

## Ergebnisse einer empirischen Studie zum Elektronengasmodell

Jan-Philipp Burde\*, Thomas Wilhelm\*

\*Institut für Didaktik der Physik, Goethe-Universität Frankfurt, Max-von-Laue-Str. 1, 60438 Frankfurt  
[burde@physik.uni-frankfurt.de](mailto:burde@physik.uni-frankfurt.de), [wilhelm@physik.uni-frankfurt.de](mailto:wilhelm@physik.uni-frankfurt.de)

### Kurzfassung

Das Ziel des Elektronengasmodells besteht darin, den Lernenden eine anschauliche Vorstellung des Potenzial- und Spannungsbegriffs zu vermitteln, da sonst ein angemessenes Verständnis elektrischer Stromkreise nur schwer zu erreichen ist. Ein auf dem Elektronengasmodell basierendes Unterrichtskonzept wurde im Rahmen einer mit 790 Schülern durchgeführten Studie im Schuljahr 2015/16 im Frankfurter Raum empirisch evaluiert. Die Studie folgte einem Prätest-Posttest-Treatment-Kontrollgruppen-Design und verwendete als Testinstrument einen psychometrisch ausgereiften zweistufigen Multiple-Choice-Test, der auch die Erhebung von typischen Fehlvorstellungen erlaubt. Die Ergebnisse der Studie sind äußerst vielversprechend: Die nach dem neuen Konzept unterrichteten Schüler entwickeln nicht nur ein deutlich besseres konzeptionelles Verständnis, sondern hatten verglichen mit ihren traditionell unterrichteten Mitschülern auch einen mehr als doppelt so hohen Lernzuwachs. Die Idee hinter dem Unterrichtskonzept sowie die grundlegenden Ergebnisse der empirischen Studie werden im Folgenden genauer vorgestellt.

### 1. Motivation

Viele Lernende neigen dazu, Stromkreise ausschließlich mit Hilfe des Strom- und Widerstandsbegriffs zu analysieren, was in der Regel mit diversen Fehlvorstellungen einhergeht. Der Grund hierfür ist, dass sie im Unterricht der Sekundarstufe I häufig keinen konzeptionell adäquaten Spannungsbegriff entwickeln, ohne den ein angemessenes Verständnis elektrischer Stromkreise jedoch nur schwer vorstellbar ist. Stattdessen ist die Vorstellung vieler Schülerinnen und Schüler von Stromkreisen in der Sek I maßgeblich vom Strombegriff geprägt, der nicht aus didaktischen sondern historischen Gründen den Unterricht in der Mittelstufe dominiert [1, S.62]. In der Folge wird die elektrische Spannung von vielen lediglich als Eigenschaft des Stroms wahrgenommen (Stichwort „Stromspannung“) [2]. Dies ist nicht nur bedauerlich, weil die Spannung eine der physikalischen Größen darstellt, welche im Alltag der Schüler eine große Bedeutung haben, sondern auch, weil ohne die Spannung ein elementares Verständnis der Elektrizitätslehre unmöglich ist. Die Spannung ist allerdings keine einfache physikalische Größe, da sie die Differenz zweier Potenzialwerte darstellt und sich immer auf zwei Punkte in einem Stromkreis bezieht. Sie ist damit komplexer als das elektrische Potenzial, das einem Punkt bzw. einem Leiterabschnitt zugeordnet werden kann. Paradoxerweise wird aber von den Schülern insbesondere im Anfangsunterricht häufig erwartet, ein Verständnis für die Spannung zu entwickeln, ohne die dahinterstehende Größe selbst, nämlich das elektrische Potenzial, zu kennen, geschweige denn zu verstehen [3, S.477]. Das Ziel des Elektronengasmodells ist es

daher, den Lernenden neben einem qualitativen Verständnis der Wirkungszusammenhänge in Stromkreisen insbesondere eine anschauliche Vorstellung des elektrischen Potenzial- und Spannungsbegriffs bereits im Anfangsunterricht zu ermöglichen.

