

Fabian Bernstein¹
 Sascha Schmeling²
 Thomas Wilhelm¹

¹ Goethe-Universität Frankfurt am Main
² CERN, Genf

Usability-Tests zur Evaluation von Experimentiermaterial

Der Research-Practice-Gap in der Physikdidaktik

Das Verhältnis zwischen didaktischer Forschung und schulischer Praxis ist seit langem Gegenstand didaktischer Reflexion, nicht selten im Modus der Kritik: Verschiedentlich wurde konstatiert und beklagt, dass didaktische Forschung („educational research“) abgehoben sei von schulischer Praxis, nicht (hinreichend) relevant und nicht implementierbar oder nicht einmal auf Implementierbarkeit bedacht (Kincheloe, 2004; Reinmann, 2005; Hirschkorn & Geelan, 2008; Burde & Wilhelm, 2018; McKenney & Reeves, 2019). Auch seien an Schulen und Universitäten so unterschiedliche Zielstellungen, Werthaltungen und Rahmenbedingungen vorzufinden, dass ein produktiver Austausch kaum zustande komme. Kincheloe meint sogar, zwei gänzlich unterschiedliche Kulturen zu erkennen, eine „craft culture“ („Handwerkskunst“) und eine „research culture“ („Forschungskultur“), die wenig Schnittmengen aufwiesen, jeweils über eigene kulturelle Codes verfügten und deren Akteure der im Grunde inkommensurablen Logik ihres jeweiligen sozialen Feldes verpflichtet seien (Kincheloe, 2004). Auch in der Physikdidaktik wurden – in voller Anerkennung der Heterogenität des Forschungsfeldes – derartige Positionen vertreten. So schreiben Wilhelm und Hopf, dass es „nicht verwunderlich sei, dass die naturwissenschaftsdidaktische Forschung häufig nur recht eingeschränkte Auswirkungen auf die Unterrichtspraxis“ habe und dass „selbst Erfolg versprechende Forschungsergebnisse [...] kaum in die Schulpraxis“ kämen (Wilhelm & Hopf, 2014). Auf der anderen Seite ist das Unbehagen mit dieser konstatierten Praxisferne didaktischer Forschung verschiedentlich als Innovationsantrieb wirksam geworden: So eröffnet beispielsweise das Positionspapier des Design-Based Research Kollektivs 2003 mit ebendieser Problembeschreibung und leitet daraus sowohl Ziele als auch Methodik des Design-based Research ab (DBRC, 2003).

Unabhängig von der Frage, ob und bis zu welchem Grad man dieser Analyse zustimmt, spricht doch einiges dafür, dass eine hohe Durchlässigkeit zwischen physikdidaktischer Forschung einerseits und unterrichtlicher Praxis andererseits ein Desiderat darstellt – mithin, dass die Frage nach der Implementierung (oder Co-Creation) von Innovationen und Forschungsergebnissen berechtigt und relevant ist. Daraus leitet sich weiter die Frage nach der methodischen Umsetzung ab: Wie kann möglichst bereits im Forschungsprozess sichergestellt werden, dass Forschungsergebnisse zu einer informierten Praxis beitragen?

Im Rahmen eines Forschungsprojektes zwischen dem CERN und der Universität Frankfurt wurde dieser Frage am Beispiel der Entwicklung von Experimentiermaterial zur modernen Physik nachgegangen. Insbesondere wurde untersucht, inwieweit das Konstrukt der „Usability“, wie es im Human-Centered Design-Framework definiert und in ISO-Normen kodifiziert ist (ISO 9241-210 Human-centered design for interactive systems, ISO 9241-11 Usability: Definitions and concepts und ISO 9241-110 Dialogue principles), dazu beitragen kann, die Wahrscheinlichkeit der Nutzung der entwickelten Materialien durch Lehrkräfte zu ma-

ximieren, um auf diese Weise einen Hebel für die Dissemination physikdidaktischer Innovation zu gewinnen.

Human-Centered Design und das Konzept der Usability

Human-Centered Design ist ein in der Industrie verbreitetes und in der ISO-Norm 9241-210 spezifiziertes Meta-Framework zur Produktentwicklung. Ausgangspunkt der Überlegung ist die Erkenntnis, dass ein Produkt nur dann erfolgreich sein kann, wenn es einen klar definierten Nutzen aus Sicht der Nutzer bietet. Aus Sicht der Produktentwicklung folgt daraus, dass eine genaue Problemanalyse – und das heißt hier: ein möglichst präzises und umfassendes Verständnis der Perspektive prospektiver Nutzer – einer Entwicklung notwendigerweise vorausgehen muss, um nicht am Bedarf vorbei zu entwickeln. Diese Überlegungen werden im Begriff des „Context of use“ zusammengefasst: Der „Context of Use“ umfasst Nutzer, deren Aufgaben, deren Ziele, die zur Verfügung stehenden Ressourcen sowie die relevanten Umgebungsfaktoren, die vor der Inangriffnahme eines Lösungsvorschlages untersucht, analysiert und in einer „Context of Use“-Description dokumentiert werden müssen.

