

HANKELN, Corinna; HAHN, Sonja, STEINKE, Jaqueline Marie & KROEHNE, Ulf

Dortmund, Frankfurt, Dortmund, Frankfurt

Lernenden-Zeichnungen automatisch auswerten mithilfe von Verstehenselementen und Computer-Vision

Graphische Antworten zur Diagnose von konzeptuellem Verständnis

Konzeptuelles Verständnis von Lernenden umfasst unter anderem die Übersetzung zum Beispiel einer verbalen oder symbolischen Repräsentation in eine graphische (Hiebert & Carpenter, 1992). Dies erfordert nicht nur die Identifikation einer passenden Darstellungsform, sondern auch deren aktive Generierung. Der „Mathe sicher können – Online-Check“ (Hankeln et al., submitted) enthält daher Aufgabenformate, in denen Lernende graphische Darstellungen erzeugen, etwa den Anteil $\frac{1}{8}$ in einem Bruchstreifen (Abbildung 1).

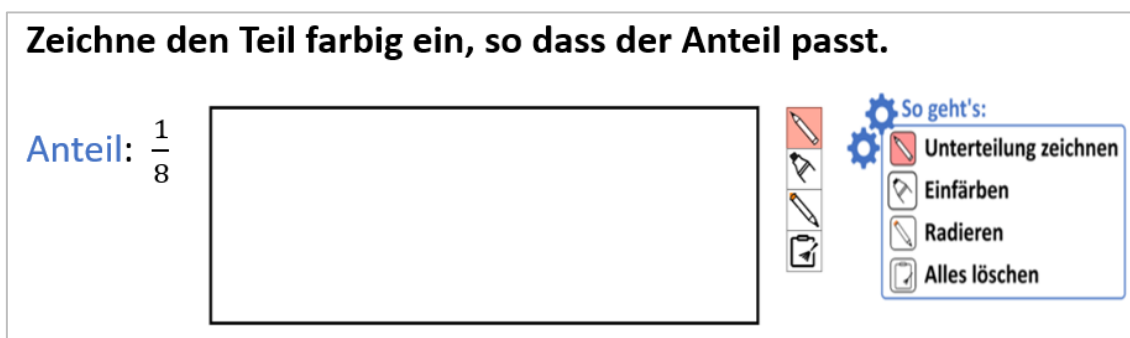


Abb. 1: Aufgabenstellung

Wird die Aufgabe korrekt gelöst, indem acht gleich große Felder eingezeichnet und eines farbig markiert wird, deutet dies auf einen vorhandenen Anteilsbegriff hin. Fehlerhafte Antworten können hingegen Hinweise auf Fehlvorstellungen liefern, etwa wenn eine ungleichmäßige Unterteilung oder fehlende Markierung vorliegt. Eine detaillierte Analyse dieser Antworten ist essenziell für passgenaue Fördermaßnahmen, erfordert jedoch fundiertes fachdidaktisches Wissen seitens der Lehrkraft. Ein unterstützendes digitales System zur Auswertung graphischer Antworten kann hier hilfreich sein. Dieser Beitrag zeigt, wie fachdidaktische Modelle die Grundlage für regelbasierte Computer-Vision-Ansätze oder KI bilden können, obwohl diese selbst kein konzeptuelles Verständnis besitzen (Hankeln, 2024).

Beispielaufgabe

Die Aufgabe in Abbildung 1 wurde im Schuljahr 2023/2024 bereits von Lehrkräften im regulären Unterricht eingesetzt: Aus den vorliegenden 62

Bildern, erstellt von Lernenden der Klassen 5 und 6, ließen sich 16 Gruppen von gleichartigen Bildern identifizieren. Nur knapp 5 % der Bilder waren völlig korrekt, weitere 20 % der Bilder waren korrekt mit kleineren Einschränkungen, wie etwa zu ungleichmäßig groß gezeichneten Feldern. Die übrigen Bilder sortierten sich zu verschiedenen Fehlertypen. Jeder dieser Bildtypen kann dabei zu einer bestimmten Diagnose des Anteilsbegriffs und einer sich daraus ergebenden Förderempfehlung zugeordnet werden. Dabei kann allerdings nicht von einer Lehrkraft erwartet werden, jeden Bildtypus selbst spontan so deuten zu können. Stattdessen soll das digitale System dabei helfen, die Bildtypen automatisch zu klassifizieren. Es wäre theoretisch möglich, eine KI auf die entsprechende Klassifizierung zu trainieren, was jedoch einen hohen Aufwand und vorab große Menge an Trainingsdaten erfordern würde (Von Davier et al., 2023).