### 2. Die Grundidee des Elektronengasmodells

Die Grundidee des Elektronengasmodells besteht darin, dass sich in Metallen wie z.B. Kupfer die Elektronen in Teilchenform befinden und sich dort ähnlich einem Gas verhalten. In einem einfachen Stromkreis bestehend aus einer Batterie und einem Lämpchen sorgt eine Batterie im Elektronengasmodell nun für eine Ungleichverteilung der Elektronen im Kupferleiter. Konkret wird sie im didaktisch vereinfachten Modell als Elektronenpumpe betrachtet, die Elektronen aus dem mit dem Pluspol verbundenen Leiter in den mit dem Minuspol verbundenen Leiter pumpt. Entsprechend des Modells sinkt damit die Elektronendichte im mit dem Pluspol verbundenen Leiterstück, während sie im mit dem Minuspol verbundenen Leiterstück ansteigt. Im Leiter kommt es aufgrund der gegenseitigen Coulomb-Abstoßung der Elektronen zu einem von der Elektronendichte abhängigen „elektrischen Druck“. Im mit dem Minuspol verbundenen Leiterstück entsteht also ein hoher elektrischer Druck, wohingegen der elektrische Druck im Leiterstück, das mit dem Pluspol verbunden ist, absinkt. Durch Gleichsetzen des „elektrischen Drucks“ mit dem elektrischen Potenzial kann die elektrische Spannung im Elektronengasmodell so als elektrischer Druckunterschied interpretiert werden.

An dieser Stelle könnte man einwenden, dass die Dichte der Leitungselektronen in einem Stromkreis konstant ist und die Leitungselektronen damit kein kompressibles Fluid darstellen. Tatsächlich sind die Elektronen in Leitern aber in Folge von Oberflächen- und Grenzladungen nicht völlig homogen über den gesamten Stromkreis verteilt (für eine fachliche Auseinandersetzung sei auf [4] verwiesen). An einem Widerstand entsteht beispielsweise an der einen Grenzfläche mit dem Leiter eine Elektronenanreicherung, während es an der anderen Grenzfläche zu einem Elektronenmangel kommt [5, S.767]. Das hier vorgeschlagene Elektronengasmodell stellt also eine didaktische Elementarisierung der unterschiedlichen Ladungsdichten in Folge von Grenz- und Oberflächenladungen dar, indem lediglich die mittlere Elektronendichte innerhalb eines Leiterabschnitts betrachtet und auf eine Unterscheidung zwischen Leitungselektronen auf der einen Seite und Oberflächen- bzw. Grenzflächenelektronen auf der anderen Seite verzichtet wird.

Eine der zentralen didaktischen Ideen des hier vorgestellten Unterrichtskonzepts besteht darin, in Anlehnung an den Ansatz von diSessa [6] an geeignetes Vorwissen der Schüler anzuknüpfen. Konkret bringen diese aus dem Alltag in der Regel bereits ein intuitives Luftdruckverständnis im Sinne von „komprimierte Luft steht unter Druck, drückt gegen die Wände und hat das Bestreben sich auszudehnen“ mit, das in der Terminologie von diSessa einen in Hinblick auf die Begriffsentwicklung des elektrischen Potentials anknüpfungsfähigen „p-prim“ darstellt. Wichtig ist an dieser Stelle anzumerken, dass dieser p-prim („phenomenological primitive“) der Theorie von diSessa entsprechend keinem ausdifferenzierten physikalischen Begriffsverständnis entspricht, sondern lediglich ein isoliertes, auf Alltagserfahrungen basierendes, kognitiv noch nicht weiter vernetztes Wissenselement darstellt. Konkret bedeutet dies, dass im Unterrichtskonzept versucht wird, den Schülern eine intuitive Vorstellung vom elektrischen Potenzial zu ermöglichen, indem an ihre alltägliche Luftdruckvorstellung angeknüpft wird, ohne jedoch ein fachlich korrektes Druckkonzept mit der in der Sek I teils schwierigen Differenzierung zwischen skalarem Druckbegriff und vektoriellem Kraftbegriff vorauszusetzen oder dieses an dieser Stelle einzuführen. Das Ziel der unterrichtlichen Bemühungen besteht darin, diesen zunächst isolierten p-prim in eine kognitive Struktur einzubetten, indem er mit weiteren p-prim wie „ein Druckunterschied führt zu einer Strömung“ verknüpft wird. Im Unterrichtskonzept wird daher ausgehend von Alltagsobjekten wie z.B. einem aufgeblasenen Fahrradreifen oder einer aufgeblasenen Luftmatratze erarbeitet, dass Luft immer von Bereichen höheren Drucks zu Bereichen niedrigeren Drucks strömt. Das Ziel dabei ist die Erkenntnis, dass Luftströmungen immer eine Folge von Druckunterschieden sind und dass zwischen den beiden Größen Druck und Druck-

unterschied konzeptionell unterschieden werden muss.

