

## Experimentieren im Physikunterricht: Wechselwirkung zwischen Überzeugungen von Lehrkräften und unterschiedlichen Experimentiersituationen

Jan Winkelmann\*, S. Franziska C. Wenzel<sup>†</sup>, Holger Horz<sup>†</sup>, Mark Ullrich<sup>†</sup>, Jeremias Weber\*,  
Roger Erb\*

\*Institut für Didaktik der Physik, Goethe-Universität Frankfurt, Max von Laue Straße 1, 60438 Frankfurt,

<sup>†</sup>Institut für Psychologie, Goethe-Universität, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt

winkelmann@physik.uni-frankfurt.de, wenzel@psych.uni-frankfurt.de, horz@psych.uni-frankfurt.de,  
m.ullrich@psych.uni-frankfurt.de, jeremias.weber@physik.uni-frankfurt.de, roger.erb@physik.uni-frankfurt.de

### Kurzfassung

Im Rahmen einer umfangreichen Längsschnittstudie zum Kompetenzerwerb im Physikunterricht (KoPhy) wurde unter anderem der Lernzuwachs durch unterschiedliche Experimentiersituationen untersucht. Aktuelle Forschungsergebnisse lassen die Vermutung gerechtfertigt erscheinen, dass nicht das Experimentieren an sich – als Schülerexperiment oder als Demonstrationsexperiment – für den Lernerfolg ausschlaggebend ist. Vielmehr scheint eine Wechselwirkung zwischen der jeweilig unterrichtenden Lehrkraft und der Experimentiersituation zu bestehen. In der hier vorgestellten Studie wird versucht, diese Wechselwirkung mit Hilfe der physikbezogenen Überzeugungen der Lehrkräfte zu erklären. In diesem Beitrag werden erste Ergebnisse präsentiert und mögliche Konsequenzen für die Unterrichtsgestaltung sowie die Physiklehrerausbildung diskutiert.

### 1. Einleitung

Umfangreiche Reviews zum Thema „Lernen in den Naturwissenschaften“ zeigen, dass „laboratory work“ lediglich kleine Beiträge zu einem verbesserten Verständnis von Naturwissenschaften zu leisten vermag (Hofstein & Lunetta, 2004; Singer, Hilton, & Schweingruber, 2006). Diese Überblicksbeiträge erfassen den Einfluss des praktischen Arbeitens auf das Gesamtverständnis von Naturwissenschaften, also zum Beispiel auch Wissen über die Arbeitsweisen dieser Wissenschaften. Ein aktuelleres Review mit stärkerem Fokus auf die Wissensvermittlung naturwissenschaftlicher Inhalte findet sich bei Millar (2010). Andere Studien, bei denen der Einsatz von Schüler- bzw. Demonstrationsexperimenten direkt verglichen wurde (z.B. Hopf, 2007; Winkelmann & Erb, im Druck) lassen erwarten, dass Unterschiede in der Unterrichtsorganisation, also auf der Ebene der Sichtstruktur, nicht entscheidend zumindest für den Unterrichtserfolg in fachlicher Hinsicht sind. Entscheidender sind vielmehr tieferliegende Prozesse und damit verbunden die Lehrperson selbst (z. B. auch Hattie, 2013).

### 2. Studiendesign

In einer umfangreichen längsschnittlichen Erhebung wurde der Einfluss dreier verschiedener Experimentiersituationen auf den Lernzuwachs von Schülerinnen und Schülern der 7. und 8. Jahrgangsstufe untersucht (Projekt KoPhy, Erb & Horz, 2017).

Eine Teilfrage des Projekts lautet: Welche Auswirkung hat die Interaktion von Lehrercharakteristika und Experimentiersituation auf die Kompetenzentwicklung von Schülerinnen und Schülern im Fach Physik? Ausgehend von bisherigen Forschungsergebnissen, wurden die folgenden zwei Hypothesen formuliert:

H.1.: Lehrkräfte mit zahlreichen konstruktivistischen Überzeugungen hinsichtlich des Lehrens und Lernens bewirken bei den Schülerinnen und Schülern dann einen größeren Fortschritt im Kompetenzbereich „Fachwissen“, wenn sie ein Treatment mit höherer Eigensteuerung unterrichten.

H.2.: Lehrkräfte mit wenigen konstruktivistischen Überzeugungen hinsichtlich des Lehrens und Lernens bewirken bei den Schülerinnen und Schülern dann einen größeren Fortschritt im Kompetenzbereich „Fachwissen“, wenn sie ein Treatment mit geringerer Eigensteuerung unterrichten.

