

Usability Tests zur Evaluation physikdidaktischer Entwicklungen

Skalierung physikdidaktischer Innovationen

Die Skalierung fachdidaktischer Innovation und der Transfer aus den geschützten Räumen von Universitäten und Forschungsprojekten in die schulische Praxis stellt seit vielen Jahren eine zentrale Herausforderung für die physikdidaktische und, allgemeiner, bildungswissenschaftliche Forschung dar und wird häufig unter dem Begriff der Transferforschung oder kritisch als „Research-Practice-Gap“ diskutiert (Kincheloe, 2004; Hirschhorn & Geelan, 2008; Gräsel, 2010; Gräsel, 2011; McKenney & Reeves, 2019; Schmiedebach & Wegner, 2021). Antworten auf die Frage zu finden, wie die „Verbreitung wissenschaftlich fundierter Innovationen im Bildungssystem“ (Gräsel 2010) gelingen kann, ist von entscheidender Bedeutung, wenn die von der didaktischen Forschung ausgehenden Impulse über den universitären Bereich hinaus in der Praxis wirksam werden und nachhaltige Veränderungen und Verbesserungen der Unterrichtspraxis anstoßen sollen. Auch in naturwissenschaftsdidaktischen Publikationen der jüngeren Zeit wurde diese Herausforderung aufgegriffen, Implementationsbarrieren untersucht und das Problem der nachhaltigen Verankerung didaktischer Innovationen in der schulischen Praxis thematisiert (Breuer et al., 2020; Breuer, 2021; Neff, 2023; Knebloch et al., 2024).

Human-Centered Design: Grundlagen und Prozesse

Ein Ansatz, der bei einem Forschungsprojekt am CERN in Kooperation mit der Goethe-Universität Frankfurt verfolgt wurde, ist, bewährte Implementationsstrategien aus anderen Disziplinen in den Blick zu nehmen und auf ihre Eignung im physikdidaktischen Kontext hin zu prüfen. Ein vielversprechender Kandidat für eine solche Implementierungsstrategie ist „Human-Centered Design“, ein in ISO-Normen kodifiziertes Industrie-Framework, das sich mit der Usability von Produkten und, mittelbar, dem Transfer von Innovationen in gesellschaftliche Praxis unter besonderer Berücksichtigung der Zielgruppen und der Rahmenbedingungen im jeweiligen Feld beschäftigt (ISO 9241-210, 2019; ISO 9241-11, ISO/TS 20282-2:2013). Für Human Centered Design spricht u.a., dass es in erheblicher Breite im industriellen Umfeld erprobt wurde, über ein begrifflich und konzeptionell ausgearbeitetes Fundament verfügt, zahlreiche validierte Instrumente und konsolidierte Methoden bereitstellt und von einer Community getragen wird, die sowohl wissenschaftlich verankert als auch im Anwendungsumfeld mit praktischen Implementationsproblemen von ökonomischer Relevanz konfrontiert ist.

Der Human-Centered Design-Prozess verläuft typischerweise in vier iterativen Stadien: Ausgangspunkt des Human-Centered Design ist eine genaue Kenntnis und Analyse der Zielgruppe oder Zielgruppen: ihrer Ziele, ihrer Aufgaben und insbesondere auch des Implementationskontextes, in dem sich eine Innovation oder ein Produkt bewähren muss. Dieser Analyseschritt ist essenziell und beruht auf der Einsicht, dass ohne eine genaue Kenntnis der Zielgruppen und der Kontexte der Produktverwendung – d.h. bspw. der Lehrkräfte und ihres schulischen Umfelds – Innovationen zum Scheitern verurteilt sind, die keine hinreichende Passung zu den Anforderungen aufweisen, die sich aus den typischen Persönlichkeitsmerkmalen und Anwendungskontexten ergeben. In einem zweiten Schritt werden Nutzungsanforderungen abgeleitet, die die Grundlage für das Design möglicher Lösungen, Produkte oder Innovationen im dritten

Schritt darstellen. Der vierte Schritt besteht in der Evaluation der entwickelten Lösungen vor dem Hintergrund der Nutzungsanforderungen: Ist das Produkt geeignet, die Ziele und Aufgaben der Zielgruppe unter Minimierung des Ressourceneinsatzes und bei Maximierung der Nutzerzufriedenheit zu erfüllen? Dies sind nach ISO 9241-210 die Schlüsselfaktoren der Usability und werden im Framework als Effektivität, Effizienz und Nutzerzufriedenheit definiert.

