

Analyse chemiedidaktisch relevanter diagnostischer Kompetenzen

Einleitende Überlegungen

Diagnostische Kompetenzen zählen zu den bedeutsamsten professionellen Fähigkeiten von Lehrer*innen (KMK 2005, 3), da sie sowohl weitreichende Konsequenzen für das Lernen als auch für die schulische Laufbahn und die spätere Berufsbiografie von Schüler*innen nach sich ziehen. Aufgrund dieser hohen Relevanz für den Berufsalltag von (angehenden) Lehrer*innen wurde das Instrument des Simulierten Klassenraums (SKR) herangezogen, um diagnostische Kompetenzen von Lehramtsstudent*innen sowie die von Lehramtsanwärter*innen mit Unterrichtsfach Chemie vergleichend zu untersuchen.

Theoretischer Hintergrund

Die diagnostische Kompetenz wird neben der Klassenführungscompetenz, der didaktischen und fachwissenschaftlichen Kompetenz als eine der vier Schlüsselkompetenzen von Lehrkräften angesehen (Helmke, Hosenfeld & Schrader 2004, 119; KMK 2004, o.S.). Ohne adäquate Kenntnisse des Leistungsvermögens und Lernvoraussetzungen ihrer Schüler*innen können Lehrkräfte Unterricht nicht adaptiv gestalten. Akkurate, diagnostisch zutreffende, Urteile sollten die Basis z. B. für die Auswahl von Unterrichtsaktivitäten, von Aufgaben mit unterschiedlichem Schwierigkeitsgrad oder von Teststrategien bilden (Südkamp & Praetorius 2017, 13). Allein darauf ist die hohe praktische Relevanz des Themas zu begründen (Helmke et al. 2004, 128; Retelsdorf & Möller 2016). Für den in der Praxis typischen Fall, dass eine Lehrperson mehrere Schüler*innen einer Klasse anhand eines Merkmals (z.B. Leistung, Motivation, Intelligenz, Prüfungsängstlichkeit) zu beurteilen hat, unterscheiden Schrader und Helmke (1987) drei Komponenten diagnostischer Genauigkeit (auch *Urteilsakkuratheit* genannt):

- Die *Niveauebene* gibt an, inwiefern die Lehrkraft das absolute Niveau eines Schülermerkmals korrekt einschätzt. Sie charakterisiert die Tendenz von Lehrern, das Niveau der eigenen Klasse in Bezug auf ein Merkmal im Vergleich zu den tatsächlichen Ergebnissen insgesamt eher zu über- oder unterschätzen (Schrader & Helmke 1987, 30). Dabei wird die Differenz zwischen der mittleren Leistungseinschätzung durch den Probanden und der mittleren tatsächlichen Leistung der simulierten Schüler*innen gebildet, weshalb eine Niveauebene nahe Null optimal ist (größer Null: Überschätzung der SuS-Leistungen, kleiner Null: Unterschätzung).
- Die *Differenzierungskomponente* kennzeichnet die Tendenz von Lehrern, die Streuung der eingeschätzten Merkmalsausprägungen im Vergleich zur Streuung der tatsächlichen Schülermerkmale zu über- oder unterschätzen. Anhand dieser Komponente kann z. B. eine „Tendenz zur Mitte“ als systematische Urteilstendenz ermittelt werden (Südkamp, Möller & Pohlmann 2008a, 262). Die Differenzierungskomponente wird als Quotient aus der Varianz der Leistungsbeurteilungen der Probanden und der Varianz der tatsächlichen Testergebnisse gebildet, weshalb sich eine optimal eingeschätzte Streuung in einer Differenzierungskomponente im Wert 1 widerspiegelt (größer als Eins: Tendenz zum Extremen, kleiner als Eins: Tendenz zur Mitte).
- Die *Rangkomponente* (auch: *Rangordnungskomponente*) gibt Auskunft, ob Lehrkräfte die Rangfolge einer relativen Merkmalsausprägung der Schüler*innen innerhalb einer Klasse korrekt einschätzen (z.B. Südkamp, Kaiser & Möller 2012). Sie wird aus der Korrelation zwischen dem von den Probanden vorhergesagten und den in der Klasse tatsächlich erzielten Leistungen gebildet (Schrader & Helmke 1987, 31).

