

Johannes Zbinden¹
Michaela Maurer¹
Philipp Pawels^{2,3}
Bruno Rütsche⁴
Lars-Jochen Thoms^{2,3}
Lennart Schalk⁴
Johannes Huwer^{2,3}
Julia Arnold^{1dd}

¹Pädagogische Hochschule FHNW
²Pädagogische Hochschule Thurgau
³Universität Konstanz
⁴Pädagogische Hochschule Schwyz

Promptengineering zur Entwicklung von Multiple Choice -Kompetenztests

Bislang fehlen reliable und valide Messinstrumente, die zur objektiven Erfassung fachspezifischer digitaler Kompetenzen von Lehrpersonen in den Naturwissenschaften herangezogen werden können. Im Rahmen des vom Schweizerischen Nationalfonds geförderten Projekts „Digitale Kompetenzen für den Lehrberuf in den Naturwissenschaften in der Schweiz“ (DiKoLAN-CH) wurde das bereits mehrfach erprobte und empirisch überprüfte Kompetenzraster DiKoLAN-DE (Becker et. al, 2020) auf die Schweizer Bildungslandschaft adaptiert (Thoms et al., 2026). Darauf aufbauend werden mehrere Testinstrumente zur Erfassung verschiedener Facetten digitalisierungsbezogener Handlungskompetenz entwickelt. Abschließend wird in einer Querschnitterhebung der Kompetenzstand angehender Lehrpersonen in der Schweiz erfasst. Hierfür wird u.a. ein Multiple-Choice-Test (MC-Test) entwickelt.

MC-Tests bieten Vorteile bei der Testdurchführung und der vereinfachten Auswertung und können somit ökonomisch für größere Gruppen von Testpersonen eingesetzt werden; gleichzeitig ist die Entwicklung von MC-Test aufwändig und erfordert typischerweise iterative Optimierung (Bühner, 2021; Krebs, 2004; Moosbrugger & Kelava, 2020). Large Language Modells (LLMs) können bei diesem Entwicklungsprozess unterstützen. Insbesondere mit Fokus auf Verkürzung der für die Item- und Distraktorengenerierung benötigten Zeit werden zunehmend LLMs zur Erstellung von MC-Aufgaben in *High Stakes Assessments* herangezogen (Law et al., 2025). Weiteres Nutzungspotential von LLMs finden sich im Bereich der Verringerung von Bias wie *Item Writing Flaws* (May et al., 2025; Moore, 2024) und persönlichem, verzerrendem Einfluss (Moore, 2024). Die schriftliche Anfrage an die LLM (Prompt) kann in Alltagssprache erfolgen, sollte aber dennoch strukturiert sein (Radford et al., 2019). In einem systematischen Review über gängige Prompts zählen Sahoo et al. (2025) über 29 disjunkte Promptengineering-Techniken auf. Die dabei zitierte Forschung ist sehr aktuell, wobei die inkludierten Artikel entweder theoretischer Natur, analytischer Art oder, wenn empirisch, dann aus dem informatischen und mathematischen Bereich sind. Im fachdidaktischen Kontext finden sich bislang kaum Hinweise zur Ausarbeitung von Prompts oder zu sinnvollen Strategien des Prompting. Die explorative Forschungsfrage dieses Beitrags lautet somit:

FF: Welche Prompt-Bestandteile werden für die Erstellung von inhaltlich adäquaten und formell korrekten Multiple-Choice-Aufgaben zu den im DiKoLAN-CH definierten Kompetenzerwartung benötigt?

Methodisches Vorgehen

Zur Beantwortung der Fragestellungen wurden für zwei exemplarische DiKoLAN-CH-Kompetenzbereiche (“Dokumentation” und “Messwert- und Datenerfassung”) auf dem Niveau “Beschreiben”) MC-Items mittels Zero-Shot- und Few-Shot-Prompting entwickelt. Beim Zero-Shot-Prompting (Radford et al., 2019) wird mit einer mehr oder weniger vorstrukturierten schriftlichen Anfrage an das LLM eine Antwort generiert, ohne dass zusätzlich Beispiele gegeben werden. Beim Few-Shot-Prompting (Brown et al., 2020) erfolgt die Anfrage angereichert mit Beispielen. Für die Entwicklung der Promptbestandteile wurde ein iteratives Vorgehen gewählt. Die von einem Ausgangsprompt generierte Antwort wurden von zwei Fachdidaktikexperten hinsichtlich inhaltlicher Passung zur Kompetenz, Adäquatheit und formaler Korrektheit beurteilt und der Prompt laufend bis zum finalen Prompt ergänzt. Es wurden die LLMs Gemini Flash 2.5 und ChatGPT 4o zur Generierung der Antworten genutzt.

