

Entwicklungsorientierte physikdidaktische Forschung
– **ausgewählte Aspekte der Vorbereitung, Durchführung und Auswertung**
entsprechender Studien –

Roland Berger

Universität Osnabrück, Barbarastraße 7, 49076 Osnabrück
r.berger@uos.de

Kurzfassung

In seinem Gastbeitrag zur Tagung „Physikdidaktik – Quo vadis?“ in 2021 hat der Erziehungswissenschaftler Ewald Terhart in seinem „Außenblick“ auf die Physikdidaktik die Stärkung „entwicklungsorientierter fachdidaktischer Forschung“ als wichtiges Ziel hervorgehoben. Als eine geeignete Möglichkeit dieses Anliegen zu stärken erscheint die Verknüpfung der Entwicklung und Implementation innovativer Konzepte und Materialien mit der empirischen Klärung relevanter Forschungsfragen. Auf der Basis eigener Erfahrungen und einschlägiger Literatur werden im Folgenden ausgewählte Aspekte der Vorbereitung, Durchführung und Auswertung entsprechender Studien diskutiert. Ein wesentliches Ziel ist dabei die Anregung und Unterstützung entwicklungsorientierter physikdidaktischer Forschung mit direktem Schulbezug.

1. Einleitung

In seinem Vortrag im Rahmen einer der Quo vadis - Tagungen der DPG im Jahr 2021 kam der Erziehungswissenschaftler Ewald Terhart in seinem Vortrag zu folgendem Schluss: *„Durch die domänenspezifisch-lernpsychologische Ausrichtung ist keineswegs die Praxisbedeutsamkeit der fachdidaktischen Forschung gesteigert worden. [...] Deshalb halte ich es für wichtig, nunmehr den Typus der entwicklungsorientierten fachdidaktischen Forschung unbedingt stärker zu machen!“* (Terhart, 2021, S. 8f.).

Schaut man sich den im Internet verfügbaren Text des Vortrags vollständig an, so erkennt man, dass Terhart keineswegs fachdidaktische Forschung gegen die Entwicklung von Unterricht ausspielen will. Vielmehr sollte sowohl Forschung *als auch* ihre Entwicklungsorientierung im Auge behalten werden, um die Schule zu stärken.

Was ist das Ziel der Verknüpfung von Forschung und Entwicklung? In meinen Augen ist das Ziel zu versuchen, die Klärung einer interessanten Forschungsfrage mit der Entwicklung von innovativen Unterrichtsinhalten (Konzepte und Materialien) zu verknüpfen, die Lehrkräfte möglichst kostenfrei erhalten können, und die als Grundlage für Lehrerfortbildungen dienen können, was oft auf hohe Akzeptanz und Interesse von Lehrkräften stößt (Lipowsky & Rzejak, 2012). Diese Verknüpfung gelingt nicht immer und erscheint auch nicht immer sinnvoll. Sie ist aber durchaus möglich, wie zahlreiche Beispiele in der Physikdidaktik zeigen. In Osnabrück wurde beispielsweise Cross-Age Tutoring intensiv untersucht. Dabei haben Schülerinnen und Schüler von Hauptschulen in Grundschulen Grundelemente des geschlossenen elektrischen Stromkrei-

ses vermittelt. Dazu gehörten vor allem die Notwendigkeit (mindestens) zweier Kabel und ferner die Nutzung unterschiedlicher Energiequellen (Solarzelle, Thermoelement, Kurbel-Generator und Essig-Batterie). Im Vorunterricht hatten die Hauptschülerinnen und -schüler auf der Basis dieser Stromkreise zusätzlich Umwandlungsprozesse bei elektrischen Energieübertragungen behandelt, z.B. die Umwandlung von chemischer Energie in elektrische Energie mit der Essig-Batterie. Im Rahmen der Studie wurden Lehrtexte, Folien und Experimente genutzt oder entwickelt, die sich Lehrkräfte von unserer Internetseite www.physikdidaktik.uni-osnabrueck.de/ vollständig und kostenfrei herunterladen können. Dieses doppelte Ziel – Forschung auf der einen Seite und Entwicklung von Unterricht auf der anderen Seite – spiegelt sich auch darin wieder, dass Ergebnisse der Studie sowohl in einer Zeitschrift für Lehrkräfte (Berger & Müller, 2015), als auch in einer dezidiert wissenschaftlichen Zeitschrift (Hänze, Müller & Berger, 2018) veröffentlicht wurden.

