die Interventionsgruppe erhält in den Stunden 2 bis 4 veränderte Instruktionen. Die letzte Stunde, in der die Interventionsgruppe ebenfalls wieder unveränderte Instruktionen erhält, soll der Beantwortung der vierten Forschungsfrage dienen und zeigen, ob mögliche Veränderungen der fachinhaltlichen Qualität der Schüleräußerungen anhalten, wenn die veränderten Instruktionen weggelassen werden. Die Kontrollgruppe erhielt über alle fünf Stunden hinweg das unveränderte Material aus der Studie von Walpuski (2006). Die Zeit wurde in beiden Gruppen gleich gehalten. Zu Beginn einer jeden Unterrichtsstunde wurde die Bearbeitung der Aufgaben den Schülerinnen und Schülern anhand eines standardisierten Einführungstexts erläutert. Nach den Kleingruppensequenzen erfolgte jeweils in beiden Gruppen ein standardisierter Kurzvortrag mit möglichen Lösungen durch den Versuchsleiter.

5 Interventionsmaßnahmen und deren Pilotierung

Basierend auf den Ergebnissen der Reanalyse wurden die Instruktionen der Kleingruppenarbeit mit dem Ziel überarbeitet, qualitativ hochwertige Schüleräußerungen zu fördern. Die Interventionsmaßnahmen sollten die drei Qualitätsmerkmale – *fachliche Richtigkeit*, *Komplexität* und *kognitive Prozesse* – anregen. Es ist allerdings schwer, auf die fachliche Richtigkeit der Schüleräußerungen Einfluss zu nehmen. Zudem zeigen die Ergebnisse der Reanalyse, dass sich die erfolgreichen von den weniger erfolgreichen Gruppen mit Blick auf die fachlich falschen Aussagen nicht signifikant unterscheiden. Aufgrund dessen wurde bei den Interventionsmaßnahmen kein expliziter Versuch zur Förderung der Richtigkeit von Aussagen unternommen, wohl aber eine Steigerung der Komplexität der Schüleräußerungen und der kognitiven Prozesse angestrebt. Um den Austausch der Schülerinnen und Schü-

ler anzuregen, ist es zunächst notwendig sinnstiftende Kommunikationsanlässe zu schaffen. Dadurch wird zusätzlich ein Wissensaustausch erreicht, so dass das Wissen einzelner Gruppenmitglieder der gesamten Gruppe zur Verfügung steht. Die Ergebnisse der Reanalyse zeigen zudem, dass insbesondere die Aussagen angeregt werden sollten, in denen Zusammenhänge verbalisiert werden. Aufgrund dessen wurden Binnenphasen eingeführt, in denen explizit dazu aufgefordert wurde, Zusammenhänge zwischen relevanten Fachbegriffen der jeweiligen Kleingruppenarbeit zu benennen. Es gibt verschiedene Möglichkeiten, diese Verbalisierung von Zusammenhängen zu erreichen. Eine Möglichkeit ist die Erstellung einer Concept-Map während der Gruppenarbeit (z. B. Kramer, 2009). Des Weiteren können sogenannte Triaden bearbeitet werden, bei denen die Zusammenhänge zwischen drei Begriffen formuliert werden müssen (siehe Abb. 8). Hierbei werden zwei

1. FORSCHERTEAM-SITZUNG (3 MIN)	
1. Paraphrasieren der Aufgabenstellung durch den Forschungsleiter	
2. Reproduzieren auf Faktenniveau	
EXPERIMENTIERPHASE (20 MIN)	
2. FORSCHERTEAM-SITZUNG (5 MIN)	
Bearbeitung der Triaden	
z. B.: Rotkohlsaft – Neutralisation – Salzsäure	

Abb. 8: Ablauf der Kleingruppenarbeit der Interventionsgruppe in den Stunden 2–4.

Fachbegriffe in einer Aussage miteinander verbunden, die in einen durch den dritten Begriff bestimmten Kontext passt (Neuroth, 2002). Durch Triaden wird im Vergleich zu Concept Maps die Anzahl an Relationen begrenzt, was eine zeitökonomische Besprechung und Sicherung im Unterricht ermöglicht. In der vorliegenden Studie wurden daher u. a. Triaden als Interventionsmaßnahme eingesetzt. Bevor die Interventionsmaßnahmen in der Interventionsstudie eingesetzt wurden, wurden diese zunächst pilotiert. Zudem wurde der Triadentest eingesetzt und pilotiert. Die Stichprobe der Pilotierung bestand aus 8 Schülerinnen und Schülern, welche in zwei Vierergruppen zusammen arbeiteten. Während der fünf Kleingruppenarbeitssequenzen wurden die beiden Kleingruppen videographiert, um eine Analyse der Instruktionsnutzung und die Entwicklung von Optimierungsmöglichkeiten zu ermöglichen. Die endgültigen Interventionsmaßnahmen werden im Folgenden dargestellt.