Usability-Tests mit Experimentiermaterial: Ablauf, Konstrukte, Operationalisierung

Zur Evaluation der Usability eines Produktes stehen verschiedene Methoden zur Verfügung. Als Goldstandard gelten Usability-Tests, also teilstandardisierte Untersuchungen, in denen repräsentative Nutzer bei der Lösung repräsentativer Aufgaben mittels eines interaktiven Systems beobachtet werden und die dazu dienen, Effektivität, Effizienz und Nutzerzufriedenheit zu messen. Usability-Tests basieren also auf der Beobachtung von Interaktionen des Nutzers mit dem zu evaluierenden System, was im Allgemeinen zu Einsichten führt, die sich ausschließlich auf dem Wege der Befragung nicht gewinnen lassen. In der Usability-Literatur wird dabei zwischen „formativen“ und „summativen“ Usability-Tests unterschieden (vgl. z. B. Rubin & Chisnell, 2008): Während formative Usability-Tests vorrangig während des Entwicklungszyklus eines Produktes eingesetzt werden und die Weiterentwicklung der Produkte durch die Systematisierung häufig auftretender Interaktionsdefekte („Usability Issues“) leiten, werden summative Usability-Tests zur Erhebung quantitativer Kennzahlen für A-B-Tests und Benchmarks eingesetzt.

Im Human-Centered Design Projekt von CERN und Goethe-Universität Frankfurt wurde exemplarisch der Frage nachgegangen, ob bzw. wie es gelingen kann, Experimente zur modernen Physik für einen zeitgemäßen Physikunterricht aufzubereiten. Dies ist nicht zuletzt getrieben von der Beobachtung, dass Verfahren und Experimente der modernen Physik bisher insgesamt wenig Eingang in die Unterrichtspraxis gefunden haben.

Im Rahmen des Projektes wurde sowohl das Experimentierverhalten von Lehrkräften in einer Dreiländerstudie auf Basis des Reasoned Action Approaches untersucht (Bernstein et al., 2020) als auch konkrete Experimentieraktivitäten entwickelt. Eines der entwickelten Experimente – ein 3D-gedrucktes Michelson-Interferometer – wurde einem vergleichenden Usability-Test an der Goethe-Universität Frankfurt unterzogen. Die Stichprobe (N = 20) bestand aus Lehrkräften, die mit ihren Klassen an Angeboten des Physik-Schülerlabors der Goethe-Universität Frankfurt teilgenommen haben. Parallel zum Schülerlabor führten die Lehrkräfte in einem separaten Raum den Usability-Test durch und erhielten die Aufgabe, nacheinander zwei Michelson-Interferometer – das 3D-gedruckte Experiment sowie ein Interferometer eines Lehrmittelherstellers – aufzubauen und mit den Aufbauten ein Interferenzmuster zu erzeugen. Die Reihenfolge, in der die Lehrkräfte mit den Interferometern konfrontiert wurden, war zwischen den Lehrkräften randomisiert; die Gesamtdauer eines Usability-Tests betrug im Mittel ca. 2,5 Stunden.