Im Folgenden werden ausgewählte Aspekte behandelt, die bei der Vorbereitung, Durchführung und Auswertung entwicklungsorientierter physikdidaktischer Forschung hilfreich sein könnten, um so weitere Studien mit diesem Ziel zu unterstützen.

2. Rolle des Lehrplans

Die Lehrpläne machen Vorgaben für Lehrkräfte für die zu unterrichtenden Inhalte. Mittelbar haben Sie damit aber auch Einfluss auf die fachdidaktische Forschung und Entwicklung. Denn eine mäßige oder gar fehlende Passung zum Lehrplan erschwert inhaltliche Innovationen, da sich eher schwerer Lehrkräfte finden, die bereit sind, an entsprechenden Studien teilzunehmen, und möglicherweise auch die

Genehmigung durch Schulbehörden erschwert sein könnte. Damit haben es Unterrichtskonzepte wie z.B. aus „Nature of Science“ schwerer, oder eine Thermodynamik für die Oberstufe, die auf dem Teilchenmodell basiert. Dadurch entsteht leicht ein *circulus vitiosus*: Das Fehlen überzeugender Unterrichtskonzepte erschwert die Aufnahme in den Lehrplan, was wiederum die inhaltliche Entwicklung des Physikunterrichts beeinträchtigen kann. Entsprechender Unterricht über den Lehrplan hinaus darf dann – sofern er überhaupt versucht wird – nur wenige Stunden umfassen, um Lehrkräfte für eine Teilnahme zu gewinnen. Dies hat zur Folge, dass Forschung und Entwicklung zu wichtigen, aber umfangreichen Unterrichtskonzepten womöglich unterbleibt. Dies ist vielleicht auch ein Grund für die Einschätzung von Terhart (2021, S. 8), wonach „der Gewinn für die Unterrichtspraxis und bis zu einem gewissen Grad auch für die Lehrerbildung nicht so hoch [ist], wie man das erwartet hat.“

Sofern aber eine gute Passung zu den Lehrplänen besteht, sind auch umfangreiche Interventionen möglich. Beispielsweise wurden in Osnabrück in den vergangenen Jahren für die Mittelstufe eine Unterrichtseinheit zum Kraftkonzept mit 15 Schulstunden entwickelt (Berger & Wöllermann, 2022) und eine sehr umfangreiche Unterrichtseinheit zum elektrischen Stromkreis mit dem Schwerpunkt der Einführung des elektrischen Potenzials mit insgesamt 34 Schulstunden (vgl. die Beiträge von M. Kahnt und A. Hindriksen in diesem Tagungsband). Gut lassen sich auch kurze Interventionen durchführen, insbesondere wenn sie Lehrplanbezug haben. Beispielsweise benötigt eine Unterrichtseinheit zur Wiederholung des Themas Schwingungen und Wellen im Rahmen des Kontexts „Mikrowellenofen“ drei Schulstunden (Berger, 2007).

3. Sicherung der inhaltlichen Qualität

Eine Stärke universitärer fachdidaktischer Forschung ist die Möglichkeit, die fachdidaktische Qualität von Unterricht mit empirischen Methoden sicherzustellen. Hierzu eignet sich die Methode der „Akzeptanzbefragung“ besonders gut. Dabei werden ca. 10 Schülerinnen und Schüler in der Regel einzeln anhand eines Interviewleitfadens entlang einer von Wiesner und Wodzinski (1996) dargestellten Grundstruktur interviewt. Im Wesentlichen geht es dabei darum, das Unterrichtsangebot vorzustellen, und durch Transferaufgaben herauszufinden, wo Verständnisschwierigkeiten bestehen. Die Ergebnisse sollten dann in die Überarbeitung des Unterrichtskonzepts einfließen. Dieses Vorgehen lässt sich im Rahmen der Didaktischen Rekonstruktion als „Erfassen der Lernerperspektive“ (Kattmann et al., 1997) und im Design-Based-Research-Ansatz als Element eines iterativen Prozesses (Haagen-Schützenhöfer & Hopf, 2020) auffassen.