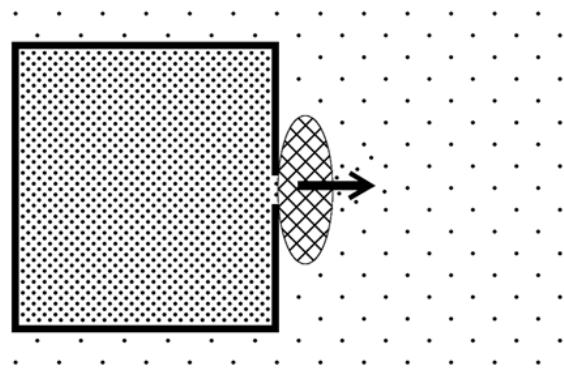


Abb. 1: Ein Luftdruckunterschied bewirkt eine Luftströmung durch einen Stoffwiderstand.

Zusätzlich findet auch eine Hinführung zur Widerstandsvorstellung statt, indem die Schülerinnen und Schüler ein Stück Stoff (z.B. ein Schal, Kragen oder Ärmel) nehmen und Luft durch dieses blasen. Je dicker das Stück Stoff dabei gefaltet wird, desto stärker ist die Hemmung bzw. Behinderung der Luftströmung (siehe Abbildung 1). Die Hemmung bzw. Behinderung der Luftströmung durch den Stoff stellt dabei einen p-prim für den elektrischen Widerstand dar und wird in Vorgriff auf die Betrachtung von Stromkreisen bereits hier als Widerstand (für die Luftströmung) bezeichnet.

Wie in Abbildung 2 dargestellt, wird der in den Leitern herrschende elektrische Druck in Anlehnung an die Luftdruckbeispiele über eine Teilchendichtedarstellung eingeführt, im weiteren Verlauf dann aber mit Hilfe eines an die alltägliche Konvention angelehnten Farbschemas direkt im Schaltplan visualisiert (rot = hoher Druck, blau = tiefer Druck). Durch Gleichsetzen des elektrischen Drucks mit dem elektrischen Potenzial kann die Spannung dann in Analogie zu Luftdruckunterschieden als elektrischer Druckunterschied und Antrieb des elektrischen Stroms eingeführt werden. Eine ausführliche Darstellung des Unterrichtskonzepts mit vielen Illustrationen und Erklärungen findet sich in [7].

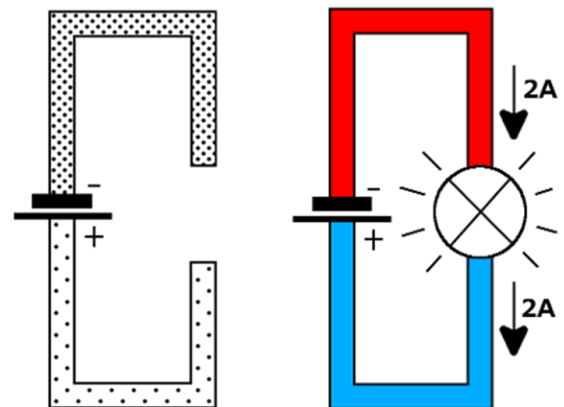


Abb. 2: Punktedichtedarstellung (links) und Farbdarstellung (rechts) des elektrischen Drucks

Gegenüber dem Modell des geschlossenen Wasserkreislaufs besteht der Vorteil des Elektronengasmodells insbesondere darin, dass die Lernenden aus dem Alltag bereits eine anschlussfähige Luftdruckvorstellung besitzen, wohingegen Wasser von den Schülern in der Regel als ein inkompressibles Kontinuum wahrgenommen wird. Damit ist für sie der Wasserdruck im geschlossenen Wasserkreislauf jedoch anschaulich nicht vorstellbar, weil sich in diesem Bild Wasser unter hohem Druck in nichts Sichtbarem von Wasser unter geringem Druck unterscheidet. Das Elektronengas hingegen stellt ein kompressibles Fluid aus Einzelteilchen dar, weshalb der Druck proportional zur Teilchendichte ist und somit für die Lernenden besser vorstellbar und visualisierbar ist [8].

Interessierte Lehrkräfte können die im Rahmen dieser Studie entwickelten Unterrichtsmaterialien unkompliziert und kostenfrei auf der Seite [www.einfache-elehre.de](http://www.einfache-elehre.de) herunterladen.

### 3. Das Design der Studie

Im Sinne von Design-Based-Research [9, 10] wurde zunächst unter Berücksichtigung früherer Studien und erforschter Schülervorstellungen ein Grundkonzept zum Elektronengasmodell erarbeitet und mit Hilfe von Teaching Experiments darauf hin überprüft, inwiefern die verwendeten Konzepte und Visualisierungen von den Schülerinnen und Schülern akzeptiert und verstanden werden [11]. Aufbauend auf den so gewonnenen Erkenntnissen wurde dann ein Unterrichtskonzept nach dem Elektronengasmodell inklusive passender Unterrichtsmaterialien wie Overheadfolien und Übungen für die Schulpraxis entwickelt. Im Rahmen einer mit 790 Schülern bzw. 36 Gymnasialschulklassen durchgeführten Studie wurde dieses Unterrichtskonzept dann im Schuljahr 2015/16 im Frankfurter Raum empirisch evaluiert.