Aus stoffdidaktischer Perspektive gibt es drei Verstehenselemente (VE) (Drollinger-Vetter, 2011) die zentral für den Aufbau eines tragfähigen Anteilsverständnisses sind (Wessel et al., 2020): (VE1) die Gleichmäßigkeit der Unterteilung in Form von gleich groß eingezeichneten Feldern, (VE2) die Darstellung des Ganzen als Anzahl der insgesamt eingezeichneten Felder und (VE3) die Darstellung des Teils als Anzahl der markierten Felder. In der zu der Aufgabe gehörenden Fördereinheit werden gezielt diese VE aufgegriffen und mit ihrer Hilfe das Verständnis eines Anteils als Beziehung zwischen Teil und Ganzem erarbeitet. Um somit aus den Lernenden-Antworten abzuleiten, welche Fördereinheit für das jeweilige Kind besonders relevant erscheint, sind somit weniger die konkreten Bildtypen relevant als vielmehr das Fehlen oder Vorhandensein der jeweiligen VE. Dies bedeutet, dass jede der Antworten auf das Vorhandensein der drei VE hin geprüft werden kann: A: Sind die eingezeichneten Felder gleich groß? B: In wie viele Felder ist der Streifen eingeteilt worden? und C: Wie viele der Felder wurden farbig markiert? Darüber hinaus fallen einige spezifische Fehler auf, bei denen systematische Fehlkonzepte anstelle der VE getreten zu sein scheinen.

Auf diese Art lässt sich ein Baum erstellen, der die unterschiedlichen Fälle, deren Unterscheidung diagnostischen Mehrwert bietet, in eine systematische Abfolge bringt (Hahn et al., eingereicht). An den Endpunkten der Äste stehen dabei 23 Endkategorien, also Bildtypen, die zu unterschiedlichen Förderhinweisen führen. Von diesen 23 Kategorien sind in der ersten Erprobung also bereits 16 Fälle aufgetreten.

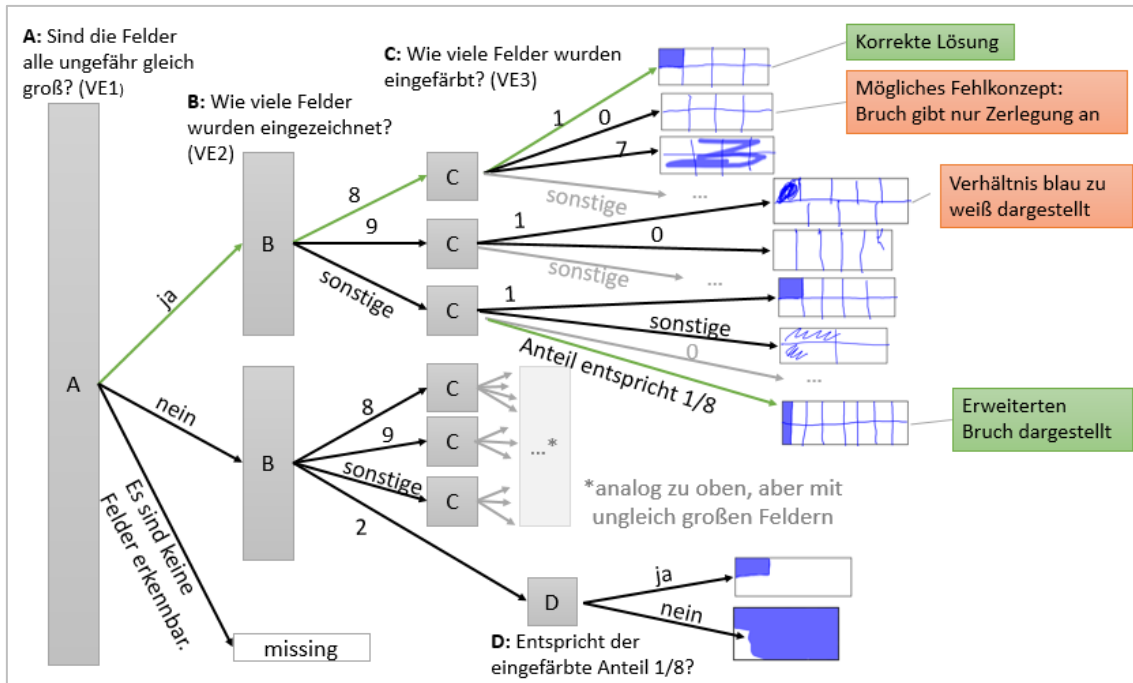


Abb. 2: Entscheidungsbaum entlang der Verstehenselemente mit ausgewählten Beispielantworten und -diagnosen

Der so entstandene Entscheidungsbaum kann nun genutzt werden, um mit Hilfe einer KI eine Lernenden-Antwort einer der Bildtypen und damit einem zugehörigen Förderhinweis zuzuordnen und gleichzeitig die Ableitung der Entscheidung aus fachdidaktischer Perspektive transparent zu machen, sowohl mit Blick auf die vorhandenen Ressourcen, an die angeknüpft werden kann als auch auf das Fehlen bestimmter Vorstellungen oder typischer Fehlvorstellungen.