Wir vermuten also eine Passung zwischen der Überzeugung der unterrichtenden Lehrkraft und einer entsprechenden Experimentiersituation, die zum Erfolg bei Schülerinnen und Schülern führen sollte.

### 3. Messinstrumente

Im Rahmen der Studie wurden verschiedene Tests zum Einsatz gebracht. Als Kontrollvariablen wurden personengebundenen Daten sowie zwei Skalen des Kognitionsfähigkeitstests (KFT, V3 & N2; Heller & Perleth, 2000) erhoben. Auf Seiten der Schülerinnen und Schüler wurden als abhängige Variablen das

aktuelle Interesse an Physik (Schulz, 2011) sowie deren Kompetenz in den Bereichen der Erkenntnisgewinnung (Glug, 2009) und des Fachwissens (Weber et al., 2017) erhoben. Die unterrichtenden Lehrkräfte wurden gebeten, Skalen zu ihren Überzeugungen zum Lehren und Lernen von Physik zu bearbeiten (Lamprecht, 2011). Da im vorliegenden Beitrag der Fokus auf der Wechselwirkung zwischen fachlichem Wissen der Schülerinnen und Schüler und den Überzeugungen der Lehrkräfte liegt, werden beide Tests im Folgenden etwas näher beschrieben.

### 3.1 Fachwissenstest geometrische Optik

Der verwendete Fachwissenstest zur geometrischen Optik geht auf Winkelmann (2015) zurück. Inhaltlich fokussiert der Test auf die Lichtbrechung. Bestehende Items des damaligen Tests wurden hinsichtlich zentraler psychometrischer Aspekte und unter Nutzung von Modellen der Item-Response-Theorie (IRT) analysiert und problematische Items überarbeitet. Überarbeitete und neu entwickelte Items wurden im Rahmen einer Pilotstudie IRT-skaliert und entsprechend ihrer psychometrischen Güte für den Fachwissenstest selektiert oder eliminiert. Zuletzt wurden aus dem so gewonnenen Itempool ( $n = 48$ ) Testhefte zusammengestellt, die eine Erfassung des Fachwissens im Bereich „Geometrische Optik“ über verschiedene Messzeitpunkte hinweg und damit die Verfolgung der Entwicklung des Fachwissens über die Zeit, erlauben. Dabei wiederholt sich jeweils nur ein Teil der Items über mehr als einen Messzeitpunkt, diese werden Ankeritems genannt. Die übrigen Items sind messzeitpunktspezifisch ausgewählt. Insbesondere wurde dabei die Schwierigkeit der einzelnen Items an die zu erwartende Personenfähigkeit angepasst: Im Pretest wurden also besonders einfache Items genutzt, im Posttest dafür schwerere Items. Um Reihenfolgeeffekte soweit wie möglich zu minimieren, wurden für jeden Messzeitpunkt fünf Testhefte anhand eines balancierten unvollständigen Testheftdesigns erstellt (Osterlind & Everson, 2009). Insgesamt weist der so entwickelte Fachwissenstest eine Bearbeitungsdauer von ca. 25 Minuten auf und erlaubt eine änderungssensitive sowie reliable Erfassung des Fachwissens.

### 3.2 Überzeugungen von Lehrkräften

Um die Überzeugungen der unterrichtenden Lehrkräfte in Hinblick auf die Physik als Unterrichtsfach und Wissenschaft zu erheben, wurden Skalen von Lamprecht (2011) verwendet. Für die Überzeugungen zum Unterrichtsfach Physik kamen die Skalen „Selbstständiges Lernen von Physik“ (SL) sowie „Rezeptartiges Lernen“ (RL) und für die Überzeugungen zur Wissenschaft Physik die Skalen „Stellenwert der Wissenschaft Physik“ (WPh) sowie „Wahrheitsanspruch und Eindeutigkeit von Physik“ (WuE) zum Einsatz.

Im Rahmen seiner Arbeit gelang es Lamprecht, seine in den Blick genommenen Lehrkräfte (Quer-

einsteiger in das Lehramt Physik) auf Grundlage ihrer Überzeugungen als drei unterschiedliche Lehrertypen zu beschreiben: den Trainingstyp, den diskursiven Typ und den Vermittlungstyp.