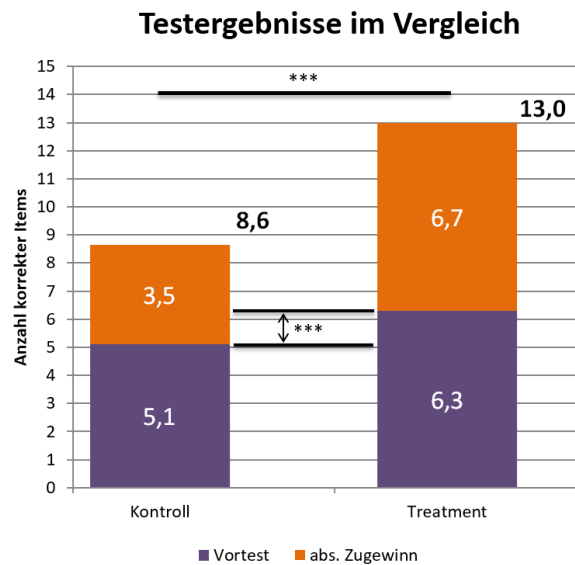
Ziel der quasi-experimentellen Studie auf Basis eines Prätest-Posttest-Treatment-Kontrollgruppen-Design war es, den Verständniszuwachs von traditionell unterrichteten Klassen (Kontrollgruppe) mit dem Verständniszuwachs von Klassen zu vergleichen, die nach dem neuen Unterrichtskonzept auf Grundlage des Elektronengasmodells unterrichtet wurden (Treatmentgruppe). Der Verständniszuwachs wurde dabei in beiden Gruppen mit Hilfe eines psychometrisch ausgereiften zweistufigen Multiple-Choice-Tests erhoben [12], der 22 Items zu Stromstärke und Widerstand umfasst und um vier weitere Items zum Potenzial- und Spannungsbegriff erweitert wurde. Der Vorteil der Zweistufigkeit des Tests besteht darin, dass die Schüler nicht nur eine Antwort ankreuzen, sondern in einer zweiten Stufe diese dann auch begründen müssen. Auf diese Weise können nicht nur typische Fehlvorstellungen erhoben werden, sondern auch falsch-positive Antworten (d.h. richtige Antworten mit falscher Begründung bzw. Vorstellung) identifiziert werden. Nur wenn

sowohl die gegebene Antwort als auch die Begründung korrekt sind, wird das jeweilige Item als korrekt gezählt, weshalb im eingesetzten Multiple-Choice-Test der maximal erreichbare Summenscore 26 beträgt.

### 4. Empirische Ergebnisse

Die Kontrollgruppe umfasste 17 Gymnasialschulklassen bzw. 357 Schülerinnen und Schüler, die von 11 Lehrkräften über durchschnittlich 23,5 Schulstunden unterrichtet wurden. Die etwas größere Treatmentgruppe umfasste 19 Gymnasialschulklassen bzw. 433 Schülerinnen und Schüler, die von 14 Lehrkräften über durchschnittlich 24,3 Schulstunden unterrichtet wurden. Die unterrichteten Klassen verteilten sich in etwa gleichmäßig auf Jahrgangsstufe 7 und Jahrgangsstufe 8, wobei das Thema „Elektrizitätslehre“ bei allen teilnehmenden Klassen zum ersten Mal im Unterricht behandelt wurde. Die beiden Gruppen sind also in Hinblick auf die Gruppengröße, die Jahrgangsstufe und die unterrichtete Stundenzahl miteinander vergleichbar.

Nach dem Elektrizitätslehreunterricht konnten die Schülerinnen und Schüler der Treatmentgruppe mit 13,0 Items im Schnitt höchst signifikant mehr Items korrekt beantworten als die Schüler der Kontrollgruppe, die lediglich 8,6 Items korrekt beantworten konnten.



