

Die Rolle der Usability für das Lernen mit digitalen Anwendungen

Einleitung

Durch die zunehmende Performanz moderner Large Language Modelle (LLMs) und anderer Anwendungen im Bereich der künstlichen Intelligenz (KI) können digitale Lernanwendungen effizienter entwickelt werden. Gleichzeitig werden Plattformen, wie z. B. Fobizz genutzt, um LLMs im Schulalltag einzusetzen. Ein wichtiges Gütekriterium einer jeden digitalen (Lern-)Anwendung ist die Usability. Sie umfasst beispielsweise, wie effizient eine bestimmte Aufgabe erledigt werden kann oder wie zufrieden die Nutzer:innen mit der Anwendung sind. Die Usabilityforschung ist ein etabliertes Forschungsfeld in der Human-Computer-Interaction (HCI). Viele Studien finden mit Studierenden oder Erwachsenen im Bereich der Informatik statt (z. B. Yuhui et al., 2019). Wenige Untersuchungen fokussieren sich bisher explizit auf die Rolle der Usability bei Lernprozessen und im Einsatz in der Schule. Die bisher durchgeführten Studien in der Biologie (Karapanos et al., 2018; Koć-Januchta et al., 2022) zeigen positive Zusammenhänge zwischen Interesse und Usability sowie Germane Load und negative Zusammenhänge mit dem Intrinsic und Extraneous Load. Dies verdeutlicht die Rolle der Usability beim Lernen. Studien im Fach Chemie fehlen bisher jedoch weitestgehend. Der Beitrag wird daher die aktuelle Forschung in diesem Feld näher beleuchten und Erkenntnisse aus einer Studie vorstellen, in der die Usability von digitalen Arbeitsblättern näher untersucht wurde und dadurch Forschungsdesiderate abgeleitet werden sollen.

Theoretische Einordnung

Begriffseinordnung

Die Usability wird häufig unter dem Begriff „Benutzerfreundlichkeit“ geführt. Dieser Begriff geht auf Nielsen (1993) zurück. Heutzutage ist die Usability nach der ISO-Norm 9241-11 (ISO, 2018) als „Gebrauchstauglichkeit“ definiert. Der Begriff verdeutlicht bereits, dass die Nützlichkeit im Fokus der Usability steht. Demnach werden verschiedene Anwendungsattribute zur Usability zusammengefasst. Nach Barnum (2021), Bevan et al. (2015) und Sulaiman & Mustafa (2019) lassen sich folgende Merkmale zusammenfassen: Effektivität und Nützlichkeit, Zufriedenheit, Effizienz und Nutzungskontext.

Die Usability ist hierbei von der User Experience abzugrenzen. Letztere umfasst auch die Ästhetik und die wahrgenommenen Emotionen der Nutzer:innen. Die Funktionalität spielt jedoch auch hierbei eine Rolle.

Messung der Usability

Die Usability lässt sich über verschiedene Methoden bestimmen. Je nach Verfahren können entweder Expert:innen oder Novizen die Usabilitymessung durchführen. Erstere können beispielsweise heuristische Evaluationen durchführen, während bei letzteren Fragebogen oder Thinking-Aloud-Techniken eingesetzt werden (Holzinger, 2015). Fragebögen haben sich hierbei als ein etabliertes Messinstrument der Usabilityforschung herausgestellt. Es existiert eine Vielzahl solcher Fragebögen, wie beispielsweise der USE (Lund, 2001) oder der SUS (Brooke, 1996). Beide Instrumente wurden auch bereits psychometrisch validiert. Fragebögen sind insbesondere zeit- und kostenökonomisch einzusetzen. Gleichzeitig können gegenüber qualitativen Methoden höhere Stichproben und damit auch eine breite Usabilitybewertung

erreicht werden. Bei speziellen Problemen sind allerdings differenzierte Rückmeldungen der Nutzer:innen nötig.

Einflussfaktoren auf die Usabilitymessung

In Bezug auf die Erhebung der Usability sind einige Dinge zu berücksichtigen. Die Literatur zeigt, zumindest für ältere Anwender:innen, dass ein Einfluss des Messzeitpunktes besteht. Demnach kann die Usability während der Nutzung eines Produkts tendenziell ab- oder zunehmen. Studien hierzu zeigen keine einheitlichen Ergebnisse. Einige Studien zeigen nur einen geringen Einfluss des Messzeitpunktes, während andere Studien eine Zu- oder Abnahme verzeichnen (Wang, 2021; Yuhui et al., 2019). Eine Zunahme der Usability mit zunehmender Nutzungsdauer erscheint nachvollziehbar, da Nutzer:innen vertrauter mit einzelnen Funktionen der Anwendung sind und zunehmend einzelne Aufgaben souveräner durchführen. Eine Erklärung für die Abnahme der Usability über einen Messzeitraum könnte mit einem negativen Ereignis während der letzten Nutzung in Zusammenhang gebracht werden. Nach den in der Literatur beschriebenen Effekten der „Peak-End-Rule“ oder dem „Recency-Effect“ werden vor allem Ereignisse bewertet, die entweder ein Etappenziel darstellen oder nahe am Bewertungszeitpunkt liegen. Demnach bewerten Nutzer:innen die Usability nicht gleichmäßig über den gesamten Zeitraum der Nutzung, sondern sie ziehen besonders wichtige Ereignisse und besonders junge Erfahrungen in die Bewertung mit ein.