Der Ablauf der Kleingruppensequenzen mit Interventionsmaßnahmen ist für jede Sequenz gleich (siehe Abbildung 8). Es wurden jeweils zwei Binnenphasen eingeführt – eine Binnenphase vor und eine Binnenphase nach der experimentellen Kleingruppenarbeit. Die Binnenphasen wurden „Forscherteamsitzungen“ genannt, um den Schülerinnen und Schülern deutlich zu machen, dass sie gemeinsam die Aufgaben bearbeiten sollen. Durch eine rotierende Rollenverteilung wurde dies zusätzlich unterstützt. In jeder Gruppe wurde täglich ein neuer „Forschungsleiter“ gewählt, welcher die Aufgabe hatte,

die Aufgabenkarten der Forscherteamsitzungen vorzulesen und darauf zu achten, dass die Fragen beantwortet wurden. Die anderen Gruppenmitglieder erhielten die Bezeichnung „Forschungsmitarbeiter“. Alle Gruppenmitglieder erhielten entsprechende Namensschilder.

In der jeweils ersten Forscherteamsitzung erhielten die Schülerinnen und Schüler neben den Aufgaben- und Informationskarten eine Materialienliste und ein Foto der Interaktionsbox. Diese erste Phase der Forscherteamsitzungen soll den Schülerinnen und Schüler die Möglichkeit geben sich an die Binnenphasen zu gewöhnen und ein gemeinsames Verständnis der Aufgabe sicherzustellen. Hierbei sollen die Lernenden zunächst in Anlehnung an die Arbeit mit gestuften Lernhilfen (Franke-Braun, 2008) die Aufgabenstellung paraphrasieren. Anschließend wurde eine Information, die bereits aus einer der vorherigen Sitzungen bekannt war oder aus dem Informationsmaterial stammte, auf *Faktenniveau reproduziert*, z. B.: „Nennt die Funktion des Rotkohlsafts.“ Die erste Binnenphase dauerte 3 Minuten. Nach dieser Zeit begann die experimentelle Kleingruppenarbeit, welche 20 Minuten dauerte. Anschließend wurde die zweite Forscherteamsitzung durchgeführt, welche 5 Minuten dauerte. Die Bearbeitung dieser zweiten Binnenphase erfolgte sowohl mündlich als auch schriftlich, um einen zusätzlichen Anlass zur Diskussion zu schaffen, denn in der Gruppe musste entschieden werden, was genau notiert wird. Es wurden jeweils zwei Triaden eingesetzt, welche für die jeweilige Kleingruppenarbeit relevante Fachbegriffe enthielten, aber

nicht im Triadentest zum Einsatz kamen. Durch den Einsatz der Triaden konnten implizit gelernte Zusammenhänge explizit formuliert werden.

5.3 Stichprobe Interventionsstudie

Die Stichprobe der Interventionsstudie bestand aus 48 Kleingruppen ($N=192$ Schülerinnen und Schüler), die von insgesamt 12 Gymnasien in Nordrhein-Westfalen stammten. Um eine erfolgreiche Zusammenarbeit der Schülerinnen und Schüler zu begünstigen, erfolgte die Gruppenzusammensetzung aufgrund der Wünsche der Schülerinnen und Schüler. Die Schülerinnen und Schüler stammten aus dem ersten Lernjahr Chemie (in NRW: 7. Jahrgangsstufe). 60,9 % der Probanden der Gesamtstichprobe waren männlich und 39,1 % weiblich. In der Kontrollgruppe waren 59,4 % männlich und in der Interventionsgruppe 62,5 %. Das Durchschnittsalter betrug 12,7 Jahre. Wie bereits erwähnt, wurden aufgrund der Ergebnisse im fachspezifischen Leistungstest und dem Kognitiven Fähigkeitstest im Pretest die Kontroll- und Interventionsgruppe balanciert. Alle Gruppen wurden während der Kleingruppenarbeit videographiert, so dass insgesamt 240 Videos aufgenommen wurden.

5.4 Ergebnisse der Interventionsstudie

Kontrolle der Rahmenbedingungen

Um sicherzustellen, dass die Balancierung erfolgreich war, wurden für die Kontroll-

und Interventionsgruppe die Mittelwerte der Punktesummen des kognitiven Fähigkeitstests und des Vorwissenstests mittels eines t-Tests verglichen. Um zu überprüfen, ob parametrische Verfahren zur Datenauswertung eingesetzt werden können, wurden die Werte zunächst auf Normalverteilung hin überprüft. Dem Kolmogorov-Smirnov-Signifikanztests zufolge kann beim fachspezifischen Leistungstest im Pretest ($p > .05$) sowie bei den kognitiven Fähigkeiten ($p > .05$) von normalverteilten Daten ausgegangen werden. Aufgrund dessen werden für die Auswertung t-Tests berechnet.

Erwartungsgemäß konnten keine signifikanten Unterschiede zwischen den beiden Gruppen bei den kognitiven Fähigkeiten ($t(189) = .376; p = .707$) und beim Vorwissen ($t(189) = -1.276; p = .204$) gefunden werden. Die 240 aufgenommenen Videos der Kontroll- und Interventionsgruppe wurden mittels des entwickelten Kategoriensystems ausgewertet. Die Videokodierung erfolgte turnbasiert mit Hilfe der Software Videograph® (Rimmele, 2004). Die Beobachterübereinstimmung wurde erneut überprüft. Dazu wurden 10 % der Videodaten (24 Videos) von zwei Ratern unabhängig voneinander kodiert. Die Beobachterübereinstimmung Cohen's Kappa liegt bei $.73 \leq \kappa \leq .99$ und liegt somit nach Wirtz und Caspar (2007) im zufriedenstellenden bis guten/sehr guten Bereich.

