

Wirkung eines KI-basierten Sprachassistenzsystems auf die kognitive Belastung von Schüler*innen beim chemischen Experimentieren

Einführung

Im Projekt *Artificial Intelligence for Chemistry Lab Learning (ArtiflCheL₂)* wird die Wirkung eines KI-gestützten Sprachassistenzsystems (*LabTwin*) auf die kognitive Belastung durch instruktives Prompting beim experimentellen Arbeiten im Fach Chemie untersucht. *LabTwin* ist eine Software, in der Experimentiervorschriften hinterlegt und per Sprachbefehl angesteuert werden können. Die Aufzeichnungen von Beobachtungen und Messwerterfassungen werden mit einem Zeitstempel hinterlegt und nach der Laborarbeit automatisch zu einem vorstrukturierten und editierbaren Protokoll zusammengeführt. Damit entfällt der Aufmerksamkeitswechsel zwischen der Arbeitsvorschrift und dem eigentlichen Handlungsschritt. Der Beitrag berichtet aus der Vorstudie, die die Kausalmodellierung der theoretischen Zusammenhänge von intrinsischer kognitiver Belastung mit Vorwissen, einen simulierten Datensatzes und die Beobachtungsdaten aus einer Klassenraumerhebung kombiniert.

Theorie

Beim Anleiten von Noviz*innen im Labor ist ein schrittweises Vorgehen angezeigt. Das liegt an der kognitiven Architektur von Menschen, die mit der Cognitive Load Theory (Sweller, 2010, 2016) beschrieben wird. Dort wird zwischen dem Langzeitgedächtnis (LG) und dem Arbeitsgedächtnis (AG) unterschieden. Das LG zeichnet sich dadurch aus, dass Wissen schnell und mühelos abgerufen werden kann. Im AG werden neue Informationen mit dem LG verknüpft. Damit nimmt das AG beim Design von Lernumgebungen eine Schlüsselfunktion ein. Im Gegensatz zum LG kann es nur wenige Einzelinformationen gleichzeitig behalten und ist störanfällig. Drei Typen von Cognitive Load (CL) werden unterschieden und als Intrinsic (ICL), Extrinsic (ECL) und Germane (GCL) Cognitive Load bezeichnet. ICL ist die Belastung, die durch die Merkmale der zu bearbeitenden Aufgabe auftritt und die kaum reduzierbar ist. Insofern ist die Schätzung von ICL beim Einsatz von mediengestützten Instruktionen ein wichtiger Indikator für die Einsatzmöglichkeiten bei Noviz*innen. ECL kann durch Lehrkräfte direkt beeinflusst werden. Dazu gehört beispielsweise die Reduktion unnötiger Zusatzinformationen. GCL bezeichnet eine gewünschte Belastung des AG, in der die aktive Verknüpfung von neuen Informationen mit dem LG stattfindet. In der Theorie sind ICL und GCL außerdem mit dem domänenspezifischen Vorwissen verbunden (Klepsch et al., 2017). Je höher das Vorwissen, desto mehr Ressourcen können aus dem LG abgerufen und mit den Informationen im AG verknüpft werden. Mit hohem Vorwissen ist ein Teil der zu verarbeitenden Informationen einer Aufgabe nicht mehr neu, muss also nicht mit Mühe im AG behalten werden und lässt somit Kapazitäten frei. Folglich ist ein positiver Zusammenhang zwischen Vorwissen und GCL sowie ein negativer Zusammenhang zwischen ICL und Vorwissen zu erwarten. Zur Schätzung des ICL bei Experimentieranweisungen durch *LabTwin* ergeben sich die folgenden Forschungsfragen: 1. *Wie verändert sich der wahrgenommene ICL von Lernenden beim Einsatz von LabTwin?* 2. *In welchem Zusammenhang steht das ICL mit dem Vorwissen der Lernenden?*