## 4. Auswertung

An der Studie nahmen insgesamt über 1000 Schülerinnen und Schüler in Hessen und Nordrhein-Westfalen teil. Die in diesem Beitrag vorgestellten, ersten Ergebnisse beziehen sich auf Daten von 944 Schülerinnen und Schülern sowie 51 Physiklehrkräften.

### 4.1 Mehrebenen-Analyse

Um eine Aussage über den Einfluss der verschiedenen Experimentiersituationen auf den Lernzuwachs der Schülerinnen und Schüler treffen zu können, wurden die Schülerleistungen zu den verschiedenen Messzeitpunkten im Rahmen einer Mehrebenen-Analyse untersucht. Eine Mehrebenen-Analyse der Schülerdaten ist aus zweierlei Gründen einer Varianzanalyse vorzuziehen. Zum einen sind die Schülerdaten voneinander abhängig (sie sind „genestet“). Dadurch würden in einer Varianzanalyse Zusammenhänge falsch benannt und Varianzanteile auf Klassenebene unterschätzt werden. Zum anderen ist auf diese Weise eine Imputation fehlender Werte möglich, was in einer Varianzanalyse mit Messwiederholung in SPSS nicht möglich ist.

Den in Abschnitt fünf vorgestellten (vorläufigen) Ergebnissen liegt ein nach dem Maximumlikelihood Verfahren geschätztes Modell zu Grunde. Dabei gehen wir nicht von einem linearen Verlauf zwischen Pre- Post- und Follow-Up-Test aus. Die Fitwerte sind überzeugend: Chi Square .59 bei 2 df; RMSEA  $< .0001$ ; CFI 1; SRMR .005 (within / Schüler) und .001 (between / Klasse).

### 4.2 Identifikation von Lehrertypen

An der hier vorgestellten Studie nahmen 51 Lehrkräfte mit ihren Klassen teil. Die Anzahl ist zu gering, um damit eine sinnvolle Clusteranalyse, wie bei Lamprecht (2011), durchzuführen. Die Skalen von Lamprecht wurden auch im Projekt proPHI der Arbeitsgruppe Korneck, Oettinghaus & Lamprecht (2016) eingesetzt. Durch die Kombination dieser Daten mit den unsrigen gelang eine zuverlässige Zuordnung der teilgenommenen Lehrkräfte zu den von Lamprecht beschriebenen Lehrertypen. Abbildung 1 zeigt eine ähnliche Ausprägung der Skalen für die jeweiligen Lehrertypen. Deutlich ist die Gegenläufigkeit von Trainings- und diskursivem Muster zu erkennen. Allerdings sind die positiven Ausprägungen in den Skalen RL und WPh etwas geringer. Das Vermittlungsmuster liegt auch bei uns über den Ausprägungen der anderen beiden Muster, allein in der Skala zum selbstständigen Lernen ist dieses

Muster in unseren Daten etwas schwächer ausgeprägt.

Die (wieder)gefundenen Lehrertypen sind in der vorliegenden Studie bisher wie folgt mit Lehrkräften besetzt: Trainingsmuster n = 13, diskursives Muster n = 23 und Vermittlungsmuster n = 14.

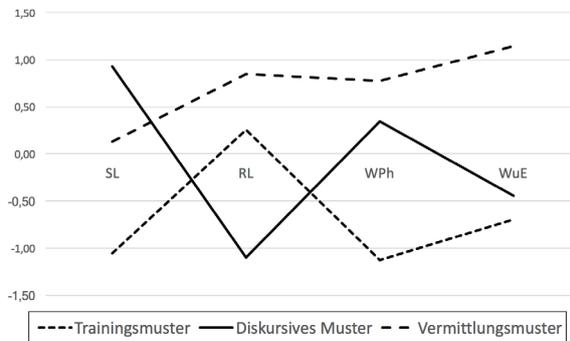


Abb. 1: In vorliegender Studie gelang eine vergleichbare Typenbildung wie bei Lamprecht (2011).

## 5. Erste Ergebnisse

Betrachtet man in der Mehrebenen-Analyse den Einfluss der verschiedenen unabhängigen Variablen (Experimentiersituation bzw. Lehrertyp), zeigen sich keine Effekte auf den Lernzuwachs der Schülerinnen und Schüler. Dies bestätigt eigene Vorarbeiten (Winkelmann, 2015). Lediglich mittels einer Dummy-Codierung der Variable Experimentiersituation (demo = 00) unterscheidet sich die Situation „Kochbuch“ im Verlauf über die drei Messzeitpunkte leicht signifikant (mit kleinem Effekt) von beiden anderen Treatments

### 5.1 Wechselwirkung

Aufgrund der Ergebnisse von Winkelmann (2015) ist zu vermuten, dass es eine Interaktion zwischen der erlebten Experimentiersituation und der unterrichtenden Lehrkraft besteht. In der damaligen Studie konnte diese Wechselwirkung berichtet, jedoch nicht erklärt werden.