Ergebnisse der Videoanalyse

Wenn mehr als ein Schüler(in) einer Gruppe fehlten, wurde diese Kleingruppenarbeitsphase von der Auswertung aus-

geschlossen, da in Dyaden ein von den Kleingruppen abweichendes Kommunikationsverhalten zu erwarten ist. Außerdem wurden einzelne Videos mit einer schlechten Tonqualität von der Auswertung ausgenommen. Insgesamt wurden 233 Videos in die endgültige Auswertung aufgenommen. Zum Vergleich der beiden Treatmentgruppen wurden jeweils die Mittelwerte der einzelnen Kategorien und Unterkategorien der Kontroll- und Interventionsgruppe miteinander mittels eines t-Tests verglichen. Bei der Analyse der Videodaten der ersten Stunde, in der noch keine Interventionsmaßnahmen eingesetzt wurden, konnten keine statistisch bedeutsamen Unterschiede in den Schüleräußerungen zwischen der Kontroll- und Interventionsgruppe festgestellt werden. Daher kann davon ausgegangen werden, dass die im Folgenden dargestellten Unterschiede auf die Interventionsmaßnahmen

zurück zu führen sind und nicht von bereits vorhandenen Unterschieden zwischen den Gruppen herrühren.

Bezogen auf die Oberkategorien zeigen sich – wie schon in der Reanalyse – keine statistisch bedeutsamen Unterschiede zwischen den Treatmentgruppen. Dieses Ergebnis ist erwartungsgemäß, da die Intervention auf die Steigerung der Qualitätsmerkmale inhaltlicher Aussagen fokussierte, die in der Sichtstrukturanalyse nicht detailliert erfasst werden.

Die weiteren Analysen der Unterkategorien deuten darauf hin, dass die Interventionsmaßnahmen durch die Schülerinnen und Schüler angenommen wurden. So zeigt sich für die Unterkategorien *Aufgabenstellung/Informationstext* in der Kategorie *organisatorisch*, dass die Interventionsgruppe hochsignifikant ($p < .001$) mehr Zeit auf das Vorlesen von Texten verwendet und signifikant ($p < .01$) mehr fachlich richtig

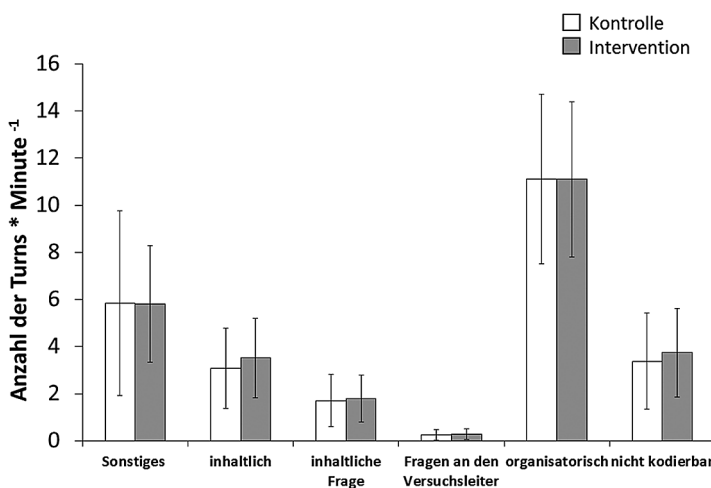


Abb. 9: Anzahl der Turns pro Minute aller Oberkategorien (Mittelwert, SA).

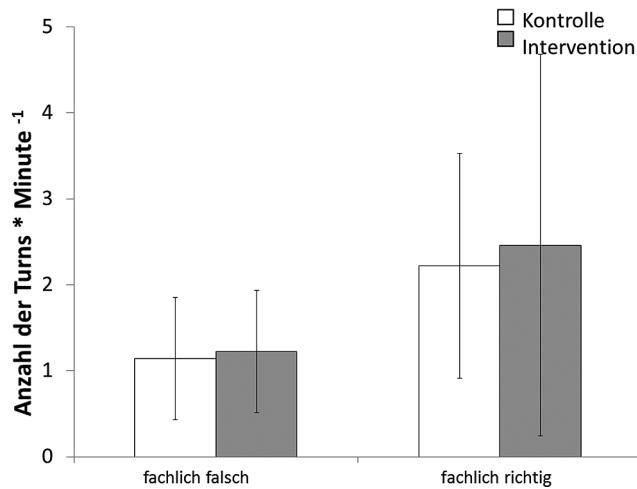


Abb. 10: Anzahl der Turns pro Minute hinsichtlich ihrer Richtigkeit (Mittelwert, SA).

sinngemäß wiedergibt. Für die Kategorie *Vorlesen* zeigen sich mit $d = .42$ sowie für *sinngemäßes Wiedergeben* mit $d = .32$ kleine bis mittlere Effekte (Cohen, 1988). Diese Ergebnisse zeigen einen erwarteten Effekt der Interventionsmaßnahmen.