Akzeptanzbefragungen zu planen, durchzuführen und auszuwerten ist anspruchsvoll und aufwändig.

Ideal erscheint es, wenn Promovierende die Akzeptanzbefragung selbst durchführen. Dies kann aber auch in universitären Abschlussarbeiten mit guten Studierenden gelingen, sofern diese bei der Vorbereitung und Durchführung sehr gut von den Betreuenden unterstützt werden.

4. Auswertung „quasi-experimenteller“ Studien

Ist das Unterrichtskonzept nun ausgearbeitet, so lassen sich auf dieser Grundlage Forschungsfragen untersuchen, die häufig quantitativ ausgewertet werden. Beispielsweise wurde in der genannten Studie zum elektrischen Potenzial die Frage untersucht, welche Art von Analogie zur Illustration des elektrischen Potenzials besonders hilfreich ist. Neben einer Kontrollgruppe wurde dabei das in Bezug auf das elektrische Potenzial neuartige Modell der elektrischen Fahrradkette mit einem Höhenmodell („Stäbchenmodell“; Burde & Wilhelm, 2021) verglichen. Derartiger Unterricht ist naturgemäß nur möglich, wenn die teilnehmenden Klassen mit ihrer jeweiligen Lehrkraft einer der drei Bedingungen zufällig zugewiesen werden („Quasi-Experiment“), und nicht die einzelnen Schülerinnen und Schüler („echtes“ Experiment). Die Schülerinnen und Schüler sind dann in ihre jeweilige Klasse „hineingeschachtelt“ („nested“). Durch die erhöhte Ähnlichkeit der Schülerinnen und Schüler innerhalb der Klassen, z.B. durch zwischen den Klassen unterschiedlich guten Unterricht, ist eine der zentralen Voraussetzungen für eine Varianzanalyse in der Regel verletzt. Die Schachtelung muss daher statistisch berücksichtigt werden. Dies kann im Rahmen einer „echten“ Mehrebenenanalyse geschehen (d.h. basierend auf iterativen Maximum Likelihood Algorithmen im Unterschied zu üblichen Varianzanalysen, die im Prinzip auf der formelbasierten Berechnung von Quadratsummen beruhen). Allerdings wird dieses mächtige und auch komplexe Werkzeug oft nicht benötigt, und es genügt eine *geschachtelte* (Ko-)Varianzanalyse („nested AN(C)OVA“; Bliese & Hanges, 2004). Berücksichtigt man die Schachtelung der Schülerinnen und Schüler in ihren Klassen nicht, so ist der verwendete Test (z.B. eine ANOVA) zu liberal, und man findet scheinbar signifikante Effekte, die in Wirklichkeit aber Artefakte sind (Bliese & Hanges, 2004). In unserer Vergleichsstudie zum elektrischen Potenzial ergibt sich mit einer ungeschachtelten ANCOVA (Syntax in SPSS: UNIANOVA Nachtest BY Modell WITH Vortest.) ein signifikanter Effekt ($F = 4.73$, $p < .01$). Die geschachtelte ANCOVA mit der SPSS-Syntax

```
UNIANOVA Nachtest BY Modell Klasse WITH Vortest
```

```
/RANDOM = Klasse
```

```
/DESIGN = Vortest Modell Klasse WITHIN Modell.
```

ergibt hingegen lediglich $F = 1.52$, $p = .25$, und damit keinen signifikanten Effekt mehr. In der angegebenen Syntax wird die Schachtelung im DESIGN-

Befehl durch „Klasse WITHIN Modell“ berücksichtigt.