Die Aufgabe aus Abb. 1 wurde zunächst durch menschliche Rater kodiert und eine Übereinstimmung von 86,6 % gefunden. Dabei sind alle Abweichungen auf eine unterschiedliche Einschätzung bei Frage (A) nach der gleichen Größe zurückzuführen, bei der die Rater unterschiedlich streng agierten. Die Übereinstimmung bei den Fragen B und C lag jeweils bei 100 %.

Die Kodierung mithilfe eines regelbasierten Computer-Vision-Systems traf die menschliche Kodierung in den ersten Versuchen zu 70,5%. Bezogen auf die einzelnen Fragen lag die Abweichung bei A mit 87,3% vergleichbar mit den menschlichen Ratern, das Auszählen der Felder gelang jeweils nur zu 82,1 % (B) bzw. 85,2 %.

Fazit

Die differenzierte und auf das Weiterlernen ausgerichtete Auswertung offener Antwortformate ist ein Bedarf, bei dem künstliche Intelligenz theoretisch zu einer Arbeitsentlastung bei gleichzeitiger Erhöhung der Informativität

führen kann. Da viele Modelle jedoch kein eigenes tiefes konzeptuelles Verständnis der mathematischen Inhalte mitbringen, sondern rein über Ähnlichkeiten zu bekannten Daten arbeiten, kann von diesen Modellen nicht per se erwartet werden, Lernenden-Antworten ohne Training so zuverlässig zu klassifizieren, wie es eine tiefgehende Diagnostik und die Ableitung von Handlungsempfehlungen erfordern würde. In unseren Erprobungen haben wir dabei festgestellt, dass der Schlüssel für eine Optimierung des Nutzens in der präzisen Auffaltung in die Bestandteile der zu messenden Konstrukte liegt. Durch die Abkehr der bloßen Identifikation von korrekten Antworten und typischen Fehlern hin zu einer auch ressourcenorientierten Auswertung entlang einer Pfadstruktur mit präziser Indikatorbeschreibung, scheint es möglich, die automatische Kodierung der graphischen Antworten nicht nur genauso zuverlässig werden zu lassen wie die menschlichen Kodierungen, sie schafft gleichzeitig auch Transparenz im Hinblick auf die Validität der Ableitung von Fördermaßnahmen und schafft es auch, die bereits vorhandenen Ressourcen der Lernenden, an die in der Förderung angeknüpft werden kann, zu erfassen. Nichtsdestotrotz bleibt festzuhalten, dass es sich bei den hier vorgestellten Erprobungen um erste Machbarkeitsstudien handelt, die es für weitere Antworten sowie Aufgabenformate zu erproben und zu adaptieren gilt.

Literatur

- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit*. Waxmann.
- Hahn, S., Hankeln, C., Gross, S., Hammerla, L. & Kroehne, U. (eingereicht) A Roadmap to Automatically Analyzing Graphical Responses in Small Scale Formative Assessments bei der *21st Conference of the European Association for Research on Learning and Instruction*, Graz, Austria
- Hankeln, C. (2024). Challenges in Using ChatGPT for Assessing Conceptual Understanding in Mathematics Education. *Journal of Mathematics Education*, 17(1), 1–15. <https://doi.org/10.26711/007577152790171>
- Hankeln, C., Kroehne, U., Voss, L., Gross, S., & Prediger, S. (submitted). Developing digital formative assessment for deep conceptual learning goals: Which topic-specific research gaps need to be closed?
- Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 65–97). Macmillan.
- Von Davier, M., Tyack, L., & Khorramdel, L. (2023). Scoring Graphical Responses in TIMSS 2019 Using Artificial Neural Networks. *Educational and Psychological Measurement*, 83(3), 556–585. <https://doi.org/10.1177/00131644221098021>
- Wessel, L., Kuzu, T., & Prediger, S. (2020). Sprachbildender Vorstellungsaufbau für Brüche in Klasse 6. In S. Prediger (Hrsg.), *Sprachbildender Mathematikunterricht in der Sekundarstufe. Ein forschungsbasiertes Praxisbuch* (S. 148–151). Cornelsen.