Usability ist in diesem Framework definiert als „extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use“ (ISO 9241-210, 2019). Effectiveness, Efficiency und Satisfaction sind ebenfalls in der ISO-Norm definiert und auch operationalisiert: Metriken oder Erhebungsinstrumente werden zwar nicht vorgeschrieben, aber doch vorgeschlagen. In der Praxis haben sich bestimmte Metriken für die Erhebung dieser Konstrukte durchgesetzt: „Effectiveness“ wird häufig als „Task Success“ (Prozentsatz der erfolgreich abgeschlossenen Aufgaben) operationalisiert, „Efficiency“ als „Time on Task“ (erforderliche Zeit für die Bewältigung einer Aufgabe) und für die Messung der „Satisfaction“ stehen verschiedene validierte und gebenchmarkte Erhebungsinstrumente zur Verfügung.

Zielstellung und Ablauf eines Usability-Tests

Um die Usability von Produkten zu evaluieren, werden in der Praxis der Produktentwicklung Usability-Test eingesetzt, die eines von mehreren Verfahren darstellen, Produkte gegen die Bedürfnisse von Nutzern zu evaluieren (ISO 9241-11, ISO/TS 20282-2:2013, ISO/IEC 25062:2006). Unter einem Usability-Test versteht man eine Evaluation, in der repräsentative Nutzer repräsentative Aufgaben mit einem interaktiven System bearbeiten und der dazu dient, Usability-Probleme zu identifizieren oder die Effektivität, Effizienz und Nutzerzufriedenheit zu messen. Charakteristisch und wesentlich für Usability-Tests ist, dass es sich primär um die Beobachtung von Verhalten in einem kontrollierten Setting handelt, und **nicht** um die Befragung von Nutzern zu ihrem Nutzungsverhalten, zu Produkteigenschaften, Einschätzungen, etc.

Verschiedene Möglichkeiten der Beobachtung und Analyse im Rahmen von Usability-Tests stehen zur Verfügung: So können Usability-Probleme beobachtet, identifiziert und klassifiziert werden, nach Häufigkeit und Schwere geordnet und so als Ausgangsbasis für die Weiterentwicklung bzw. Iteration eines Produktes genutzt werden („Formative Usability Test“). Soll hingegen ein Vergleich eines zu evaluierenden Produktes mit einem Benchmark oder einem anderen Produkt im Rahmen eines A-B-Tests erfolgen, bieten sich quantitative Maße wie Task Success, Time on Task oder spezielle Fragebögen zur Erhebung der Nutzerzufriedenheit.

denheit, wie die weitverbreitete System Usability Scale oder der User Experience Questionnaire, an („Summative Usability Test“).

A-B-Vergleich zweier Michelson-Interferometer im Rahmen eines Usability-Tests

Um die Usability eines am CERN entwickelten 3D-gedruckten Michelson-Interferometers (<https://cern.ch/laserlab3D>) zu evaluieren, wurden ein A-B-Vergleich mit einem Michelson-Interferometer eines führenden deutschen Lehrmittelherstellers vorgenommen. Dabei wurden einerseits auftretende Usability-Probleme dokumentiert, andererseits quantitative Maße wie Task Success und Time on Task erhoben sowie Fragebögen (für das Satisfaction-Konstrukt) wie die Single Ease Question, die System Usability Scale und der User Experience Questionnaire eingesetzt.

Als Testpopulation wurden Physiklehrkräfte genutzt, die mit ihren Klassen an das Physik-Schülerlabor der Goethe-Universität Frankfurt kommen. Diese Lehrkräfte wurden mit Hilfe von Szenarien und Testkarten insbesondere gebeten, die Interferometer funktionstüchtig aufzubauen. Alle Lehrkräfte haben dabei mit beiden Interferometern in randomisierter Reihenfolge gearbeitet. Die Test-Sessions dauerten jeweils ca. zwei Stunden. Bisher konnten fünf Usability-Tests abgeschlossen werden, wobei beabsichtigt ist, insgesamt zwanzig Usability-Tests durchzuführen (zur Stichprobengröße siehe z.B. Sauro & Lewis, 2016).