Eine Möglichkeit, die unterrichtenden Lehrkräfte näher zu beschreiben, liegt in der Analyse ihrer Überzeugungen zum Lehren und Lernen von Physik. Für die Analyse möglicher Wechselwirkungen zwischen den Lehrkräften und den Experimentiersituationen wurden sämtliche möglichen Kombinationen mit einer ausgewählten Kombination verglichen (Dummy-Codierung mit „Training x Demo“ = 0). Abbildung 2 fasst die so aufgespannte Matrix zusammen.

Es zeigt sich, dass es in der oberen Zeile (nur Trainingsstyp-Lehrkräfte) sowie in der ersten Spalte (nur Experimentiersituation Demo) leicht signifikante Unterschiede zur „Null-Zelle“ gibt. Eine vorsichtige Interpretation kann lauten: Die Kombination aus Lehrertyp und Experimentiersituation (insbesondere mit starker Lenkung = Demo) spielt eine Rolle für den Leistungsverlauf der Schüler. Außerdem er-

Interaktionsterme dummy codiert (demo / training = 00)

(Training x Demo)	Training x Kochbuch *	Training x Guided *
Vermittlung x Demo *	Vermittlung x Kochbuch	Vermittlung x Guided
Diskursiv x Demo *	Diskursiv x Kochbuch	Diskursiv x Guided

Abb. 2: Zusammenfassung der Interaktionsterme für die Wechselwirkung zwischen Lehrertyp und Experimentiersituation.

scheint insbesondere das Trainingsmuster für Interaktionen interessant bzw. anfällig.

Allerdings müssen diese Ergebnisse unter folgenden Einschränkungen betrachtet werden: Sämtliche gefundenen Unterschiede sind nur knapp signifikant. Die einzelnen Zellen der Matrix (Abb. 2) sind nicht gleich stark und auch nur schwach besetzt (in die Analyse konnten bisher nur Daten von 38 Lehrkräften aufgenommen werden). Noch ist zudem nicht geklärt, in welche Richtung die leicht signifikanten Unterschiede zu interpretieren sind. D.h. es ist noch nicht klar, ob die Kombinationen ein besseres oder schlechteres Ergebnis erzielten.

## 6. Zusammenfassung

Aktuell erwarten wir noch Daten von 14 weiteren Klassen, die mit in die Analyse aufgenommen werden sollen. Mit dem Hinweis auf die Vorläufigkeit der hier beschriebenen Analysen kann zusammengefasst werden:

### 1. Frühere Befunde können bestätigt werden:

Es ist zum einen gelungen, die Lehrertypen von Lamprecht auch in den unsrigen Daten wieder zu finden. Darüber hinaus konnten die Ergebnisse (eigener) früherer Arbeiten zu Auswirkungen des Experimentierens auf den Lernzuwachs bestätigt werden: Ob mit Schüler- oder Demonstrationsexperimenten unterrichtet wird, hat keine Bedeutung für den Lernzuwachs.

### 2. Lehrertypen und Treatment haben jeweils keinen Effekt auf das Fachwissen.

Neben der Art des Experimentierens scheint auch der Lehrertyp für sich alleine genommen keinen Einfluss auf den Lernerfolg von Schülerinnen und Schülern zu haben.

### 3. Interaktionen zwischen Lehrertypen und Experimentiersituation

Die Interaktion wird bei starker Lenkung (durch die Lehrkraft oder durch das Treatment) ansatzweise signifikant. Dieses Ergebnis kann zum jetzigen Zeitpunkt allerdings nur sehr vorsichtig interpretiert werden. Es stehen weitere Analysen aus.