Auch das Geschlecht nimmt Einfluss auf die Bewertung der Usability. Tendenziell bewerten Anwenderinnen die Usability schlechter als Anwender (Karapanos et al., 2018; Ong & Lai, 2004). Dies könnte mit der technikbezogenen Selbstwirksamkeit zusammenhängen (Ong & Lai, 2004). Es finden sich jedoch in der Literatur auch Beispiele, in denen keine Unterschiede festgestellt wurden (Ibli & Billingham, 2019).

Verschiedene Studien verdeutlichen die Relevanz der Usability beim Lernen mit digitalen Anwendungen. Studien in verschiedensten Settings zeigen hier Zusammenhänge zum Interesse (Karapanos et al., 2018), der Motivation (Álvarez-Xochihua et al., 2017), dem Cognitive Load (Koć-Januchta et al., 2022) oder auch zu Emotionen (Stefancova et al., 2018) – Variablen, die beim Lernen eine wichtige Rolle spielen.

Usability digitaler Arbeitsblätter

Methodik

Es wurde eine quasi-experimentelle Feldstudie über vier reguläre Unterrichtsstunden in der Mittel- und Oberstufe im Fach Chemie durchgeführt. Die Schüler:innen lernten in dieser Zeit mit speziellen digitalen binnendifferenzierten Arbeitsblättern, sogenannten HyperDocs (Fitting et al., 2023). Fachinhalte waren Farbstoffe als Zusatzstoffe, Auswertung von Messreihen sowie der Konzentrationsbegriff (Mittelstufe) bzw. das Prinzip der Photometrie (Oberstufe).

Die digitalen Arbeitsblätter wurden auf iPads der 10. Generation bearbeitet. Dazu wurden Geräte mit einem Apple Pencil bei jeder Erhebung zur Verfügung gestellt. Die Schüler:innen hatten keinen Zugang zum Internet, da ein lokales W-LAN eingesetzt wurde.

Nach jeder Stunde wurden Cognitive Load (Klepsch et al., 2017) und Variablen der intrinsischen Motivation nach dem *Intrinsic Motivation Inventory* erhoben. Die Usability wurde nach einer gekürzten Form des USE (Lund, 2001) erhoben. An der Studie nahmen über 300 Mittelstufenschüler:innen (2/3 Gymnasium und 1/3 Gesamtschule) und über 100 Oberstufenschüler:innen (überwiegend Gymnasium) teil. Die Geschlechter waren gleich verteilt. Aufgrund von Fallausschlüssen ergeben sich folgende Stichprobengrößen: Mittelstufe

(Cognitive Load: 300, Alter: 14.73; Motivation: 306, Alter: 14.73) und Oberstufe (Cognitive Load: 100, Alter: 16.93; Motivation: 101, Alter: 16.91).

Ergebnisse und Diskussion

Die Zusammenhänge (Spearman's Rho) zwischen Usability (Nützlichkeit und Zufriedenheit) und Motivation (Interesse / Spaß und Kompetenzerleben) bzw. Cognitive Load wurden nach der ersten und vierten Stunde bestimmt. Im Nachfolgenden werden die insgesamt gefundenen Zusammenhänge als Wertebereich für die Mittel- und Oberstufe sowie die beiden Messzeitpunkte angegeben.

Tabelle 1: Ergebnisse der Korrelationsanalysen nach Spearman's rho.

	Nützlichkeit	Zufriedenheit
Intrinsic Load	-.09 bis -.30**	-.07 bis -.32**
Extraneous Load	-.15 bis -.31**	-.22** bis -.31**
Germane Load	.09 bis .42**	.32** bis .47**
Interesse	.36** bis .47**	.57** bis .77**
Kompetenzerleben	.20* bis .40**	.29** bis .49**

* $p < 0.05$, ** $p < 0.01$

Die Zusammenhänge zum Extraneous und Germane Load lassen sich theoretisch begründen. Eine ungeordnete, unübersichtliche Benutzeroberfläche mit häufigen Fehlern kann demnach den Extraneous Load erhöhen, während der Germane Load verringert wird. Gleichzeitig kann dadurch auch das Erleben von Kompetenz oder Autonomie abnehmen. Letztlich führt dies zu einem verminderten Interesse. Eine gut strukturierte Benutzeroberfläche hingegen, mit Hinweisen und Tooltips, mit denen Nutzer:innen Aufgaben effizient erledigen können, kann den Germane Load und die Variablen der Motivation erhöhen. Damit könnte sowohl ein direkter Einfluss der Usability auf die Motivation als auch eine Mediation über den Cognitive Load stattfinden. Die Rolle des Intrinsic Loads hingegen ist theoretisch weniger leicht einzuordnen. Letztlich kann jedoch keine Aussage über die Wirkungsrichtung der Variablen getroffen werden.