Design

Vier Facetten wurden für die Studie im Sinne eines Value-added Design bei der Untersuchung digitaler Tools (Buchner & Kerres, 2023) besonders betont. A) Die Versuchsvorschriften für die grundlagenorientierte Feldstudie wurden gemeinsam mit den Lehrkräften er- und überarbeitet. B) Es sollten Prozessdaten für den CL-Verlauf im Klassenraum erhalten werden. C) Eine Kausalanalyse mit *Directed Acyclical Graphs* (DAGs; Cinelli et al., 2022; Rohrer, 2018; Weidlich et al., 2023) sollte mit einem bayesianischen Mehrebenenmodell kombiniert werden, das durch die natürlichen Cluster von Teilnehmenden, Klassen und Schulen impliziert ist. D) Das nicht-experimentelle Design erfordert den Vergleich mit einem simulierten Datensatz, in dem die aus dem Kausalmodell abgeleiteten Zusammenhänge hinterlegt sind, um empirische Evidenzen ableiten zu können (Goldfeld & Wujciak-Jens, 2020; McElreath, 2020). Das gewählte Experiment war eine Säure-Base-Titration in zwei Schulen. Aus Schule 1 nahmen drei Klassen der 10. Schulstufe eines Gymnasiums teil, in dem Chemie als Regelunterricht 2-stündig pro Woche unterrichtet wird. Es lag kaum Erfahrung mit nasschemischen Analysen vor. Aus Schule 2 nahmen zwei Klassen der 11. Schulstufe teil, die an berufsbezogenen Laborunterricht teilnahmen und bereits viel Vorerfahrung bei Titrationen besaßen ($N_{gesamt}=55$). Der simulierte Datensatz bestand aus 3 Schulen und 5 Klassen ($N_{gesamt}=39$). Das Vorwissen wurde über einen Multiple-Choice-Test vor dem Experiment erhoben (Raschmodell, $M2=119.8(104)$, $p=.13$; $CFI=.94$; $RMSEA=.054$, $95\% CI[0;0.093]$; $SRMSR=.14$, $Rel.WLE=.78$). Der CL wurde zu drei Messzeitpunkten über einen adaptierten Fragebogen (Klepsch et al., 2017) erfasst (Graded Response Model, $EAP-PV_{ICL/ECL/GCL}=.83/.92/.94$). Abbildung 1 zeigt einen Ausschnitt des Kausalmodells, dessen

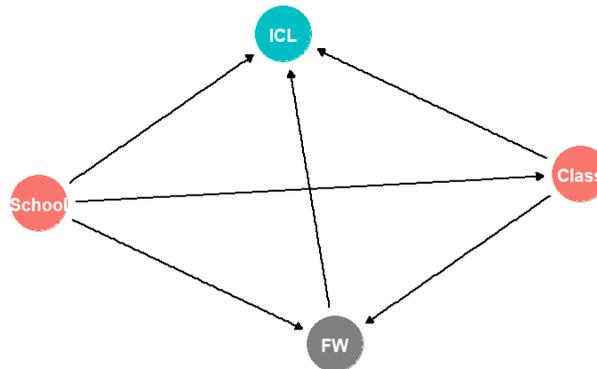


Abb. 1: Teil-DAG der Zusammenhänge von Schule, Klassenraum und Fachwissen auf ICL. Die durch Pfeile dargestellten Wirkrichtungen implizieren die Kontrolle für Fachwissen unter Berücksichtigung einer Mehrebenenstruktur.

Ergebnisse hier berichtet werden.

Ergebnisse

Abbildung 2 zeigt den Verlauf des ICL über die Messzeitpunkte (MP). Abbildung 3 zeigt den Zusammenhang von Vorwissen und ICL für beide Datensätze zu MP2.

Diskussion

Die unerfahrene Gruppe wusste nicht, was sie im Experiment erwartete, die Erfahrenen waren mit Titrationen bereits vertraut. In der Folge zeigten beide Gruppen zum Beginn keinen Unterschied im wahrgenommenen CL. Dies ändert sich durch die Konfrontation mit der tatsächlichen Arbeit bei MP2. Das Mehrebenenmodell reproduzierte den Zusammenhang von Vorwissen und ICL aus dem Simulationsdatensatz zu MP2 sehr gut. Dasselbe Modell findet