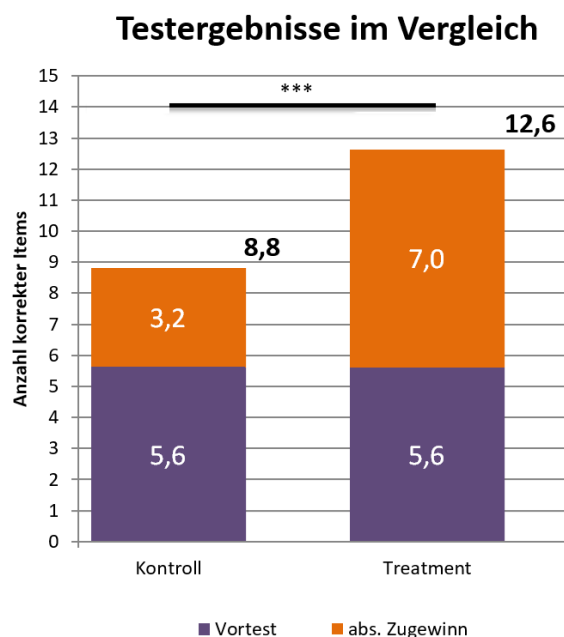
**Abb. 3:** Testergebnisse von Kontroll- und Treatmentgruppe im Vergleich

Wie in Abbildung 3 zu sehen, unterscheiden sich die beiden Gruppen allerdings auch höchst signifikant hinsichtlich ihres Vortestergebnisses. Vor Beginn des Elektrizitätslehreunterrichts konnten die Schüler der Kontrollgruppe durchschnittlich 5,1 Items und die Schüler der Treatmentgruppe durchschnittlich 6,3 Items korrekt beantworten. Aufgrund dieses signifikanten Unterschieds bereits vor der Durchführung der Unterrichtseinheit bietet es sich an, die Differenz der Veränderung in der Experimental- und der Kontrollgruppe zu berechnen, um auf diese Wei-

se den „Nettoeffekt“ des Unterrichts zu ermitteln [13, S.559]. Hier zeigt sich, dass der durchschnittlich erzielte absolute Lernzuwachs in der Treatmentgruppe mit 6,7 Items fast doppelt so hoch ausfällt wie in der Kontrollgruppe, wo er lediglich 3,5 Items beträgt. Auch hier ist der Unterschied zwischen den beiden Gruppen höchst signifikant.

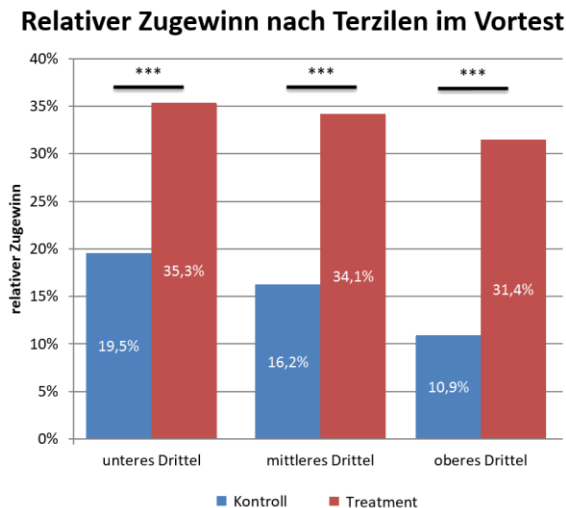
Nun könnte man einwenden, dass der höhere absolute Lernzuwachs in der Treatmentgruppe auf das höhere Vorwissen der Schüler zurückzuführen ist („Matthäus-Effekt“) oder umgekehrt, dass Schüler mit hohem Vorwissen nicht mehr so viel zulegen können. Eine einfache Möglichkeit dies zu untersuchen, stellt die Parallelisierung der beiden Stichproben bezogen auf das Vortestergebnis dar, indem sogenannte „matched samples“ gebildet werden. Da die Bildung eines „matched samples“ einerseits zwangsläufig mit der gezielten Entfernung von Schülern aus der Stichprobe einhergeht, andererseits aber die grundlegende Struktur der Stichprobe soweit wie möglich erhalten bleiben sollte, wurde sich für ein Matching auf Klassenebene entschieden. Bei diesem Verfahren wird so lange aus jeder Klasse der Kontrollgruppe der Schüler mit dem schlechtesten Vortestergebnis und aus jeder Klasse der Treatmentgruppe der Schüler mit dem besten Vortestergebnis entfernt, bis die beiden Gruppen in Bezug auf das Vortestergebnis möglichst genau übereinstimmen. Auf diese Weise kann der Einfluss des Vorwissens als Störvariable (auch „Confounder“ genannt) auf das Nachtestergebnis weitgehend ausgeschlossen werden [13, S.527].

