

Empirische Befunde zur Lernförderlichkeit des Elektronengasmodells

Motivation

Viele Unterrichtskonzepte zur Elektrizitätslehre legen den Fokus auf den Strom- statt auf den Spannungsbegriff. Es wird vermutet, dass dies möglicherweise ein späteres tieferes Verständnis elektrischer Stromkreise erschwert, da es Lernenden auch in der Sek I oftmals nicht gelingt, ein eigenständiges Spannungskonzept zu entwickeln (Cohen, Eylon & Ganiel, 1983; Rhöneck, 1986). Das Ziel des Unterrichtskonzepts auf Basis des Elektronengasmodells besteht deshalb darin, den Lernenden eine anschauliche und intuitive Vorstellung des Potenzial- und Spannungsbegriffs zu vermitteln, indem das elektrische Potenzial mit dem Luftdruck bzw. dem in Leitern herrschenden „elektrischen Druck“ verglichen wird. Ziel ist es dann, in Analogie zu Alltagsbeispielen wie Luftmatratzen oder Fahrradreifen, bei denen Luftdruckunterschiede die Ursache für Luftströmungen darstellen, die elektrische Spannung als „elektrischen Druckunterschied“ und somit Ursache und Antrieb des elektrischen Stroms einzuführen (Burde & Wilhelm, 2017).

Beschreibung der Studie und der Stichprobe

Im Rahmen einer mit 790 Schülern durchgeführten Studie wurde das beschriebene Unterrichtskonzept im Schuljahr 2015/16 im Frankfurter Raum empirisch evaluiert. Die quasi-experimentelle Studie folgte einem Prätest-Posttest-Treatment-Kontrollgruppen-Design und verwendete als Testinstrument einen Multiple-Choice-Test (Urban-Woldron & Hopf, 2012). Dieser zweistufige und psychometrisch ausgereifte Test bestand aus 22 Items zu Stromstärke und Widerstand und wurde um vier weitere Items zum Spannungsbegriff erweitert. Der maximal erreichbare Summenscore im Test beträgt damit 26 Punkte. Ziel der Studie war es u.a., 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).

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.

Um den erzielten Lernzuwachs in Kontroll- und Treatmentgruppe miteinander vergleichen zu können, wird im Folgenden der Nettoeffekt des Treatments betrachtet, d.h. wie viel Punkte im Test die Treatmentgruppe durch den Unterricht im Durchschnitt mehr erzielt hat als die Kontrollgruppe. Dieser Nettoeffekt des Treatments soll exemplarisch mit verschiedenen in der fachdidaktischen Forschung gängigen statistischen Methoden berechnet und deren jeweilige Vor- und Nachteile diskutiert werden. Konkret wird die Betrachtung der Differenzvariablen „Zugewinn“ mit einer Kovarianzanalyse und einer Mehrebenenanalyse verglichen.

Vergleich des erzielten absoluten Zugewinns in KG und TG

Im Folgenden wird der in Kontroll- und Treatmentgruppe erzielte absolute Lernzuwachs, definiert als Differenz des arithmetischen Mittelwerts von Post- und Pretest-Ergebnis, miteinander verglichen, um den signifikant unterschiedlichen Pretest-Ergebnissen Rechnung zu tragen. Zur Ermittlung des Nettoeffekts des Treatments gegenüber dem klassischen Unterricht wird von Bortz und Döring (2006, S. 559) vorgeschlagen, den Wert des absoluten Lernzuwachses der Kontrollgruppe (3.5 Punkte) von dem der Treatmentgruppe (6.7 Punkte) abzuziehen. Für den Nettoeffekt des neuen Unterrichtskonzepts ergibt sich somit ein Wert von 3.2 Punkten ($SE = 0.5$). Eine univariate zweifaktorielle Varianzanalyse mit Messwiederholung (gemischtes Design) zeigt, dass ein höchst signifikanter Interaktionseffekt zwischen Gruppen- und Messwiederholungsfaktor vorliegt ($p < .001$), sich Kontroll- und Treatmentgruppe also bzgl. des durchschnittlich erzielten absoluten Lernzuwachses höchst signifikant unterscheiden. Die Effektstärke beträgt $d = .62$, was einen mittleren Effekt darstellt.