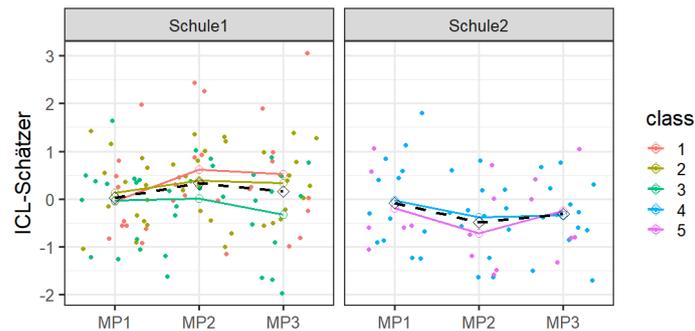


Abb. 2: Schätzwerte für ICL über den Verlauf der Messzeitpunkte. Auffällig sind die schulbezogenen Unterschiede zu MP2.

im Realdatensatz allerdings einen umgekehrten Zusammenhang. Die angenommene Wirkrichtung von Wissen auf ICL kann folglich nicht aufrechterhalten werden.

Ausblick

Aus den Ergebnissen wird die Hypothese abgeleitet, dass beim Einsatz des Sprachassistenzsystems eine fokussierte Unterstützung insbesondere in der Mitte des Experimentierprozesses sinnvoll ist. Dies ist die Interventionsbedingung für die Hauptstudien erhebung. Der nicht theoriekonforme Zusammenhang mit dem Vorwissen muss darüber hinaus durch weitere DAGs und Simulationen erkundet werden. Durch den Einsatz eines Simulationsdatensatzes für eine ökologisch valide Klassenraumstudie und die Berücksichtigung der Mehrebenenstruktur, geht das Projekt über oberflächliche Medienvergleichsstudien (Buchner & Kerres, 2023) hinaus. Es kann als anwendungsorientierte Grundlagenstudie (Klahr, 2019) eingeordnet werden, das als Framework für andere Felduntersuchungen nutzbar ist.

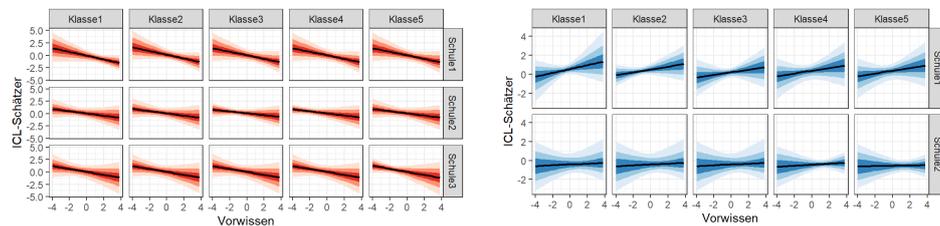


Abb. 3: Grenzeffekte des Zusammenhangs von Vorwissen und ICL zu MP2 unter Berücksichtigung der Mehrebenenstruktur (Simulationsdaten: rot; Realdaten: blau). Der Zusammenhang im Realdatensatz ist nicht theoriekonform und mit hohen Messunsicherheiten verbunden.

Referenzen

- Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education, 195*, 104711. <https://doi.org/10.1016/j.compedu.2022.104711>
- Cinelli, C., Forney, A., & Pearl, J. (2022). A Crash Course in Good and Bad Controls. *Sociological Methods & Research, 004912412210995*. <https://doi.org/10.1177/00491241221099552>
- Goldfeld, K., & Wujciak-Jens, J. (2020). simstudy: Illuminating research methods through data generation. *Journal of Open Source Software, 5*(54), 2763. <https://doi.org/10.21105/joss.02763>
- Klahr, D. (2019). Learning Sciences Research and Pasteur's Quadrant. *Journal of the Learning Sciences, 28*(2), 153–159. <https://doi.org/10.1080/10508406.2019.1570517>
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology, 8*, 1997. <https://doi.org/10.3389/fpsyg.2017.01997>
- McElreath, R. (2020). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan* (Second Edition). CRC Press.
- Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review, 22*(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Sweller, J. (2016). Working Memory, Long-term Memory, and Instructional Design. *Journal of Applied Research in Memory and Cognition, 5*(4), 360–367. <https://doi.org/10.1016/j.jarmac.2015.12.002>
- Weidlich, J., Hicks, B., & Drachsler, H. (2023). Causal reasoning with causal graphs in educational technology research. *Educational Technology Research and Development*. <https://doi.org/10.1007/s11423-023-10241-0>