Wie in Abbildung 4 zu sehen ist, beträgt das durchschnittliche Vortestergebnis der beiden Gruppen im matched sample nun 5,6 Punkte. Im Nachtest konnten die Schüler der Treatmentgruppe mit 12,6 Items höchst signifikant mehr Items korrekt beantworten als die Schüler der Kontrollgruppe, die lediglich 8,8 Items korrekt beantworten konnten. Der absolute Lernzuwachs im matched sample fällt damit in der Treatmentgruppe mit 7,0 Items mehr als doppelt so hoch aus wie in der Kontrollgruppe, wo er lediglich 3,2 Items beträgt. Der Unterschied im durchschnittlichen Lernzuwachs zwischen den beiden Gruppen ist dabei höchst signifikant. Da die beiden Gruppen nun in Bezug auf das Vortestergebnis parallelisiert sind, konnte gezeigt werden, dass der höhere absolute Lernzuwachs in der Treatmentgruppe nicht mit dem Matthäus-Effekt erklärt werden kann, also nicht auf das unterschiedliche Vortestergebnis zurückzuführen ist.



**Abb. 4:** Testergebnisse des „matched sample“ von Kontroll- und Treatmentgruppe im Vergleich

Eine weitere interessante Fragestellung besteht darin, wie der im Unterricht erzielte Lernzuwachs vom Vorwissen der Schüler abhängt. Hierzu wurden die Schüler der Kontroll- und Treatmentgruppe in Abhängigkeit ihres Vortestergebnisses zunächst in die Terzile „unteres Drittel“, „mittleres Drittel“ und „oberes Drittel“ eingeteilt. Da ein Schüler mit einem hohen Vortestergebnis prinzipiell keinen so hohen absoluten Lernzugewinn im Test mehr erzielen kann wie ein Schüler mit einem niedrigen Vortestergebnis, wird im Folgenden der durch den Unterricht erzielte relative Zugewinn betrachtet. Der relative Zugewinn gibt dabei an, wie groß der absolute Lernzugewinn in Relation zum für den Schüler überhaupt erzielbaren Zugewinn ausfällt (relativer Zugewinn = absoluter Zugewinn dividiert durch den möglichen Zugewinn). Wie aus Abbildung 5 hervorgeht, erzielen Schüler, die nach dem neuen Unterrichtskonzept unterrichtet wurden, unabhängig vom Vortestergebnis einen deutlich höheren relativen Zugewinn als die traditionell unterrichteten Schüler. Der Unterschied zwischen den beiden Gruppen ist dabei für jedes Terzil höchst signifikant und fällt im oberen Drittel besonders hoch aus. Dies ist darauf zurückzuführen, dass der relative Lernzuwachs in der Kontrollgruppe umso geringer ausfällt, desto besser die Schülerinnen und Schüler im Vortest abgeschnitten haben, wohingegen der relative Lernzuwachs in der Treatmentgruppe unabhängig vom Vortestergebnis nahezu konstant hoch ausfällt.



**Abb. 5:** Relativer Zugewinn nach Terzilen im Vortest (Kontrollgruppe: blau; Treatmentgruppe: rot)

Ein Vergleich des von den jeweiligen Terzilen einer Gruppe erzielten relativen Lernzuwachses mit Hilfe einer Varianzanalyse (ANOVA) ergibt, dass sich die einzelnen Terzile der Treatmentgruppe nicht signifikant voneinander unterscheiden, wohingegen der relative Lernzuwachs des dritten Terzils in der Kontrollgruppe signifikant geringer ausfällt als im ersten Terzil. Im direkten Vergleich zum traditionellen Unterricht scheinen daher insbesondere Schülerinnen und Schüler mit hohem Vorwissen von dem neuen Unterrichtskonzept zu profitieren. Anzumerken ist auch hier, dass der sogenannte „Matthäus-Effekt“ nicht zu beobachten ist: Ein gutes Abschneiden im Vortest führt sowohl relativ als auch absolut gesehen nicht zu einem höheren Lernzuwachs.

Mit Hilfe des zweistufigen Multiple-Choice-Tests nach Urban-Woldron konnten ferner auch die Fehlvorstellungen der Schülerinnen und Schüler erhoben werden. Ein Vergleich der bei den Schülerinnen und Schülern nach dem jeweiligen Unterricht noch vorhandenen Fehlvorstellungen zeigt, dass in der Treatmentgruppe 8 von 11 Fehlvorstellungen signifikant geringer ausgeprägt sind als in der Kontrollgruppe und bei den restlichen drei Fehlvorstellungen kein signifikanter Unterschied feststellbar ist. Im direkten Vergleich zum traditionellen Unterricht scheint das neue Unterrichtskonzept auf Basis des Elektronengasmodells insbesondere zu einem besseren Verständnis für den Differenzcharakter der elektrischen Spannung zu führen und der Stromverbrauchsvorstellung erfolgreich entgegen zu wirken.