Neben diesen Ergebnissen zeigt sich, dass die Interventionsgruppe zudem hochsignifikant ($p < .001$) mehr Aussagen zur Regelung der Aufgabenverteilung und mehr Aussagen, die der Aufforderung zur Mitarbeit dienen, tätigt. Die Effektstärken liegen bei $d = .36$ bei der Kategorie *Aufgabenverteilung* sowie $d = .39$ bei der Kategorie *Aufforderung zur Mitarbeit* und sind damit klein (Cohen, 1988).

In den Interventionsgruppen wurden verschiedene Rollen (Forschungsleiter und Forschungsmitarbeiter) verteilt. Dies könnte das häufigere Auftreten der Subkategorie *Aufgabenverteilung* bewirkt haben. Damit kann zudem auch der Unterschied zwischen den beiden Gruppen bei den re-

gulativen Aussagen (*Aufforderung zur Mitarbeit*) begründet werden. Die Zuteilung bestimmter Rollen könnte zu regulativen Aussagen motiviert haben, wenn eine Aufgabe durch das zuständige Gruppenmitglied nicht erfüllt wurde. Dieser Befund deutet dementsprechend darauf hin, dass die Interventionsmaßnahmen angenommen wurden.

Um zu untersuchen, ob der Einsatz der Intervention auch bezogen auf die Qualitätsmerkmale der inhaltlichen Aussagen gegriffen hat, werden diese im Folgenden detailliert in den Blick genommen.

Die Auswertung hinsichtlich des Qualitätsmerkmals der *fachlichen Richtigkeit* der inhaltlichen Aussagen zeigt erwartungsgemäß keine signifikanten Unterschiede zwischen der Kontroll- und Interventionsgruppe, da diese zwar als Qualitätsmerkmal definiert, in der Intervention aber nicht explizit gefördert wurde (siehe Abbildung 10).

Die Auswertung der Videodaten hinsichtlich der *Komplexität* der inhaltlichen Aussagen über alle Stunden hinweg zeigt auf Faktenniveau keine Unterschiede. Allerdings tätigt die Interventionsgruppe höchst signifikant ($p < .001$) mehr Aussagen des höheren Komplexitätsniveaus, dem *Zusammenhangsniveau* (siehe Abbildung 11). Die Effektstärke liegt bei $d = .71$ und ist damit mittel bis groß (Cohen, 1988).

Die Betrachtung des Qualitätsmerkmals der kognitiven Prozesse zeigt, dass die Interventionsgruppe signifikant häufiger ($p < .01$) Aussagen der höheren kognitiven Prozesse (*selegieren*, *organisieren* und *integrieren*) macht. Die Effektstärken sind mit $d = .33$ für *selegieren*, $d = .69$ für *organisieren* und $d = .57$ für *integrieren* klein bis mittel. Bezieht man die beiden Qualitätsmerkmale *Komplexität* und *kognitive Prozesse* inhaltlicher Aussagen in die Auswertung

mit ein, so zeigt sich, dass auf Faktenniveau keine statistisch bedeutsamen Unterschiede vorliegen (siehe Abbildung 12).

Für die Aussagen auf Zusammenhangsniveau zeigen sich für alle kognitiven Prozesse höchst signifikante Unterschiede ($p < .001$) zugunsten der Interventionsgruppe (siehe Abbildung 13).

Die Effektstärken liegen zwischen .55 und .72 und sind damit mittel bis groß. Betrachtet man neben der Komplexität und kognitiven Prozesse auch die Richtigkeit der inhaltlichen Aussagen, so bestätigen sich die hier dargestellten Befunde.

Um die vierte Forschungsfrage beantworten zu können, wurden die Videodaten der fünften Stunde separat hinsichtlich der drei Qualitätsmerkmale für *inhaltliche Aussagen* untersucht. Hierbei sollte der Frage nachgegangen werden, ob eine Steigerung der fachinhaltlichen Qualität

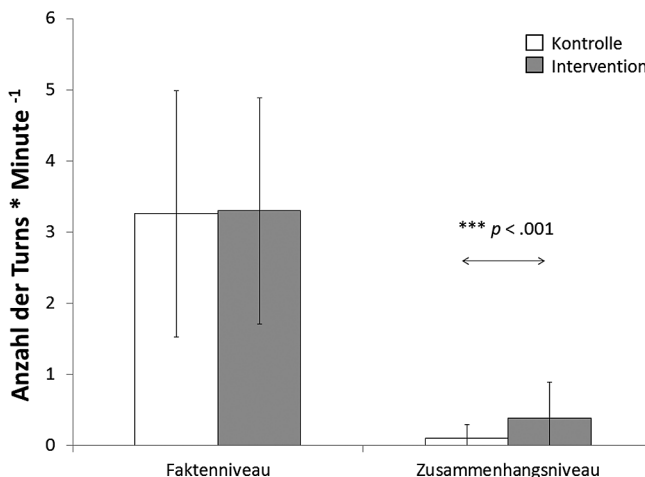


Abb. 11: Anzahl der Turns pro Minute hinsichtlich ihrer Komplexität (Mittelwert, SA).