Die Rangkomponente ist im Gegensatz zur Niveau- und Differenzierungskomponente als „originäres Kompetenzmerkmal der Lehrkraft zuzuschreiben“ (Südkamp & Praetorius 2017,

23). Deshalb wird in vielen Studien oftmals die These aufgegriffen, dass die Rangkomponente der eigentliche Indikator diagnostischer Kompetenz sei. Deshalb fokussierten Forschungsarbeiten in diesem Gebiet lange Zeit auf die Bestimmung der Rangkomponente. Zwei Gründe sprechen mittlerweile gegen die Überbetonung der Rangkomponente: Zum einen bleibt die Frage nach der Struktur, Generalität und Spezifität diagnostischer Kompetenzen offen, wenn nur ein Aspekt der Urteilsakkuratheit betrachtet wird (Südkamp & Praetorius 2017, 24). Zum anderen werden bei alleiniger Betrachtung der Rangkomponente über mehrere Aufgaben und Lernende hinweg Mittelwerts- und Streuungsunterschiede außer Acht gelassen, was zu einer Überschätzung diagnostischer Kompetenzen führen kann (Südkamp & Praetorius 2017, 25). Südkamp, Möller und Pohlmann (2008a, 2008b) plädieren u. a. deshalb dafür, zusätzlich zu den genannten drei Komponenten, auch das *globale Abweichungsmaß* zu betrachten.

- Das *globale Abweichungsmaß* gibt Auskunft darüber, wie stark die Einschätzungen der Lehrkraft von den tatsächlichen Merkmalsausprägungen der einzelnen Schüler*innen abweichen. Das globale Abweichungsmaß trägt dazu bei, dass Unterschiede im Ausmaß des Verschätzens der Proband*innen unabhängig von deren Urteilstendenzen sichtbar werden und Über- und Unterschätzungen sich nicht neutralisieren (Kaiser & Möller 2017, 58).

Fragestellungen

Aufgrund der dargelegten Überlegungen formulieren wir die folgenden Forschungsfragen:

1. *Inwiefern sind Lehramtsstudent*innen sowie Lehramtsanwärter*innen in der Lage, mündlich erbrachte Schüler*-Leistungen – trotz der im Unterricht zu beobachtenden Interaktionsdynamik – korrekt einzuschätzen und zu beurteilen? (Grundsätzlicher Forschungsfokus)*
2. *Inwieweit lässt sich die erreichte diagnostische Performanz mit den Ergebnissen früherer Untersuchungen vergleichen? (Vergleichbarkeitsvermutung)*
3. *Inwiefern zeigen sich **Modifikationseffekte** auf der Grundlage unterschiedlich konzipierter simulierter Klassenräume (z. B. im Naturwissenschaftlichen Unterricht 5/6 oder im Chemieunterricht der Jahrgangsstufe 8/9)?*
4. *Inwiefern lassen sich **Trainingseffekte** identifizieren, wenn Proband*innen die Simulation eines Klassenraums zweimal durchlaufen?*
5. *In welcher Weise bringen die Analysen eine Art **Erfahrungsvorteil** zum Vorschein; d. h. zeigen Lehramtsanwärter*innen bessere diagnostische Performanz als Lehramtsstudierende?*

Methode

Der simulierte Klassenraum ist eine speziell für die Untersuchung der diagnostischen Kompetenz gut geeignete Methode (Fiedler, Freytag & Unkelbach, 2007; Fiedler, Walther, Freytag & Plessner, 2002). Bei der verwendeten Version handelt es sich um eine HTML-5-gestützte Computersimulation eines Klassenraums, bei der die Proband*innen die Rolle einer Lehrkraft übernehmen. Dabei interagieren sie virtuell mit ihren Schüler*innen. Im Anschluss an ein simuliertes Klassengespräch sind die Proband*innen gehalten, die vorab experimentell festgelegte Qualität der simulierten Schüler*innen-Leistung als auch deren Motivation beurteilen. Zusätzlich müssen die Proband*innen den Schüler*innen eine Note auf einer Skala von eins bis sechs (typische Schulnoten) zuordnen. Um den Fokus auf die Einschätzung der SuS-Leistungen zu legen, wurde der die Häufigkeit der Meldungen simulierende Motivationsparameter für alle Schüler*innen auf 50% fixiert. Die Leistungsparameter wurden für je eine/n der 6 Jungen und 6 Mädchen zwischen 90% und 40% Klasse in abgestuft 10%-Schritten justiert, um ein breites Leistungsspektrum und eine eindeutige Rangfolge zu simulieren.