Geschachtelte Varianzanalysen berücksichtigen neben der quasi-experimentell variierten Bedingung (im Beispiel zum elektrischen Potenzial die Art der Analogie) die „Klasse“ als weiteren Faktor. Dieser Faktor muss statistisch als zufälliger Faktor behandelt werden, da in der Regel keine Aussagen nur über bestimmte (für die Untersuchung herangezogene) Klassen gemacht werden sollen, sondern möglichst auf die Grundgesamtheit von Klassen verallgemeinert werden soll (vgl. Bortz, 2005, S.302). Aus statistischer Sicht weist Bortz darüber hinaus auf einen Artikel von Zucker (1990) hin der zeigt, dass die geschachtelte ANOVA zu liberal ist, wenn man „Klasse“ nicht als zufälligen, sondern als festen Faktor behandelt.

Es ist (als „vertrauensbildende Maßnahme“) sehr instruktiv, die geschachtelte Varianzanalyse mit dem von Bortz (2005, S. 389-391) im Kapitel „Hierarchische und teilhierarchische Versuchspläne“ gegebenen Beispiel zu prüfen. Bortz rechnet „von Hand“ eine zweifaktorielle Varianzanalyse (wobei ein Faktor der „Schachtelfaktor“ ist) mit einem „handlich“ kleinen Datensatz durch Berechnung der Quadratsummen. Mithilfe der (an das Beispiel von Bortz angepassten) oben angegebenen SPSS-Syntax kommt man (wie es sein muss) zum gleichen Ergebnis.

5. Qualität von Leistungstests

Ein häufig angegebenes psychometrisches Maß für die Qualität von Fragebögen, aber auch von Tests ist Cronbach's α . Insbesondere bei Leistungstests liegt der Wert nicht selten mehr oder weniger deutlich unterhalb von 0.7, einem Wert, der häufig als Mindestwert für die interne Konsistenz einer Skala aufgefasst wird, z.B. das im Rahmen von mehreren Aufgaben erfasste Verständnis des elektrischen Spannungsbegriffs. Da es im schulischen Kontext aus Zeitgründen häufig nicht möglich ist, viele Aufgaben zu einem Konstrukt zu stellen, beeinträchtigt dies den Wert von Cronbach's α , da er von der Anzahl der Items abhängt. Clark und Watson (1995) empfehlen daher (ergänzend) die Angabe der (von der Itemzahl unabhängigen) mittleren Inter-Item-Korrelation, die (je nach „Bandbreite“ des Konstrukts) im Bereich von .15 bis .50 liegen sollte. Um Cronbach's α in einen „akzeptablen“ Bereich anzuheben könnte man versucht sein, die Auswahl der Items allein aufgrund von psychometrischen Kriterien vorzunehmen. Schecker (2014) schreibt hierzu: „Es wäre fatal, wenn zugunsten hoher α -Werte zu große Abstriche bei der inhaltlichen Ausrichtung der Items in Kauf genommen werden. Fachdidaktisch inhaltlich anspruchsvolle Leistungstests, bei denen die Itemzahl wegen der begrenzten Testzeit nicht beliebig erhöht werden kann, haben Schwierigkeiten, hohe α -Koeffizienten zu erreichen.“

In Anbetracht der Schwierigkeit, inhaltlich und psychometrisch sehr gute Leistungstests zu entwickeln, ist es umso problematischer, dass in vielen Veröffentlichungen (wenn überhaupt) nur einige ausgewählte Aufgaben aus dem in der entsprechenden Studie eingesetzten Aufgabenpool berichtet werden. Das hat zur Folge, dass Tests immer wieder neu aufwändig entwickelt werden müssen. Darüber hinaus lässt sich die Aussagekraft von Studien schwer einschätzen, und der Vergleich mit anderen Studien ist erschwert. Vor allem aber lassen sich Studien entgegen einer zentralen Anforderung an wissenschaftliche Forschung nicht replizieren, wenn die Aufgaben nicht vollständig bekannt sind.