Ergebnisse Promptentwicklung

Die adäquatesten und formell korrektesten Antworten wurden erzielt, wenn die in Tab. 1 dargestellten Elemente und Unterelemente im Prompt enthalten waren.:

Tab. 1 Elemente und Unterelemente des Prompts. Die farbigen Markierungen zeigen, wie die (Unter-)Elemente konkret im finalen Prompt umgesetzt wurden.

Element	Unterelement
Verweis	-Ausformulierte Kompetenz
Formelle Hinweise	-Länge der Antwortoptionen überall gleich (Bühner, 2021; Haladyna, 2015; Krebs, 2004)
Inhaltliche Hinweise	-Keine Negierung in den Distraktoren -Keine Abschwächung der Distraktoren -Keine graduelle und absolute Modifikatoren -Situation im Stamm (Krebs, 2004) -Verweis Zielstufe -Orientierung Stamm am Lehrplan 21 -Hinweis kein Hinzufügen von Inhalten
Konkreter Auftrag	-Kompetenz Spezifikation -Vorgabe Frageformat -Vorgabe Antwortformat -Block Erläuterung
Optionales Beispiel	-Beispiel-Item

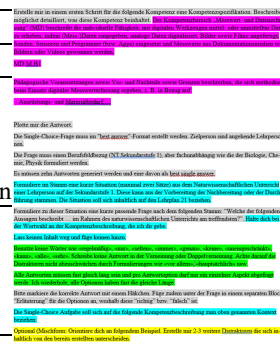


Abb. 1 Screenshot des finalen Prompts

Eine zusätzliche Erweiterung des Prompts verbesserte die Antwort nicht qualitativ. Zur Illustration des Vorgehens wird in Tabelle 2 ein mit dem finalen Prompt generiertes (zwecks besserer Darstellung gekürztes) Item und dessen qualitative Einschätzung dargestellt.

Tab. 2: Darstellung eines Beispielitems, das mittels den in Abb. 1 dargestellten Prompt-(Unter-)Elementen erstellt wurde, sowie dessen qualitative Bewertung (basierend auf der Einschätzung von zwei Fachdidaktikexpert:innen).

Item	Qualitative Bewertung
Eine Lehrperson plant im NT-Unterricht ein Experiment zur Fotosynthese der Wasserpest mit digitalen CO ₂ - bzw. O ₂ -Sensoren. Welche der folgenden Aussagen beschreibt	Die formellen Hinweise sind berücksichtigt: Es finden sich keine Item Writing Flaws. Die inhaltlichen Hinweise sind berücksichtigt.

Item	Qualitative Bewertung
eine methodische Überlegung beim geplanten Einsatz der digitalen Messwerterfassung?	Die Kompetenz MD.M.B1 wird thematisch aufgenommen.
a) Die Lehrperson beherrscht die Funktionsweise der Sonde und der Software.	Die Distraktoren sind disjunkt. Inhaltlich weisen die Antwortoptionen
b) Der Aufbau der Versuchsanordnung mit digitalen Sonden benötigt zusätzliche Zeit.	Schwächen auf: Die richtige Antwort (fett) zeichnet sich vor allem durch das treffende
c) Die Echtzeit-Visualisierung der Daten steigert das Interesse der Lernenden am Versuch.	Wording und ihre allgemeine Formulierung aus. Die Distraktoren beziehen sich jeweils nur auf einzelne methodische Teilaspekte (z. B. Organisation oder Interesse der Lernenden).
d) Die erforderliche Medienkompetenz der Lernenden beeinflusst die didaktische Planung.	Somit wird die korrekte Antwort eher durch das Verständnis des Begriffs <i>methodische Überlegung</i> erkannt als durch eine tatsächlich fachdidaktische Reflexion

Erkenntnisse, Ausblick und Limitierungen

Die explorativen Ergebnisse zur Promptgestaltung für die Erstellung von MC-Aufgaben zum Kompetenzraster DiKoLAN-CH bieten Hinweise für die Aufgabenentwicklung bei ähnlichen Vorhaben in fachdidaktischen Kontexten. Es empfiehlt sich, die aufgezählten Elemente und die formellen Hinweise zu verwenden, sowie die weiteren Unterelemente entsprechend dem Projekt anzupassen. Weiter ist es hilfreich mit Beispielen zu arbeiten (Few-Shot Prompting), wobei diese Beispiele aber sehr nahe am Zielformat sein sollten. Grundsätzlich kann angemerkt werden, dass je genauer der Befehl formuliert wird, desto qualitativ besser fällt die Antwort aus, wobei insbesondere im fachdidaktischen Bereich auch die nachfolgend dargestellten Grenzen aufgezeigt werden konnten.