Um ein umfassendes Bild der Usability der Experimente zu erhalten, wurden sowohl qualitative Daten zu Usability-Problemen erhoben als auch quantitative Instrumente eingesetzt. Alle Usability-Tests wurden videographiert, um die Protokolle der Usability-Tests ex-post verifizieren zu können. Die detaillierte Operationalisierung der Konstrukte und der Ablauf der Usability-Tests wird aus Abbildung 1 ersichtlich: Während Task Success und Time on Task unmittelbar der Beobachtung zugänglich sind, muss zur Messung der latenten Nutzerzufriedenheit auf Befragungsinstrumente zurückgegriffen werden. Neben der „Single Ease Question“ zur Messung der wahrgenommenen Aufgabenschwierigkeit kamen die System Usability Scale (Lewis, 2018) und der User Experience Questionnaire (Laugwitz et al., 2008; Schrepp

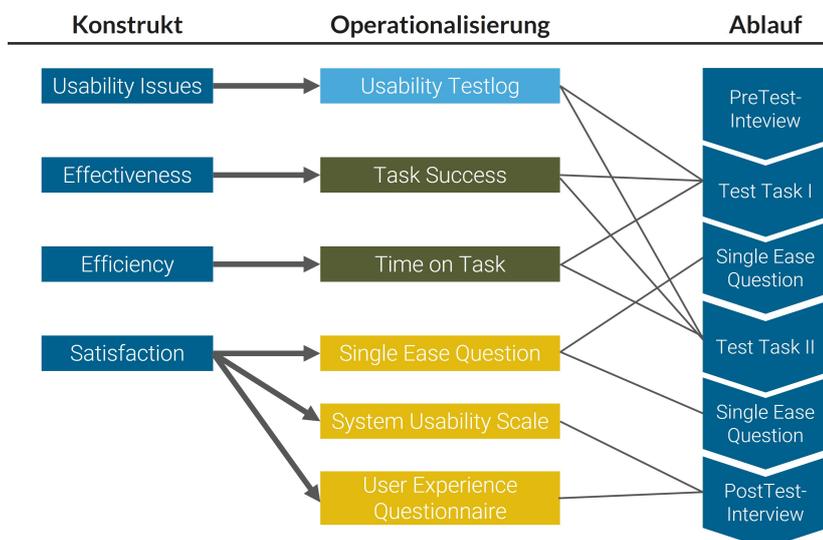


Abb.1: Konstrukte, Operationalisierung und Ablauf der Usability-Tests

et al., 2017) zum Einsatz. Die System Usability Scale erlaubt dabei ein robustes Benchmarking, da sie seit Jahrzehnten für die Evaluation von Produkten eingesetzt wird (Sauro, 2011; Lewis & Sauro 2018), wohingegen der User Experience Questionnaire mit seinen fünf Subskalen „Durchschaubarkeit“, „Effizienz“, „Steuerbarkeit“, „Stimulation“ und „Originalität“ differenziertere Einblicke in die subjektiv empfundene pragmatische und hedonische Qualität eines Produktes ermöglicht (Schrepp, 2015).

Resultate und Ausblick

Die Auswertung der Usability-Tests zeigt, dass signifikante Unterschiede in allen Usability-Dimensionen zwischen den beiden Interferometern bestehen. So ist mit Blick auf die Effektivität festzustellen, dass das Standard-Lehrmittel nur von fünf Lehrkräften in der gegebenen Zeit erfolgreich aufgebaut werden konnte, während dies beim 3D-gedruckten Lehrmittel elf Lehrkräften gelang. Auch bei der Effizienz werden klare Unterschiede deutlich: Während die erfolgreichen Lehrkräfte für die Einrichtung des Standard-Lehrmittels im Median ca. 29 Minuten benötigten, gelang das für das 3D-gedruckte Interferometer in der Hälfte der Zeit. Signifikante Unterschiede zeigen sich schließlich in der System Usability Scale und im User Experience Questionnaire. Für das Standard-Lehrmittel ergab sich im SUS ein Mittelwert von ca. 41 Punkten – dieser Wert wird in der Usability-Literatur als „inakzeptabel“ interpretiert (Sauro, 2011). Das 3D-gedruckte Interferometer erreichte im Mittel ca. 76 Punkte, entsprechend einer „guten“ Usability. Die Befunde im User Experience Questionnaire sind analog. Die vorliegenden Resultate deuten auf eine deutlich bessere Usability des 3D-gedruckten Interferometers gegenüber dem Standard-Lehrmittel hin. Offen bleibt jedoch, ob die gesteigerte Usability des 3D-gedruckten Experiments in der Praxis zu einer höheren Adoptionsrate und letztlich zu einem häufigeren Unterrichtseinsatz führen würde. Ein Zusammenhang zwischen der Usability von Experimenten und ihrem Unterrichtseinsatz erscheint zwar plausibel, ist aber empirisch nachzuweisen, zumal davon auszugehen ist, dass in die Entscheidung für oder gegen einen Einsatz im Unterricht weitere Faktoren einfließen dürften.