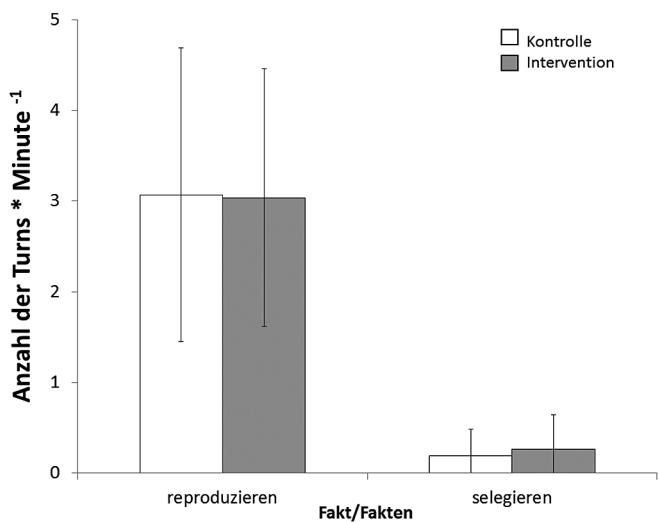


Abb. 12: Anzahl der Turns pro Minute auf *Faktenniveau* unterschieden nach *kognitiven Prozessen* (Mittelwert, SA).

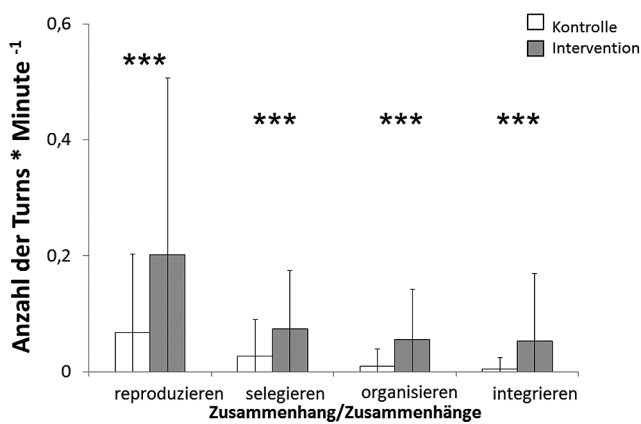


Abb. 13: Anzahl der Turns pro Minute auf *Zusammenhangsniveau* unterschieden nach *kognitiven Prozessen* (Mittelwert, SA).

der Schüleräußerungen aufrechterhalten werden kann, wenn die Anregung dazu weggelassen wird. Analog zu den Ergebnissen über alle Stunden hinweg konnten jeweils bezogen auf die *fachliche Richtigkeit* und die Anzahl der *inhaltlichen Aussagen* auf *Faktenniveau* keine Unterschiede zwischen den Treatmentgruppen gefunden werden. Jedoch zeigen sich auf *Zusammenhangsniveau* auch in der fünften Kleingruppensequenz für alle kognitiven Prozesse signifikante Unterschiede (siehe Abbildung 14). Die Effektstärken sind groß und liegen zwischen $d = .85$ und $d = 1.01$. Das Verhältnis der *inhaltlichen Aussagen* auf *Faktenniveau* zu den Aussagen auf *Zusammenhangsniveau* bleibt, im Vergleich zu der Auswertung über alle Stunden hinweg, gleich.

Insgesamt zeigen die Ergebnisse der Videoanalyse, dass durch die Interventi-

onsmaßnahmen qualitativ hochwertigere Aussagen bezüglich der beiden Qualitätsmerkmale inhaltlicher Aussagen, der *Komplexität* und der *kognitiven Prozesse* erfolgreich gefördert werden konnten. Die Steigerung der hochwertigeren Aussagen hält zudem mindestens noch eine weitere Stunde nach dem Weglassen der Instruktionen an.

Ergebnisse der schriftlichen Tests

In einem weiteren Schritt wurden die schriftlichen Tests insbesondere mit Fokus auf den Lernerfolg ausgewertet, um zu überprüfen, ob die Steigerung der qualitativ hochwertigeren Aussagen auch mit einer Steigerung des Lernerfolgs einhergeht. Zunächst wurde anhand eines Vergleichs der Pre- und Posttestdaten der Gesamtstichprobe mittels eines t-Tests

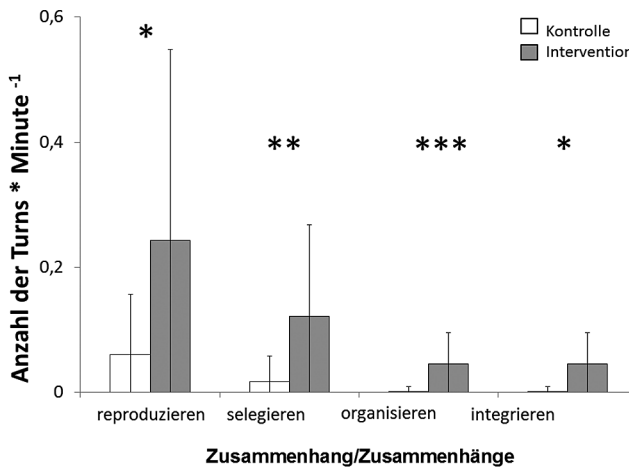


Abb. 14: Anzahl der Turns pro Minute auf *Zusammenhangsniveau* unterschieden nach *kognitiven Prozessen* in der 5. Stunde (Mittelwert, SA).