Taber (2018, S. 1294) betont die hohe Bedeutung der Augenscheinvalidität, die eine Sichtung des vollständigen Aufgabenpools erfordert. Er kommt zu dem Schluss: „Research that seeks to be considered scientific needs to be reported in sufficient detail to allow others to build on it – and that clearly requires the availability of full instruments rather than simply reporting sample items.“

Eine Veröffentlichung des gesamten Aufgabenpools ist heutzutage technisch leicht möglich (z.B. als elektronischer Anhang in Zeitschriften wie der *Phy-Did*, aber auch in Repositorien wie dem „Forschungsdatenzentrum Bildung“). Um den Anschluss an aktuelle Standards nicht zu verlieren erscheint eine Orientierung zum Beispiel an der Deutschen Gesellschaft für Psychologie sinnvoll. Sie empfiehlt auf der Basis der Leitlinien zum Umgang mit Forschungsdaten der Deutschen Forschungsgemeinschaft: „Mit Erscheinen einer Publikation soll die Person oder Gruppe, die die Daten erhoben hat (die Datenbereitstellenden), alle Primärdaten sowie die dazugehörigen Metadaten bereitstellen, die zur Reproduktion der publizierten Ergebnisse notwendig sind, unabhängig davon, in welchem Kontext (bspw. drittmittelfinanziertes Forschungsprojekt oder studentische Abschlussarbeit) diese entstanden sind.“ (Gollwitzer et al., 2020, S. 10)

6. Literatur

- Berger, R. (2007). Das Gruppenpuzzle am Beispiel des Mikrowellenofens. *Praxis der Naturwissenschaften*, 56 (2), 5-11.
- Berger, R. & Müller, M. (2015). Erzeugung und Übertragung elektrischer Energie. Eine Unterrichtseinheit mit Lernzirkel für die Sekundarstufe I. *Naturwissenschaften im Unterricht Physik*, 26, 11-19.
- Berger, R. & Wöllermann, J. (2022). Freischneiden. *Naturwissenschaften im Unterricht Physik (Heft 187)*, 33, 22-25.
- Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, 7(4), 400-417.

- Bortz, J. (2005). Statistik (6. Auflage). Springer.
- Burde, J.-P. & Wilhelm, T. (2021). Unterrichtskonzeptionen zu elektrischen Stromkreisen. In T. Wilhelm, H. Schecker & M. Hopf (Hrsg.), Unterrichtskonzeptionen für den Physikunterricht (231-278). Springer.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319.
- Gollwitzer, M., Abele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A. M., Schönbrodt, F. D., & Steinberg, U. (2020). Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen. <https://doi.org/10.31234/osf.io/hcxtm>
- Haagen-Schützenhöfer, C. & Hopf, M. (2020). Design-based research as a model for systematic curriculum development: The example of a curriculum for introductory optics. *Physical Review Physics Education Research*, 16, 020152.
- Hänze, M., Müller, M., & Berger, R. (2018). Cross-age tutoring: how to promote tutees' active knowledge-building. *Educational Psychology*, 38, 915-926.
- Kattmann, U., Duit, R., Gropengießer, H. & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion. Ein Rahmen für naturwissenschafts-didaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3 (3), 19-34.
- Lipowsky, F. & Rzejak, D. (2012). Lehrerinnen und Lehrer als Lerner – Wann gelingt der Rollentausch? Merkmale und Wirkungen effektiver Lehrerfortbildungen. *Schulpädagogik heute*, 5(3), 1-17.
- Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbachs α . Online Zusatzmaterial zu D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftlichen Forschung*. Berlin: Springer. https://www.researchgate.net/publication/313220515_Uberprufung_der_Konsistenz_von_Itemgruppen_mit_Cronbachs_alpha [14.02.2023]
- Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in Science Education. *Research in Science Education*, 48, 1273-1296.
- Terhart, E. (2021). Fachdidaktik Physik 2021 – ein (Rück- und Aus-) Blick von außen. Vortrag im Rahmen von „Physikdidaktik Quo vadis: Schlaglichter auf physikdidaktische Forschung“. <http://www.schulpool.uni-wuppertal.de/quo-vadis/> [14.02.2023]
- Wiesner, H. & Wodzinski, R. (1996): Akzeptanzbefragung als Methode zur Untersuchung von Lernschwierigkeiten und Lernverläufen. In R. Duit & C. von Rhöneck (Hrsg.), *Lernen in den Naturwissenschaften* (S. 250-274). Kiel: IPN.
- Zucker, D. M. (1990). An analysis of variance pitfall. The fixed effects analysis in a nested design. *Educational and Psychological Measurement*, 50, 731-738.