## 7. Ausblick

Um einen tieferen Einblick in die vorliegenden Daten zu erhalten, sollen weitere Analysen durchgeführt werden. So steht eine Kontrolle der Dummy-Codierung aus: Werden sich ähnliche signifikante Unterschiede zeigen, wenn andere Kombinationen als „Null-Zelle“ definiert werden? Außerdem wurde bisher nicht mit dem vollständigen Datensatz gerechnet. Statt der bisherigen 38 Klassen werden am Ende Daten von 50 Klassen bzw. etwa 1250 Schülerinnen und Schülern zur Verfügung stehen. Mit dem vollständigen Datensatz sollen dann weitere statistische Analyseverfahren zum Einsatz gebracht werden, etwa eine Piece-wise Analyse der Leistungsentwicklung zwischen Pre- und Posttest sowie zwischen Post und verschiedenen Follow-Up-Zeitpunkten. Auch sollen weitere Kontrollvariablen wie das Geschlecht oder die Lehrererfahrung in das Modell der Mehrebenen-Analyse aufgenommen werden.

## 8. Literatur

- Erb, R. & Horz, H. (2017). Experimentieren in Physik. In BMBF (Hrsg.), *Optimale Lern- und Bildungschancen für Kinder und Jugendliche. Aktuelle Forschungsergebnisse für die Bildungspraxis*, S. 8-9. Zuletzt aufgerufen am 12.5.2018 unter: [https://www.bmbf.de/pub/Optimale\\_Lern\\_und\\_Bildungschancen\\_fuer\\_Kinder\\_und\\_Jugendliche.pdf](https://www.bmbf.de/pub/Optimale_Lern_und_Bildungschancen_fuer_Kinder_und_Jugendliche.pdf)
- Hattie, J. (2013). *Lernen sichtbar machen*. Schneider Verlag Hohengehren.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R)*. Göttingen: Beltz Test GmbH.
- Glug, I. (2009). *Entwicklung und Validierung eines Multiple-Choice-Tests zur Erfassung prozessbezogener naturwissenschaftlicher Grundbildung*. Kiel: IPN.
- Hofstein, A., & Lunetta, V. N. (2004). The Laboratory in Science Education: Foundations for the Twenty-First Century. *Science Education*, 88, 28-54.
- Hopf, M. (2007). Problemorientierte Schülerexperimente. In H. Niedderer, H. Fischler, E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen*. Band 68. Berlin: Logos.
- Korneck, F., Oettinghaus, L., & Lamprecht, J. (2016). *Komponenten professioneller Handlungskompetenz von Quereinsteigern und Lehramtsstudierenden des Haupt- und Realschullehrerstudiums im Fach Physik (proPHI)*. Version: 1. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. [http://doi.org/10.5159/IQB\\_proPHI\\_v1](http://doi.org/10.5159/IQB_proPHI_v1)
- Lamprecht, J. (2011). Ausbildungswege und Komponenten professioneller Handlungskompetenz. Vergleich von Quereinsteigern mit Lehramtsabsolventen für Gymnasien im Fach Physik. In H. Niedderer, H. Fischler, E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen*, Band 125. Berlin: Logos Verlag.
- Millar, R. (2010). Practical work. In J. Osborne & J. Dillon (Eds.), *Good practice in science teaching: What research has to say* (pp. 108-134). Maidenhead, Great Britain: Open University Press.
- Osterlind, S. J. & Everson, H. T. (2009). *Differentialitem functioning* (Vol. 161). Sage Publications.
- Schulz, A. (2011). Experimentierspezifische Qualitätsmerkmale im Chemieunterricht. Eine Videostudie. In H. Niedderer, H. Fischler, E. Sumfleth (Hrsg.), *Studien zum Physiklernen*. Band 113. Berlin: Logos Verlag.
- Singer, S. R., Hilton, M. L., & Schweingruber, H. A. (2006). *America's Lab Report. Investigations in High School Science*. Washington, DC: The National Academies.
- Weber, J., Winkelmann, J., Erb, R., Wenzel, F., Ullrich, M., & Horz, H. (2017). Entwicklung eines Fachwissenstests in der geometrischen Optik. In H. Grötzebauch & V. Nordmeier (Hrsg.), *PhyDid B – Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung des Fachverbands Didaktik der Physik in Hannover 2016*. Beitrag DD. 17.03.
- Winkelmann, J. & Erb, R. (2018). Der Einfluss von Schüler- und Demonstrationsexperimenten auf den Lernzuwachs in Physik. *Phydid A - Physik und Didaktik in Schule und Hochschule*, (1) 17, 21-33.
- Winkelmann, J. (2015). Auswirkungen auf den Fachwissenszuwachs und auf affektive Schülermerkmale durch Schüler- und Demonstrationsexperimente im Physikunterricht. In H. Niedderer, H. Fischler, & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen*. Band 179. Berlin: Logos Verlag.