Testergebnisse im Vergleich

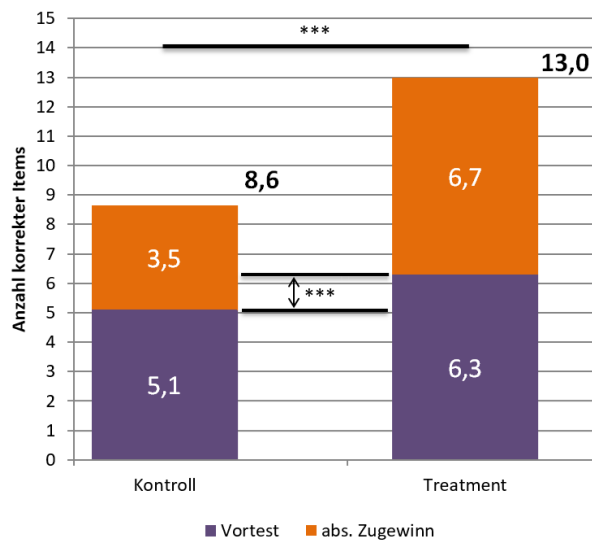


Abb. 1: Die Testergebnisse im Vergleich (Signifikanzniveau: *** $p < .001$)

Nach Bortz und Döring (2006, S. 560) ist allerdings davon auszugehen, dass sich die im Vortest gezeigte Diskrepanz zwischen Kontroll- und Treatmentgruppe im Nachtest alleine aufgrund von Regressionseffekten verkleinert. Dies hätte im vorliegenden Fall zur Folge, dass der Nettoeffekt des Treatments tendenziell unterschätzt wird. Unabhängig von der Diskrepanz im Vortest ist die alleinige Betrachtung des Zugewinns nach Todman und Dugard (1995, S. 182) nicht ausreichend, um den Einfluss des Pretest-Ergebnisses zu kontrollieren, da der absolute Zugewinn mit den Pretest-Ergebnissen korreliert. Eine Korrelationsanalyse zeigt, dass in dieser Studie der absolute Zugewinn negativ mit dem Pretest-Ergebnis korreliert ist. Dies bedeutet also, dass der absolute Zugewinn tendenziell umso geringer ausfällt, desto höher das Pretest-Ergebnis ist.

Kontrolle des Pretest-Ergebnisses mittels Kovarianzanalyse

Ein statistisches Verfahren, das diese auf Regressionseffekte zurückgehenden Schwierigkeiten vermeidet, stellt die Kovarianzanalyse (ANCOVA) dar. Der Vorteil der Kovarianzanalyse besteht insbesondere darin, dass das Pretest-Ergebnis der einzelnen Lernenden als Kova-

riate berücksichtigt und somit kontrolliert werden kann, was eine adäquatere Schätzung des Nettoeffekts des Treatments ermöglicht. Da die Kovarianzanalyse das Pretest-Ergebnis kontrolliert, wird der Nettoeffekt des Treatments hier über die Differenz der mittleren Posttest-Ergebnisse von Kontrollgruppe (9.0 Punkte) und Treatmentgruppe (12.7 Punkte) ermittelt. Für den Nettoeffekt des Treatments ergibt sich mit Hilfe einer Kovarianzanalyse somit ein Wert von 3.7 Punkten (SE = 0.5), was einen höchst signifikanten Unterschied darstellt ($p < .001$) und einer mittleren Effektstärke von $d = .75$ entspricht. Da die Kovariate „Pretest“ nicht völlig unabhängig von der Gruppenzugehörigkeit ist, kann allerdings davon ausgegangen werden, dass der so ermittelte Nettoeffekt des Treatments tendenziell zu gering ausfällt, weil die Kovariate ein Teil der Varianz des Posttest-Ergebnisses aufklärt, der eigentlich auf die Gruppenzugehörigkeit zurückgeht (Field, 2011, 397f).

Mehrebenenanalytische Berechnung des Nettoeffekts

Die bisher beschriebenen Verfahren gehen davon aus, dass die einzelnen Messungen statistisch gesehen unabhängig voneinander sind. Diese geforderte statistische Unabhängigkeit ist aber bei fachdidaktischer Feldforschung oftmals nicht gegeben, da die Schülerinnen und 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 bzw. dem gleichen Lernklima) überzufällig ähnlich sind (Kuhn, 2014, 297f). In der Terminologie der Mehrebenenanalyse (bzw. der „Hierarchisch Linearen Modellierung“) sind die Schüler in Schulklassen „geschachtelt“.