## 5. Ausblick

Zur bisherigen Auswertung der Daten der Studie wurde auf die konventionellen statistischen Verfahren des t-Tests und der Varianzanalyse zurückgegriffen. Eine zentrale Voraussetzung der diesen Verfahren zugrunde liegenden t-Statistik bzw. F-Statistik ist jedoch, dass die einzelnen Messungen statistisch gesehen unabhängig voneinander sind. Diese geforderte statistische Unabhängigkeit ist aber bei fachdi-

daktischer Feldforschung strenggenommen nicht gegeben, da die Schüler im realen Unterricht gemeinsam in Klassen lernen und sich somit beispielsweise bzgl. ihres Lernerfolges in Folge bestimmter gemeinsamer Einflüsse (wie z.B. dem gemeinsamen Unterricht, dem gleichen Lernklima und anderen sozialen Interaktionen innerhalb einer Klasse) überzufällig ähnlich sind [14, S.297f]. Darüber hinaus ist davon auszugehen, dass auch der Lernerfolg der verschiedenen Klassen nicht völlig unabhängig voneinander ist, sondern u.a. von der unterrichtenden Lehrkraft abhängt, weshalb man von einer hierarchischen Datenstrukturierung ausgehen muss. Eine für die statistische Interpretation der Daten bedeutende Konsequenz einer solchen hierarchischen Strukturierung der Daten besteht darin, dass ein zusätzlicher Schüler einer bestimmten Klasse aufgrund seiner statistischen Abhängigkeit zu seinen Mitschülern nun nicht einen vollwertigen neuen Informationsbeitrag von 100% zur statistischen Schätzung liefert, sondern nur einen verminderten Informationsbeitrag von beispielsweise 80% [15, S.33; 16, S.9]. Um wie viel der Informationsbeitrag vermindert ist, hängt von der sogenannten Intraklassenkorrelation ab. Sie wird mit Hilfe des Intraklassenkoeffizienten gemessen, der hier z.B. als das Verhältnis der Varianz zwischen den Klassen zur Gesamtvarianz definiert ist und umso größer ist, desto stärker die Schüler innerhalb einer Klasse korreliert sind [15, S.14f]. Da in Folge der Intraklassenkorrelation nicht alle Schüler einen vollständigen Informationsbeitrag liefern, fällt der effektive Stichprobenumfang entsprechend geringer aus, was aber aufgrund der Abhängigkeit des Standardfehlers vom Stichprobenumfang auch mit einer höheren Unsicherheit bzgl. des gefundenen Effekts einhergeht. Wird die hierarchische Datenstrukturierung wie bei t-Tests oder Varianzanalysen vernachlässigt, besteht deshalb die Gefahr, gewisse Effekte für statistisch signifikant zu halten, obwohl sie es tatsächlich nicht sind, womit man einen Fehler 1. Art begehen würde [14, S.298].

Ein vor diesem Hintergrund sehr interessantes statistisches Verfahren stellt die Mehrebenenanalyse dar, da diese nicht nur die hierarchische Datenstrukturierung adäquat berücksichtigt, sondern darüber hinaus auch die Klärung komplexer inhaltlicher Fragen erlaubt, die mit konventionellen varianzanalytischen Methoden wie z.B. einer ANCOVA nicht zu beantworten sind [14, S.309f; 16, S.4]. Die Grundidee der Mehrebenenanalyse im Kontext der Unterrichtsforschung besteht insbesondere darin, dass zwischen der Individualebene der Schüler, der Kontextebene der Klassen und der Lehrerebene unterschieden wird („Drei-Ebenen-Design“). Dabei ähnelt die Mehrebenenanalyse prinzipiell einer linearen bzw. multiplen Regression, unterscheidet sich von dieser jedoch dadurch, dass für jede Ebene eine eigene Modellgleichung formuliert wird und die Regressionsparameter, z.B. zwischen den verschiedenen Klassen,

variieren können. Die adäquate Berücksichtigung der Mehrebenenstruktur der Daten ermöglicht nicht nur eine systematische sukzessive Modellentwicklung mit entsprechenden Prädiktoren auf verschiedenen Ebenen und ggf. die Untersuchung von möglichen Cross-Level-Interaktionen, sondern geht im Gegensatz zu konventionellen Varianz- und Regressionsanalysen insbesondere mit einer genaueren Berechnung der entsprechenden Effekte und einer adäquateren Schätzung ihrer Unsicherheit einher [15, S.30ff].