Fazit und Ausblick

Die Ergebnisse der bisherigen Forschung im Schulunterricht konnten auch für ein Einsatzbeispiel im Chemieunterricht und mit digitalen Arbeitsblättern bestätigt werden. Demnach bestehen mittelstarke Zusammenhänge zum Interesse / Spaß, Kompetenzerleben und Cognitive Load. Die Wirkungsrichtung der Zusammenhänge bleibt allerdings offen. Die Usability ist demnach ein wichtiges Attribut beim Lernen mit digitalen Anwendungen, wie beispielsweise digitalen Arbeitsblättern, und sollte bei einer Erhebung immer in Betracht gezogen werden.

In Ausblick zu stellen sind gezielte Interventionen und Beobachtungsstudien, um die genaue Wirkung der Usability zu ergründen, um ein besseres Verständnis für die Wirkung der Motivation zu entwickeln. Ferner sind besonders nützliche Attribute einer Anwendung zum Lernen zu identifizieren, um Best-Practice-Ansätze zu formulieren und bestehende Heuristiken zu erweitern. Durch den Einzug von LLMs in das Bildungswesen ist die Evaluation der Usability verschiedener Chatbot-Plattformen durch Schüler:innen ebenfalls ein wichtiges Desiderat. Ferner ist zu prüfen, wie die Integration von LLMs in bestehende Anwendungen die Usability beeinflusst. Eine neuere Studie zeigt beispielsweise, dass die Vertrauenswürdigkeit von Chatbots von deren Usability abhängig ist (Can & Nguyen, 2025).

Literatur

- Álvarez-Xochihua, O., Muñoz-Merino, P., Muñoz-Organero, M., Kloos, C. & González-Fraga, J. (2017). Comparing usability, user experience and learning motivation characteristics of two educational computer games. *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) - Volume 3*. SCITEPRESS, 143-150.
- Barnum, C. M. (2021). *Usability testing essentials: Ready, set...test!* (2nd Edition). Morgan Kaufmann.
- Bevan, N., Carter, J. & Harker, S. (2015). ISO 9241-11 Revised: What have we learnt about usability since 1998? In M. Kurosu (Eds.), *Human-Computer Interaction: Design and Evaluation*. Heidelberg: Springer International Publishing, 143-151.
- Brooke, J. (1996). SUS: A „Quick and Dirty“ Usability Scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability Evaluation In Industry*. CRC Press, 189-194.
- Can, V.-D., & Nguyen, V.-H. (2025). The relationship between perceived usability and perceived credibility of middle school teachers in using AI chatbots. *Cogent Education*, 12(1).
- Fitting, N., Ulber, R., Czubatinski, L. & Hornung, G. (2023). Evaluation digitaler Arbeitsblätter im Chemieunterricht in Hinblick auf Usability und Interesse. In J. Roth, M. Baum, K. Eilerts, G. Hornung, & T. Trefzger (Eds.), *Die Zukunft des MINT-Lernens – Band 2: Digitale Tools und Methoden für das Lehren und Lernen*. Berlin: Springer, 93 – 108.
- Holzinger, A. (2005). Usability engineering methods for software developers. *Communications of the ACM*, 48(1), 71–74.
- Ibili, E. & Billinghamurst, M. (2019). Assessing the relationship between cognitive load and the usability of a mobile augmented reality tutorial system: A study of gender effects. *International Journal of Assessment Tools in Education*, 6(3).
- ISO. (2018). *ISO 9241-11:2018(en), Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>
- Karapanos, M., Becker, C. & Christophel, E. (2018). Die Bedeutung der Usability für das Lernen mit digitalen Medien. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 32, 36–57.
- Klepsch, M., Schmitz, F. & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8, 1–18.
- Koć-Januchta, M. M., Schönborn, K. J., Roehrig, C., Chaudhri, V. K., Tibell, L. A. E. & Heller, H. C. (2022). "Connecting concepts helps put main ideas together": Cognitive load and usability in learning biology with an AI-enriched textbook. *International Journal of Educational Technology in Higher Education*, 19, 11.
- Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8(2), 3–6.
- Nielsen, J. (1993). *Usability Engineering*. Ap Professional.
- Ong, C.-S. & Lai, J.-Y. (2006). Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Computers in Human Behavior*, 22(5), 816–829.
- Sulaiman, W. N. A. W. & Mustafa, S. E. (2020). Usability elements in digital textbook development: A systematic review. *Publishing Research Quarterly*, 36(1), 74–101.
- Stefancova, E., Moro, R. & Bielikova, M. (2018). Towards detection of usability issues by measuring emotions. In M. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, & M. Winckler (Eds.), *Human-computer interaction – INTERACT*. Heidelberg: Springer, 60-70.
- Wang, Y. (2021). Gaps between continuous measurement methods: A longitudinal study of perceived usability. *Interacting with Computers*, 33(3), 223–237.
- Yuhui, W., Chunfu, L., Tian, L. & Qin, H. (2019). Longitudinal study on retrospective assessment of perceived usability: A new method and perspectives. *Interacting with Computers*, 31(4), 393–412.