die Lernwirksamkeit der Unterrichtsreihe überprüft. Die Ergebnisse zeigen einen hochsignifikanten Lernzuwachs sowohl für die Kontrollgruppe als auch für die Interventionsgruppe. Über die Gesamtgruppe zeigt sich dementsprechend auch ein hoch signifikanter Lernzuwachs ($p < .001$) mit einer Effektstärke von $d = 1.95$. Die Lernwirksamkeit der Unterrichtsreihe konnte somit analog zu Walpuski (2006) als nachgewiesen angesehen werden. Jedoch lassen sich an dieser Stelle keine signifikanten Unterschiede zwischen Kontroll- und Interventionsgruppe nachweisen. Dieser Befund ist allerdings nicht unbedingt erwartungswidrig, da der Schwerpunkt der Interventionsmaßnahmen auf dem Verbalisieren von Zusammenhängen lag, während der fachspezifische Leistungstest hauptsächlich Faktenwissen erhebt. Die weitere Auswertung der Leistungstests erfolgte analog zur Reanalyse auf Ebene der Gruppenaggregate. Die Ergebnisse der Studententests, welche ebenfalls hauptsächlich Faktenwissen abfragen, untermauern diesen Befund. Auch hierbei finden sich keine Unterschiede zwischen den Treatmentgruppen. Die Ergebnisse des NAW-Tests von pre nach post bestätigen ebenfalls die Lernwirksamkeit der Unterrichtsreihe bezogen auf die naturwissenschaftlichen Arbeitsweisen ($p = .004$; $d = .21$) (Cohen, 1988). Zwischen den Treatmentgruppen liegen auch hier keine signifikanten Unterschiede im Lernzuwachs vor. Dieser Befund lässt sich damit erklären, dass die Interventionsmaßnahmen keinen Fokus auf naturwissenschaftliche Arbeitsweisen legen. Insgesamt kann jedoch davon

ausgegangen werden, dass die Beschäftigung mit den Interaktionsboxen in beiden Treatmentgruppen implizit die Kenntnisse über naturwissenschaftliche Arbeitsweisen fördern. Der Triadentest wurde nur im Posttest eingesetzt, da dieser fachspezifisches, vernetztes Zusammenhangswissen abprüft, das im Pretest nicht zu erwarten ist. Die Auswertung erfolgte durch einen Vergleich der Punktsummen zwischen den beiden Treatmentgruppen im Posttest. Es wurde der t-Test verwendet, um mögliche Unterschiede auf ihre statistische Bedeutsamkeit hin zu untersuchen. Über alle Triaden konnten signifikante bis höchst signifikante Unterschiede bezüglich der Punkte zwischen den beiden Treatmentgruppen festgestellt werden. Die Mittelwerte der Interventionsgruppe sind für jede Triade signifikant höher als die Mittelwerte der Kontrollgruppe. Der Unterschied zwischen Kontroll- und Interventionsgruppe ist höchst signifikant ($t(186) = 5.375$; $p < .001$) und stellt mit $d = .85$ einen großen Effekt dar. Abbildung 15 stellt die Punktsummen des Triadentests dar.

Auch der Follow-up Test, der 4-5 Monate nach der Durchführung der Kleingruppenarbeit durchgeführt wurde, bestätigt die Befunde. Im fachspezifischen Leistungstest sowie im NAW-Test zeigen sich keine Unterschiede zwischen den Gruppen. Im Triadentest finden sich allerdings immer noch höchst signifikante Unterschiede zugunsten der Interventionsgruppe. An dieser Stelle ist es wichtig, festzuhalten, dass der Triadentest zwar eine gewisse Ähnlichkeit zur Interventionsmaßnahme aufweist, die sich ebenfalls

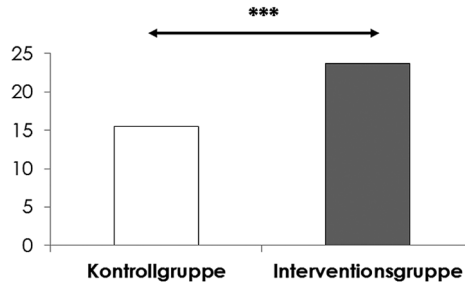


Abb. 15: Mittelwerte der Punktsummen im Triadentest im Posttest.

der Triaden zur Kommunikationsförderung bedient, allerdings wurde sichergestellt, dass in der Intervention verwendete Triaden nicht im Test vorkommen. Zudem stellen Triaden kein anspruchsvolles Testformat dar, sodass diese ohne vorhergehendes Training eingesetzt werden können. Vor der Bearbeitung des Triadentests wurde dennoch jeweils standardisiert der Umgang mit dem Testformat für alle Schülerinnen und Schüler an einem alltagsweltlichen Beispiel demonstriert, so dass ein in der Interventionsmaßnahme begründeter Lernvorsprung der Probanden minimiert werden konnte.