Aufgrund des in der vorliegenden Studie vergleichsweise großen Effekts des Treatments ist davon auszugehen, dass die hier präsentierten Ergebnisse trotz der mit den verwendeten statistischen Verfahren verbundenen Problemen – insbesondere bzgl. ihrer überschätzten statistischen Sicherheit – die grundsätzlichen Befunde korrekt widerspiegeln. Nichtsdestotrotz ist die Durchführung einer Mehrebenenanalyse aufgrund der vorliegenden hierarchischen Datenstrukturierung die angemessenere Analyseverfahren, sofern wie im vorliegenden Fall mit 23 Lehrer bzw. 36 Klassen die Voraussetzungen an die Anzahl der Einheiten auf den höheren Aggregatebenen erfüllt sind [17]. Da in der vorliegenden Stichprobe eine große und statistisch gesehen signifikante Varianz der Posttestergebnisse sowohl zwischen den Klassen als auch zwischen den Lehrern vorliegt, ist geplant, dieser Heterogenität durch eine mehrebenenanalytische Betrachtung Rechnung zu tragen. Ein weiterer Vorteil der Mehrebenenanalyse besteht darin, dass sie eine Kontrolle der Vortestunterschiede erlaubt [18, S.70], wohingegen das bisher verwendete Matchingverfahren auf Klassenebenen durch die Entfernung von Schülern nicht nur mit einem Informationsverlust, sondern auch mit einer nicht erstrebenswerten Veränderung der Struktur der Stichprobe einhergeht.

## 6. Literatur

- [1] Gleixner, C. (1998). Einleuchtende Elektrizitätslehre mit Potenzial, Dissertation LMU München.
- [2] Rhöneck, C. v. (1986). Vorstellungen vom elektrischen Stromkreis und zu den Begriffen Strom, Spannung und Widerstand, *Naturwissenschaften im Unterricht-Physik*, 34 (1986) 13, S.108-112.
- [3] Herrmann, F.; Schmälzle, P. (1984). Das elektrische Potential im Unterricht der Sekundarstufe I. In: *MNU* 37 (1984) 8, S.476-482.
- [4] Burde, J.-P.; Wilhelm, T.; Wiesner, H. (2014). Das Elektronengasmodell in der Sekundarstufe I. In: *PhyDid-B - Didaktik der Physik – Frühjahrstagung Frankfurt 2014*.
- [5] Chabay, R. W.; Sherwood, B. A. (2011). *Matter & Interactions*. Hoboken N.Y.: Wiley.
- [6] diSessa, A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10 (2-3), S.105-225.
- [7] Burde, J.-P.; Wilhelm, T. (2016). Die Elektrizitätslehre mit dem Elektronengasmodell. In: *Praxis der Naturwissenschaften – Physik in der Schule* 65, Nr. 8, 2016, S.18-24.
- [8] Burde, J.-P.; Wilhelm, T. (2016). Moment mal ... (22): Hilft die Wasserkreislaufanalogie? In: *Praxis der Naturwissenschaften – Physik in der Schule* 65, Nr. 1, 2016, S.46-49.
- [9] Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research Ansatz. *Unterrichtswissenschaft*, 33(1), S.52-69.
- [10] Wilhelm, T., Tobias, V., Waltner, C., Hopf, M. & Wiesner, H. (2012). Design-Based Research am Beispiel der zweidimensional-dynamischen Mechanik. In S. Bernholt (Hrsg.): *Konzepte fachdidaktischer Strukturierung für den Unterricht*, Jahrestagung der GDGP in Oldenburg 2011, Reihe: Gesellschaft für Didaktik der Chemie und Physik, Band 32, Lit-Verlag, Münster, S.31-47.
- [11] Burde, J.-P.; Wilhelm, T. (2015). Akzeptanzbefragung zum Elektronengasmodell. In: *PhyDid B - Didaktik der Physik – Frühjahrstagung Wuppertal 2015*.
- [12] Urban-Woldron, H. & Hopf, M. (2012). Entwicklung eines Testinstruments zum Verständnis in der Elektrizitätslehre. *Zeitschrift für die Didaktik der Naturwissenschaften* 18, S.201-227.
- [13] Bortz, J., Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Auflage. Berlin: Springer.
- [14] Kuhn, J. (2014). Mehrebenenanalyse am Beispiel der Lernwirkung von Aufgaben. In: D. Krüger, I. Parchmann und H. Schecker (Hg.): *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin: Springer Spektrum, S.297-310.
- [15] Twisk, J.W.R. (2006). *Applied Multilevel Analysis. A Practical Guide*. Cambridge: Cambridge University Press.
- [16] Kreft, I.; Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: SAGE.
- [17] Maas, C.J.M; Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. In: *Methodology* 1 (3), S.86-92.
- [18] Paccagnella, O. (2006). Centering or not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects. In: *Evaluation Review* 30 (1), S.66-85.