

BAUER, Thomas & BÖLLING, Niklas
Marburg

Multiple-Choice-Fragen zu konzeptuellem Wissen: Mit welchen Überlegungen gelangen Studierende zu Antworten?

Einleitung

Multiple-Choice-Fragen (MC-Fragen) werden in der Hochschullehre zu verschiedenen Zwecken eingesetzt: zum einen als Element in bestimmten Lernarrangements, zum anderen als Testinstrument, z.B. in Klausuren. In unserem Projekt sind wir an MC-Fragen interessiert, die konzeptuelles Wissen adressieren. Solche Aufgaben spielen eine wesentliche Rolle als Lernaufgaben im Rahmen von Peer Instruction (Bauer et al., 2023) oder als Testaufgaben im RACI-Test (Real Analysis Concept Inventory; Bauer et al., 2024).

Wenn Studierende MC-Fragen in einem Test zu konzeptuellem Wissen richtig beantworten, dann möchte man idealerweise darauf rückschließen, dass sie über bestimmtes Wissen zu den angesprochenen Konzepten verfügen. Um einzuschätzen, in welchem Maße solche Rückschlüsse gerechtfertigt sind, müsste man Erkenntnisse darüber haben, mit welchen Überlegungen die Studierenden in dem Test zu ihren Antworten gelangt sind: Ist die Antwortentscheidung durch sachgerechte Argumente auf Basis von konzeptuellem Wissen gefallen oder waren andere Aspekte beteiligt oder gar ausschlaggebend? Es scheint daher dringend nötig, mehr Einblick in die Überlegungen von Studierenden bei der Bearbeitung von MC-Fragen zu erhalten, denn die Interpretation von Testergebnissen beruht wesentlich darauf, von welchen Annahmen über deren Überlegungen man ausgeht.

Hintergrund und Fragestellung

MC-Fragen bieten hinsichtlich der Durchführung und Auswertung von Tests eine Reihe von praktischen Vorteilen. Verschiedentlich wird allerdings kritisch hinterfragt, ob sie als Instrument geeignet sind, wenn man an kognitiven Prozessen interessiert ist, die über Erinnern und Erkennen hinausgehen. Untersuchungen weisen jedoch darauf hin, dass es durchaus möglich ist, MC-Aufgaben zu konstruieren, die „higher-order thinking“ messen (siehe Scully, 2017, sowie auch Lindner et al., 2015).

Diese Überlegungen sind von besonderer Bedeutung, wenn ein Test entwickelt und eingesetzt wird, in dem konzeptuelles Wissen (siehe Hiebert & Lefevre, 1986) auf höheren kognitiven Stufen (Anderson et al., 2001) erfasst werden soll. Im RACI-Projekt (Bauer et al., 2024) wird dieses Ziel für den Bereich der reellen Analysis verfolgt. Wir bearbeiten das Forschungsanliegen, das im Titel dieses Beitrags formuliert wurde, hier für den RACI-Test.

In: L. Schick, M. Platz & A. Lambert (Hrsg.),
Beiträge zum Mathematikunterricht 2025.

58. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.
<https://doi.org/10.37626/GA9783959873307.0>

Dabei verfolgen wir derzeit zwei Aspekte; zum einen können wir für jede einzelne Aufgabe (bzw. jedes einzelne Item) des Tests fragen:

(F1) *Aufgabenbezogener Fokus*: Welches konzeptuelle Wissen setzen Studierende bei der Bearbeitung der Aufgabe / des Items ein? Welche Fehlüberlegungen kommen bei der Bearbeitung vor?

Antworten auf diese Frage sind auf den mathematischen Inhalt der einzelnen Aufgabe bezogen und insofern an sie gebunden. Unser Ziel bei (F1) ist es, zu ermitteln, welche Variation zwischen den Studierenden bei einer gegebenen Aufgabe bzw. einem gegebenen Item vorkommt.

Zum anderen fragen wir aufgabenübergreifend:

(F2) *Testbezogener Fokus*: Welche Arten von adäquaten/inadäquaten Überlegungen stellen Studierende im RACI-Test an, wenn sie zu korrekten Antworten gelangen? Welche, wenn sie zu inkorrekten Antworten gelangen?