6 Zusammenfassung und Diskussion der Ergebnisse

Zusammenfassend kann für die Ergebnisse der Tests festgestellt werden, dass die Interventionsmaßnahmen insbesondere das komplexere und vernetztere Zusammenhangswissen fördern, und das auch über einen längeren Zeitraum hinweg. Insgesamt lässt sich festhalten, dass

durch die gezielte Anregung zu fachlichen Äußerungen durch die Interventionsmaßnahmen die Qualität und der Umfang der fachlichen Aussagen erhöht werden konnte. Die Steigerung der Qualität der fachinhaltlichen Aussagen ging zudem mit einer Steigerung des fachlichen Lernens – insbesondere des komplexeren Zusammenhangswissens – einher. Die Steigerung der Qualität der fachinhaltlichen Aussagen hält weiter für mindestens eine weitere Stunde an.

Durch die vorliegende Studie war es zum einen möglich die Kleingruppenarbeitsphasen mit Interaktionsboxen zu optimieren. Zum anderen wurde ein Analyseinstrument zu entwickelt, mit dem eine ökonomische Auswertung der Qualität fachinhaltlicher Schüleräußerungen möglich ist. Dieses Instrument kann für weitere Forschungen verwendet werden, wie beispielsweise die Validierung schriftlicher Leistungstests des Kompetenzbereichs Kommunikation. Des Weiteren konnte durch die Arbeit nicht nur ein korrelativer Zusammenhang zwischen der Qualität fachlicher Aussagen und dem

Lernerfolg bestätigt werden, sondern auch die Wirkrichtung bestimmt werden. So konnte gezeigt werden, dass eine Optimierung der Qualität fachinhaltlicher Schüleräußerungen mit einer Steigerung des Lernerfolgs einhergeht.

Dieser Befund ist in großem Maße für die Unterrichtspraxis relevant. Häufig wird bemängelt, dass im Unterricht lediglich isoliertes Faktenwissen gelernt wird und nur selten dieses Wissen vernetzt wird (z. B. Glemnitz, 2007). In der vorliegenden Studie konnte gerade dieses vernetzte Wissen gefördert werden. Dementsprechend kann eine Implementation der Ergebnisse in der Schulpraxis zum einen zu einer Optimierung von kooperativer Kleingruppenarbeit zum Thema Säuren und Basen führen, und zum anderen damit verbunden auch zu einer Förderung des vernetzten Wissens der Schülerinnen und Schüler in diesem Themengebiet. Darüber hinaus können die Ergebnisse der vorliegenden Studie als Anlass zu weiteren Forschungsvorhaben dienen. So sollte die Zusammensetzung der Gruppen näher in den Blick genommen werden. Hierbei wäre beispielsweise eine Analyse heterogen und homogen zusammengesetzter Gruppen erstrebenswert. Des Weiteren könnte eine Analyse der Daten mittels einer Mehrebenenanalyse erfolgen, welche die Abhängigkeit der Schülerinnen und Schüler innerhalb der Kleingruppen näher in den Fokus nimmt.

Danksagung

Wir danken der DFG für die Förderung des Projektes WA2829/1-1 und den beteiligten Lehrerinnen und Lehrern sowie Schülerinnen und Schülern für die Bereitschaft zur Mitarbeit.

Literatur

- Abelson, R. P. & Levi, A. (1985). Decision making and decision theory. In Lindzey, G. & Aronson, E. (Hrsg.), *The Handbook of Social Psychology* (S. 231–309). New York: Random House.
- Aufschnaiter, C. von, Erduran, S., Osborne, J. & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*, Heidelberg: Springer.
- Campbell, D. T. & Stanley J. C. (1963). Experimental And Quasi-Experimental Designs For Research on Teaching. In N. Gage (Hrsg.), *Handbook for research on teaching* (S. 171–246). Chicago: Rand McNally.
- Chi, M. T., Leeuw, N. de, Chiu, M.-H. & LaVancher, C. (1994). Eliciting Self-explanations Improves Understanding. *Cognitive Science*, 18, 439–477.
- Chi, M. T. & Bassok, M. (1989). Learning from examples via self-explanations. In R. Glaser & L.B. Resnick (Hrsg.), *Knowing, learning, and instruction. Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*, Hillsdale, NJ: Erlbaum.
- Driver, R., Newton, P. & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312.