Erste qualitative Ergebnisse der Usability-Tests

Die bisher durchgeführten Usability-Tests lassen bereits wiederkehrende Problemfelder erkennen. Bei dem kommerziellen Lehrmittel wird bspw. deutlich, dass eine ausschließliche Zurverfügungstellung einer englischsprachigen Anleitung von einigen Lehrkräften kritisch gesehen wird. Auch die Ausgestaltung dieser Anleitung ist in mehrerlei Hinsicht problematisch: So führt die Verdichtung der schematischen Abbildung des Aufbaus dazu, dass mehrere mögliche Aufbauten in einer Grafik zusammenfasst wurden. Dies bedeutet, dass nicht alle abgebildeten Bauteile für den Aufbau des Michelson-Interferometers erforderlich sind. Darüber hinaus finden sich in der Anleitung nicht erklärte Abkürzungen, die eine Quelle der Verunsicherung der Lehrkräfte darstellen. Des Weiteren sind im Standard-Lehrmittel die optischen Halter so gestaltet, dass der Laser nicht über den auf der optischen Grundplatte abgebildeten Linien verläuft, was ebenfalls zur Verunsicherung der Lehrkräfte beiträgt.

Eine schwerwiegende Fehlerquelle sind die unterschiedlichen Längen der Spiegelhalter: Die vier erforderlichen Spiegel weisen drei verschiedene Längen auf, wobei aus der Anleitung nicht hervorgeht, welcher Spiegel an welche Stelle gesetzt werden muss. Dies führt zu Trial-and-Error, Ineffizienz und Frustration bei den Lehrkräften. Schließlich ist als problematisch anzusehen, dass der Höhenverstellbereich der zuletzt einzusetzenden Linse so klein ist, dass meistens eine komplette Neujustage des Aufbaus erforderlich ist, da erst im letzten Schritt ersichtlich wird, dass die Linse nicht auf der richtigen Höhe in den Strahlengang eingebracht werden kann.

Für das 3D-gedruckte Interferometer hat sich gezeigt, dass die Form des Laserpointer-Halters zu Fehlbedienungen führte, da aus der Gestalt des Halters selbst seine Verwendung nicht klar hervorgeht (mangelnde „Affordance“). Ferner führte der konstruktiv begrenzte Verstellbereich der Spiegelhalter zu Schwierigkeiten bei der Justage des Aufbaus.

Insgesamt lassen sich bereits aus diesen ersten qualitativen Ergebnissen Erkenntnisse für eine weitere Optimierung der Aufbauten im Sinne einer Maximierung der Usability gewinnen.

nen. Ob eine erhöhte Usability darüber hinaus zu einer wahrscheinlicheren Nutzung der Experimente im Unterricht durch die Lehrkräfte führte, ist – wenngleich plausibel – durch weitere empirische Untersuchungen zu stützen.

Literatur

- Burde, J.-P., & Wilhelm, T. (2018). Einfache Stromkreise mit Potenzial. Ein neues Unterrichtskonzept zeigt, wie fachdidaktische Entwicklungsforschung arbeitet. *Physik Journal*, 17 (5), 27-30
- Hirschhorn, M. & Geelan, D. (2008). Bridging the Research-Practice Gap: Research Translation and/or Research Transformation, 54. ISO 9241-11 (2018): Usability: Definition and concepts
- ISO 9241-11 (2018). Usability: Definition and concepts
- ISO 9241-110 (2006). Dialogue principles
- ISO 9241-210 (2019). Human-centred design for interactive systems
- ISO/TS 20282-2 (2013). Usability of consumer products and products for public use — Part 2: Summative test method
- ISO/IEC 25062 (2006). Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports
- Kincheloe, J.L., (2004). The bizarre, complex, and misunderstood world of teacher education. In J.L. Kincheloe, A. Bursztyrn, & S. Steinberg (Hrsg.), *Teaching teachers: Building a quality school of urban education*, 1-49, New York: Peter Lang.
- McKenney, S. & Reeves, T. C. (2019). *Conducting educational design research*, 2. Aufl. Routledge. <https://doi.org/Susan>
- Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. *Unterrichtswissenschaft*, 33 (1), 52–69.
- Sauro, J., & Lewis, J. R. (2016). *Quantifying the User Experience: Practical Statistics for User Research*, 2. Aufl. Cambridge: Morgan Kaufmann.
- The Design-Based Research Collective (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32 (1), 5–8. <https://doi.org/10.3102/0013189X032001005>
- Wilhelm, T. & Hopf, M. (2014). Design-Forschung. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.): *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin & Heidelberg: Springer.