Eine bedeutende Konsequenz einer solchen „geschachtelten“ Datenstruktur besteht darin, dass ein zusätzlicher Schüler einer bestimmten Klasse aufgrund seiner Ähnlichkeit zu seinen Mitschülern keinen vollständig neuen Informationsbeitrag von 100% zur statistischen Schätzung liefert, sondern nur einen verminderten Informationsbeitrag von beispielsweise 80% (Kreft & Leeuw, 1998, S. 9; Twisk, 2006, S. 33). Je ähnlicher sich die Schülerinnen und Schüler innerhalb einer Klasse sind, je höher also die Korrelation zwischen ihnen ist, desto geringer fällt der zusätzliche Informationsbeitrag eines weiteren Schülers innerhalb einer solchen Klasse aus (Twisk, 2006, 14f). Um wie viel der Informationsbeitrag vermindert ist, hängt von der sogenannten Intraklassenkorrelation („Intraclass Correlation“) ab. Da der Standardfehler vom Stichprobenumfang abhängt, der effektive Stichprobenumfang mit steigender Intraklassenkorrelation in Folge des verminderten Informationsbeitrags aber abnimmt, bedeutet eine höhere Intraklassenkorrelation auch immer eine höhere Unsicherheit bzgl. des gefundenen Effekts. Wird die hierarchische Datenstrukturierung wie z.B. bei Kovarianzanalysen vernachlässigt, besteht deshalb die Gefahr einer Alpha-Fehler-Inflation, in deren Folge gewisse Effekte für statistisch signifikant gehalten werden, obwohl sie es tatsächlich nicht sind (Kreft & Leeuw, 1998, 9f). Die adäquate Berücksichtigung der hierarchischen Datenstruktur mittels einer Mehrebenenanalyse ermöglicht somit nicht nur eine genauere Bestimmung der untersuchten Effekte sondern auch eine adäquatere Schätzung ihrer statistischen Unsicherheit (Twisk, 2006, 30ff).

Die Erstellung eines Hierarchisch Linearen Modells mit drei Ebenen (Level-1: Schülerebene; Level-2: Klassenebene; Level-3: Lehrerebene) zeigt, dass lediglich der Level-2-Prädiktor „Gruppenzugehörigkeit“ und der Level-1-Prädiktor „Pretest-Ergebnis“ nach dem Bayesschen Informationskriterium einen positiven Einfluss auf die Modellgüte haben. Die Mehrebenenanalyse ergibt einen Nettoeffekt des Treatments von 3.9 Punkten (SE = 0.8), was einen höchst signifikanten Unterschied darstellt ($p < .001$) und einem großen Effekt von $d = .93$ entspricht. In dem vergleichsweise großen Standardfehler des Nettoeffekts des Treatments (0.8 Punkte bei der Mehrebenenanalyse vs. 0.5 Punkte bei Zugewinn und ANCOVA) spiegelt sich die adäquatere Schätzung der statistischen Unsicherheit wider.

Literatur

- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Auflage). Berlin: Springer.
- Burde, J.-P. & Wilhelm, T. (2017). Die Elektrizitätslehre mit dem Elektronengasmodell. *Praxis der Naturwissenschaften – Physik in der Schule*.
- Cohen, R., Eylon, B. & Ganiel, M. (1983). Potential difference and current in simple electric circuits: A study of students' concepts. *American Journal of Physics*, 51 (5), 407-412.
- Field, A. (2011). *Discovering Statistics using SPSS. (and sex and drugs and rock 'n' roll)* (3. Aufl.). London: SAGE.
- Kreft, I. & Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: SAGE.
- Kuhn, J. (2014). Mehrebenenanalyse am Beispiel der Lernwirkung von Aufgaben. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 297-310). Berlin: Springer Spektrum.
- Rhöneck, C. v. (1986). Vorstellungen vom elektrischen Stromkreis und zu den Begriffen Strom, Spannung und Widerstand. *Naturwissenschaften im Unterricht - Physik*, 34 (13), 10-14.
- Todman, J. & Dugard, P. (1995). Analysis of Pre- test- Post- test Control Group Designs in Educational Research. *Educational Psychology*, 15 (2), 181-198.
- Twisk, J.W.R. (2006). *Applied Multilevel Analysis. A Practical Guide*. Cambridge: Cambridge University Press.
- Urban-Woldron, H. & Hopf, M. (2012). Entwicklung eines Testinstruments zum Verständnis in der Elektrizitätslehre. *Zeitschrift für die Didaktik der Naturwissenschaften*, 18, 201-227. Verfügbar unter http://archiv.ipn.uni-kiel.de/zfdn/pdf/18_Urbahn.pdf