- Erduran, S., Simon, S. & Osborne, J. (2004). TAPing into argumentation: science discourse. *Science Education*, 88(6), 915–933.
- Ferguson-Hessler, M.G. & Jong, T. de (1990). Studying Physics Texts: Differences in study processes between good and poor performers. *Cognition and Instruction*, 7(1), 41–54.
- Franke-Braun, G. (2008). *Aufgaben mit gestuften Lernhilfen. Ein Aufgabenformat zur Förderung der sachbezogenen Kommunikation und Lernleistung für den naturwissenschaftlichen Unterricht*, Berlin: Logos.
- Gijlers, H. & Jong, T. de (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching*, 42(3), 264–282.
- Hänze, M. & Berger, R. (2007). Kooperatives Lernen im Gruppenpuzzle und im Lernzirkel. *Unterrichtswissenschaft*, 35(3), 227–240.
- Heller, K.A. & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4. bis 12. Klassen. Revision, Göttingen.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*, Berlin: Logos.
- Klos, S., Henke, C., Kieren, C., Walpuski, M. & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik*, 54, 304–321.
- Knobloch, R., Sumfleth, E. & Walpuski, M. (2011). Analyse der Schüler-Schüler-Kommunikation im Chemieunterricht Entwicklung und Erprobung eines Kategoriensystems. *CHEMKON*, 18 (2), 65–70.
- Kramer, G. (2009). Entwicklung und Überprüfung eines Strukturmodells der fachlichen Kommunikationskompetenz im Biologieunterricht. <http://d-nb.info/1019954035/34> [zuletzt abgerufen am 25.10.2013].
- Kuhn, D. & Reiser, B. J. (2005). *Students constructing and defending evidence-based scientific explanations*, Dallas, TX.
- Lizotte, D. J., McNeill, K. L. & Krajcik, J. (2004). Teacher practices that support students' construction of scientific explanations in middle school classrooms. In Y. Kafai, W. Sandoval, N. Enyedy, & A. Nixon (Hrsg.), *Proceedings of the 6th International Conference of the Learning Sciences*. Norwood N.J: Erlbaum.
- Means, M. & Voss, J. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction*, 1, 337–350.
- Neuroth, J. (2002). Triaden-Test als Methode zur Ermittlung des Verknüpfungsgrades des Schülerwissens. Eine Verfahrensmodifikation zum Verknüpfungstest. Schriftliche Hausarbeit im Rahmen der ersten Staatsprüfung für das Lehramt für die Sekundarstufen I und II, Universität Essen.
- Osborne, J., Erduran, S. & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020.
- Renkl, A. (1997). Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science*, 21(1), 1–29.
- Rimmele, R. (2004). Videograph – Multimedia-Player zur Kodierung von Videos, Kiel: IPN – Leibnitz Institut für die Pädagogik der Naturwissenschaften.
- Ropohl, M. (2010). *Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion. Entwicklung und Analyse von Testaufgaben*, Berlin: Logos.
- Rotering-Steinberg, S. (1995). Kooperative Formen des Lehrens und Lernens in der Erwachsenenbildung. *Unterrichtswissenschaft*, 23(4), 332–346.
- Rumann, S. (2005). *Kooperatives Experimentieren im Chemieunterricht. Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik*, Berlin: Logos.
- Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, 41(5), 513–536.

- Sangin, M., Molinari, G., Dillenbourg, P., Rebetez, C. & Bétrancourt, M. (2006). Collaborative learning with animated pictures: The role of verbalizations. *International Conference on Learning Science*.
- Schulz, A. (2011). *Experimentierspezifische Qualitätsmerkmale im Chemieunterricht. Eine Videostudie*, Berlin: Logos.
- Solomon, J. (1988). What is DISS? Discussion of issues in school science. *Education in Science*(129), 18–19.
- Sumfleth, E. & Pitton, A. (1998). Sprachliche Kommunikation im Chemieunterricht: Schülervorstellungen und ihre Bedeutung im Unterrichtsalltag. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(2), 4–20.
- Sumfleth, E., Neuroth, J. & Leutner, D. (2010). Concept Mapping - eine Lernstrategie muss man lernen. *CHEMKON*, 17(2), 66–70.
- Toulmin, S. (1958). *The uses of argument*, Cambridge: Cambridge University Press.
- Walpuski, M., Kauertz, A., Kampa, N., Fischer, H.E., Mayer, J., Sumfleth, E. & Wellnitz, N. (2010). ESNaS - Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe. In A. Gehrmann (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht*. Bad Heilbrunn: Klinkhardt.
- Walpuski, M. (2006). *Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback. Eine empirische Studie*, Berlin: Logos.
- Walpuski, M. & Sumfleth, E. (2007). Strukturierungshilfen und Feedback zur Unterstützung experimenteller Kleingruppenarbeit im Chemieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 13.
- Wirtz, M. A. & Caspar, F. (2007). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*, Göttingen: Hogrefe.
- Zohar, A. & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1), 35–62.

KONTAKT

Prof. Dr. Maik Walpuski
Schützenbahn 70
45127 Essen
maik.walpuski@uni-due.de

AUTORENINFORMATION

Dr. Rebecca Knobloch arbeitet als Lehrerin für die Fächer Chemie und Biologie an einem Warendorfer Gymnasium. Zuvor hat sie in der Didaktik der Chemie promoviert und war Kollegiatin im Graduiertenkolleg Naturwissenschaftlicher Unterricht an der Universität Duisburg-Essen. Sie studierte die Fächer Chemie und Biologie an der Universität Duisburg-Essen und legte 2008 das 1. Staatsexamen und 2013 das 2. Staatsexamen für das Lehramt an Gymnasien und Gesamtschulen ab.

Dr. Elke Sumfleth ist Professorin für Didaktik der Chemie und war Sprecherin des DFG-Graduiertenkollegs „Naturwissenschaftlicher Unterricht“ an der Universität Duisburg-Essen. Ihre Forschungsschwerpunkte sind Empirische Lehr-Lern-Forschung in Chemie mit unterschiedlichen Schwerpunkten und Entwicklung und Evaluation von Materialien für den Chemieunterricht.

Dr. Maik Walpuski ist Professor für Didaktik der Chemie an der Universität Duisburg-Essen. Seine Forschungsschwerpunkte sind die Kompetenzmodellierung und Kompetenzmessung für das Fach Chemie sowie die prozessanalytische Untersuchung des Chemieunterrichts, insbesondere in Experimentierphasen.