Um „Arten von Überlegungen“ zu konzeptualisieren, unterscheiden wir in Anlehnung an Lithner (2000) *logical reasoning* (LR), *plausible reasoning* (PR) und *reasoning based on established experiences* (EE). Zur Analyse verwenden wir aufgabenunabhängige Kategorien und können auf diese Weise nicht nur die Variation zwischen Studierenden, sondern auch zwischen Aufgaben/Items betrachten. So lässt sich zum Beispiel der Frage nachgehen, bei welchen Aufgaben im RACI-Test das (aus Perspektive der Testkonstruktion unerwünschte) Phänomen vorkommt, dass Studierende mit inadäquaten Plausibilitätsüberlegungen zu richtigen Antworten gelangen.

Methodisches Vorgehen

Material. Wir gehen im vorliegenden Beitrag auf Ergebnisse zu folgendem Item aus einer Aufgabe zur Folgenkonvergenz im RACI-Test ein.

Aufgabe. Es sei $(a_n)_{n \in \mathbb{N}}$ eine Folge reeller Zahlen, die gegen eine Zahl $a \in \mathbb{R}$ konvergiert. Lassen sich die folgenden Aussagen daraus folgern?

[...]

(3) Für jedes $n \in \mathbb{N}$ liegen im Intervall $[a - \frac{1}{n}, a + \frac{1}{n}]$ alle bis auf höchstens endlich viele Folgenglieder.

ja nein

Im Rahmen der Entwicklung des RACI-Tests (Bauer et al., 2024) wurden A-priori-Analysen möglicher Argumentationen durchgeführt. Beim vorliegenden Item könnte eine Person, die ihre Entscheidung stichhaltig begründet, beispielsweise folgende Überlegungen anstellen:

1) Sie drückt die vorausgesetzte Konvergenz der Folge durch die Bedingung $\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n \geq N: |a_n - a| < \varepsilon$ aus.

2) Sie spezialisiert – auf das Ziel der Aufgabe hin gerichtet – die allquantifizierte Variable ε zu $1/n$.

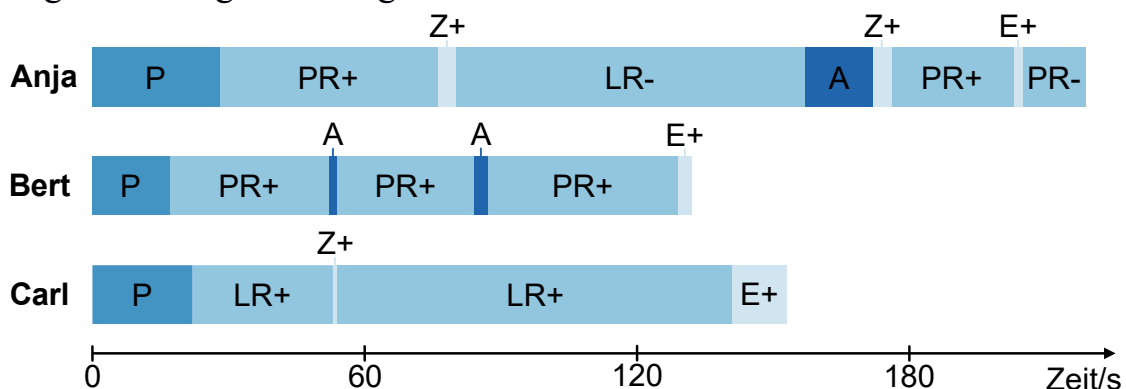
3) Sie versteht (oder erinnert), dass die Quantifizierung $\exists N \in \mathbb{N} \forall n \geq N$ inhaltlich äquivalent als „alle bis auf höchstens endlich viele“ ausgedrückt werden kann.

Insgesamt wären solche Überlegungen geeignet, stichhaltig zu begründen, dass die zu betrachtende Aussage eine Spezialisierung (und insofern Folgerung) der Voraussetzung ist. Selbstverständlich sind auch andere stichhaltige Überlegungen denkbar, etwa indem die Konvergenzbedingung (in Formeln oder verbal) unter Verwendung des Umgebungsbegriffs ausgedrückt wird.