Literatur

- Bernstein, F., Schmeling, S., Wilhelm, T., & Woithe, J. (2020). Saliente Überzeugungen von Physiklehrkräften zum Experimentieren. *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen*, 86.
- Breuer, J., Vogelsang, C., & Reinhold, P. (2020). Implementation und Nutzung von Unterrichtsmaterialien im schulischen Unterricht – Eine Bestandsaufnahme der mathematisch-naturwissenschaftlichen Fächer. *PhyDid A - Physik und Didaktik in Schule und Hochschule*, 1(19), Article 19.
- Breuer, J. (2021). Implementierung fachdidaktischer Innovationen durch das Angebot materialgestützter Unterrichtskonzeptionen. Fallanalysen zum Nutzungsverhalten von Lehrkräften am Beispiel des Münchener Lehrgangs zur Quantenmechanik. Logos Verlag Berlin. <https://doi.org/10.30819/5293>
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13(1), 7–20. <https://doi.org/10.1007/s11618-010-0109-8>
- Gräsel, C. (2011). Die Verbreitung von Innovationen als Aufgabe der Unterrichtsforschung. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung: Traditionslinien und Perspektiven* (S. 320–328). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-94025-0_23
- 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
- Kincheloe, J. L. (2004). The bizarre, complex, and misunderstood world of teacher education. In J.L. Kincheloe, A. Burszty, & S. Steinberg (Hrsg.), *Teaching teachers: Building a quality school of urban education*, 1-49, New York: Peter Lang.
- Kneblach, J., Hengel, C., Kyas, J., Hansch, A., & Wilhelm, T. (2024). Mangelnde Implementation physikdidaktischer Innovationen. In H. Grötzebauch, & S. Heinicke (Hrsg.): *PhyDid B, Didaktik der Physik*.
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Hrsg.), *HCI and Usability for Education and Work* (S. 63–76). Springer. https://doi.org/10.1007/978-3-540-89350-9_6
- Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. *Journal of Usability Studies*, 13(3).
- Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, 34(7), 577–590. <https://doi.org/10.1080/10447318.2018.1455307>
- McKenney, S. & Reeves, T. C. (2019). *Conducting educational design research*, 2. Aufl. Routledge. <https://doi.org/Susan>
- Neff, S. (2023). *Transfer digitaler Innovationen in die Schulpraxis: Eine explorative Untersuchung zur Förderung der Implementation*. Logos Verlag Berlin. <https://doi.org/10.30819/5687>
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests* (2nd ed). Wiley Pub.
- Sauro, J. (2011). *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Measuring Usability LLC.
- Schmiedebach, M., & Wegner, C. (2021). Design-Based Research als Ansatz zur Lösung praxisrelevanter Probleme in der fachdidaktischen Forschung. *Bildungsforschung*, 2, 1–10.
- Schrepp, M. (2015). *User Experience Questionnaire Handbook*. <https://doi.org/10.13140/RG.2.1.2815.0245>
- Schrepp, M., Thomaschewski, J., & Hinderks, A. (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). <https://doi.org/10.9781/ijimai.2017.445>