Das Kompetenzraster DiKoLAN-CH spricht als Zielgruppe Fachdidaktiker:innen an. Die ausformulierten Kompetenzerwartungen richten Ansprüche an Studierende. Diese Formulierungen bedürfen einer fachdidaktischen Interpretation, um sie in Form von Lernzielen oder Lerninhalten den Studierenden zu kommunizieren. Solche fachdidaktischen Übersetzungen sind bislang noch kaum vorhanden und damit vermutlich auch nicht in den Trainingssets von derzeit existierenden LLMs enthalten. Daher konnte die fachdidaktisch gewünschte Adäquatheit auch bei funktionierendem Prompt nicht immer vollumfänglich erreicht werden. Eine vollautomatisierte Item-Generierung im Bereich der Fachdidaktik auf tertiärer Stufe scheint bei den verwendeten LLMs nicht zielführend. Ähnliches stellt auch Law et al. (2025) in seiner Untersuchung von LLM generierten MC-Aufgaben in einem anderen Fachbereich fest. Limitierend muss für die vorliegende Arbeit angemerkt werden, dass die fachdidaktische Adäquatheit durch zwei Experten im Diskurs eingeschätzt wurde.

Das hier beschriebene Vorgehen beim Promptengineering wurde im Zeitraum März – Juni 2025 mit zwei weit verbreiteten LLMs (Gemini Flash 2.5 und ChatGPT 4o) entwickelt. Es ist entsprechend auf diese LLM(-versionen) angepasst. Bei der derzeitigen rasanten Entwicklung von LLMs (z. B. Integration von Reasoning) ist davon auszugehen, dass Anpassungen des Vorgehens notwendig sind. Unser Vorgehen eignet sich aber als produktiver Ausgangspunkt für zukünftige LLM-gestützte Entwicklungen von MC-Aufgaben für Tests in den Fachdidaktiken.

Literatur

- Becker, S., Bruckermann, T., Finger, A., Huwer, J., Kremser, E., Meier, M., Thoms, L.-J., Thyssen, C. & Kotzebue, L. von. (2020). Orientierungsrahmen Digitale Kompetenzen Lehramtsstudierender der Naturwissenschaften – DiKoLAN. In S. Becker, J. Messinger-Koppelt & C. Thyssen (Hrsg.), *Digitale Basiskompetenzen: Orientierungshilfe und Praxisbeispiele für die universitäre Lehramtsausbildung in den Naturwissenschaften* (S. 14–43). Joachim Herz Stiftung.
- Brown, T. B., Mann, B... Amodei, D. (2020). Language Models are Few-Shot Learners (No. arXiv:2005.14165). arXiv. <https://doi.org/10.48550/arXiv.2005.14165>
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4., korrigierte und erweiterte Auflage). Pearson.
- Diao, S., Wang, P., Lin, Y., Pan, R., Liu, X., & Zhang, T. (2024). Active Prompting with Chain-of-Thought for Large Language Models (No. arXiv:2302.12246). arXiv. <https://doi.org/10.48550/arXiv.2302.12246>
- Haladyna, T. M. (2015). *Developing and validating multiple-choice test items* (Third edition, first issued in paperback). Routledge, Taylor & Francis Group.
- Krebs, R. (2004). *Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung*.
- Law, A. K., So, J., Lui, C. T., Choi, Y. F., Cheung, K. H., Kei-ching Hung, K., & Graham, C. A. (2025). AI versus human-generated multiple-choice questions for medical education: A cohort study in a high-stakes examination. *BMC Medical Education*, 25(1), 208. <https://doi.org/10.1186/s12909-025-06796-6>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing (No. arXiv:2107.13586). arXiv. <https://doi.org/10.48550/arXiv.2107.13586>
- May, T. A., Fan, Y. K., Stone, G. E., Koskey, K. L. K., Sondergeld, C. J., Folger, T. D., Archer, J. N., Provinzano, K., & Johnson, C. C. (2025). An Effectiveness Study of Generative Artificial Intelligence Tools Used to Develop Multiple-Choice Test Items. *Education Sciences*, 15(2), 144. <https://doi.org/10.3390/educsci15020144>
- Moore, S. (2024). *Toward Automation to Support Creation and Evaluation of Pedagogically Valid Multiple-Choice Question Assessments at Scale*.
- Moosbrugger, H., & Kelava, A. (Hrsg.). (2020). *Testtheorie und Fragebogenkonstruktion*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-61532-4>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.
- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2025). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications (No. arXiv:2402.07927). arXiv. <https://doi.org/10.48550/arXiv.2402.07927>
- Thoms, L.-J., Pawels, P., Rüttsche, B., Maurer, M., Zbinden, J., Arnold, J., Schalk, L., & Huwer, J. (2026). Expertenbefragung zur Adaption des DiKoLAN für die Schweizer Sek I. In P. Pollmeier (Hrsg.), *KI in der naturwissenschaftlichen Bildung* (Bd. 46, S. 393-396). Gesellschaft für Didaktik der Chemie und Physik.

Förderhinweis

Das Projekt “Digitale Kompetenzen für den Lehrberuf in den Naturwissenschaften in der Schweiz (DiKoLAN-CH)” wird vom Schweizerischen Nationalfonds (SNF) gefördert (Projektnummer 219778).