Datenerhebung und -auswertung. Um Einblick in die Denkprozesse der Studierenden zu erhalten, wurden bisher mit zwölf Personen Interviews durchgeführt, in denen diese den Test mit lautem Denken bearbeiten. Bei der Durchführung der Interviews wurde Geschriebenes und Gesprochenes aufgezeichnet. In den transkribierten Bearbeitungsprozessen codieren wir Bearbeitungsphasen in Anlehnung an Lithners „four-step structure“ (Lithner, 2000) und unterscheiden dabei die Kenntnisnahme einer Problemsituation (P), die Auswahl einer Strategie (A), die Durchführung der Strategie sowie Einschätzungen und Entscheidungen. In den Durchführungen codieren wir die Art der Überlegung (LR, PR, EE) und stufen diese jeweils als adäquat (+) oder inadäquat (–) ein. Einschätzungen und Entscheidungen codieren wir als korrekt (+) oder nicht korrekt (–). Neben endgültigen Entscheidungen (E) berücksichtigen wir auch Zwischeneinschätzungen (Z).

Erste Ergebnisse und Diskussion

Beispielhaft gehen wir auf drei Bearbeitungen des oben vorgestellten Items ein (Anja, Bert, Carl). Die Codierung nach Bearbeitungsphasen ist im nachfolgenden Diagramm dargestellt.



Wir gehen exemplarisch auf Anjas Bearbeitung ein. Sie betrachtet die zu untersuchende Aussage zunächst in Einzelfällen (PR+) und nimmt daraufhin eine korrekte Zwischeneinschätzung vor (Z+). Anschließend versucht sie,

die Konvergenzbedingung mit logischen Symbolen auszuformulieren, stellt dabei aber den Zusammenhang zwischen n und ε sowohl in der Formel als auch ihrer verbalen Formulierung falsch dar (LR-). Sie entscheidet sich dann, vorübergehend zu einem anderen Item zu wechseln (A). Anschließend gibt sie eine Zwischeneinschätzung (Z+), führt dann eine adäquate, vorstellungsbezogene Überlegung durch (PR+) und trifft daraufhin eine korrekte Entscheidung (E+). Sie schließt mit der (allerdings irrigen) Annahme, dass sich die behauptete Aussage aus der vorher von ihr fehlerhaft formulierten Konvergenzbedingung durch Umformen gewinnen ließe (PR-).

Die Bearbeitungen der anderen beiden Probanden sind deutlich kürzer und weisen eine einfachere Phasenstruktur auf: Bert führt adäquate Überlegungen anhand einer konkreten Beispielfolge durch (PR+) und kommt auf dieser Grundlage zu einer (korrekten) Entscheidung (E+). Während er keinerlei Überlegungen vom Typ LR anstellt, sind die Überlegungen von Carl vollständig von diesem Typ: Nach einer kurzen Elaboration der zu folgernden Aussage (LR+) nimmt er eine korrekte Zwischeneinschätzung vor (Z+) und bildet dann eine vollständige Argumentation (LR+).

In Bezug auf (F1) und (F2) zeigen bereits die drei gezeigten Bearbeitungen eine große Spannweite auf: Korrekte Entscheidungen im Test können auf der verständigen, zielführenden Nutzung der Definition der Folgenkonvergenz basieren (Carl), aber auch auf konzeptuellem Verständnis auf intuitiv-anschaulicher Ebene (Bert) oder auf einer (im vorliegenden Fall widersprüchlichen) Kombination von beidem (Anja).

Literatur

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.
- Bauer, T., Biehler, R., & Lankeit, E. (2023). ConcepTests in Undergraduate Real Analysis: Comparing Peer Discussion and Instructional Explanation Settings. *Int. J. Res. Undergrad. Math. Ed.*, 9, 426–460.
- Bauer, T., Biehler, R., & Lankeit, E. (2024). Designing a Concept Inventory for Real Analysis. INDRUM2024.
- Hiebert, J., & Lefevre, P. (1986). Conceptual and Procedural Knowledge in Mathematics: An Introductory Analysis. In J. Hiebert (Hrsg.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (S. 1–27). Lawrence Erlbaum Associates.
- Lindner, M. A., Strobel, B., & Köller, O. (2015). Multiple-Choice-Prüfungen an Hochschulen? *Zeitschrift für Pädagogische Psychologie*, 29(3–4), 133–149.
- Lithner, J. (2000). Mathematical reasoning for task solving. *Educ. Stud. Math.*, 41, 165–190.
- Scully, D. (2017). Constructing Multiple-Choice Items to Measure Higher-Order Thinking. *Practical Assessment, Research, and Evaluation*, 22(4).