In dieser Untersuchung wurden zwei unterschiedliche Klassenräume verwendet. Der bereits erprobte Klassenraum Nawi 5/6 (Bolte et al. 2011; 2012) wurde um den Klassenraum Chemie im Jahrgang 8/9 ergänzt. Die Fragen und die Antworten des simulierten Nawi-Klassenraums entstammen aus zur Schülervorstellungsforschung (Benedict & Bolte 2008; Erb & Bolte 2009); die Inhaltfelder und Themen des simulierten Chemieunterrichts wurden auf Grundlage

des aktuellen Rahmenlehrplans (SenBJF 2015) entwickelt; Das Niveau entspricht dem Unterricht an Gymnasien. Diese konzeptionelle Entscheidung soll sicherstellen, dass die Proband*innen keine Unterscheidung zwischen leichten und schweren Fragen vornehmen und Antworten unterschiedlich gewichten. Aufgrund der Beschränkung auf Multiple-Choice-Fragen, bei denen immer eine der vier vorgegebenen Antwortmöglichkeiten als fachlich richtig einzustufen sind, liegen in beiden simulierten Klassenräumen mit klar eingegrenzten Merkmalen vor, die eine hohe Urteilsakkuratheit ermöglichen (nach Lintorf et al. 2016).

Ergebnisse

Die Stichprobe umfasst 37 Lehramtsstudierende, die im Wintersemester 2017/18 das Modul „Einführung in die Fachdidaktik der Chemie“ belegten, sowie 19 Lehramtsanwärterinnen, die sich im Sommer 2018 im Vorbereitungsdienst des Landes Berlin befanden und das Fach Chemie unterrichteten. Das Alter der Lehramtsstudierenden betrug im Mittel 23,8 Jahre ($SD = 4,1$ Jahre), das der Lehramtsanwärter im 33,6 Jahre ($SD = 5,3$ Jahre).

Die Analysen und Prüfung auf statistische Signifikanz (ANOVA) sprechen für die Vergleichbarkeitsvermutung und bestätigen einen positiven Trainingseffekt in aufeinanderfolgenden Durchläufen. Ein statistisch signifikanter Modifikationseffekt wurde nicht festgestellt, ebenso wenig konnte die Erfahrungsvorteilsvermutung der Lehramtsanwärter*innen gegenüber den Lehramtsstudierenden statistisch untermauert werden.

	Optim.	Stollin & Bolte (2020)		Bolte u.a. (2011)	Südkamp u.a. (2008)	Südkamp & Möller (2009)	Spinath (2005)
		Nawi 5/6	Chem.8/9				
Rangkomponente	1	0,37 0,33	0,54 0,45	0,40; 0,26	0,62; 0,68 0,54; 0,53	0,64; 0,66	Median 0,40
Niveauelemente	0	0,04 0,02	0,01 0,04	-0,07; -0,09	0,04; 0,03 0,09; 0,07	0,07; 0,09	0,03
Differenzierungskomponente	1	0,49 0,57	0,61 0,57	0,79; 0,71	0,76; 0,92 0,74; 0,81	0,81; 0,82	0,84
globales Abweichungsmaß	0	0,26 0,23	0,21 0,22	0,17; 0,19	0,20; 0,15 0,21; 0,20	0,20; 0,19	./.

Tab. 1: Ergebnisse zur Analyse diagnostischer Kompetenzen – differenziert nach Beurteilungskomponente und Referenzstudien¹

Fazit

Die Lehramtsanwärter*innen und -student*innen der Stichprobe sind in der Lage, das Niveau simulierter Schüler*innen passend einzuschätzen. Defizite zeigen sich allerdings in der Einschätzung der Streuung der Leistungen sowie der Rangreihenfolge (FF 1). Die Ergebnisse des neu aufgelegten Klassenraums lassen sich mit den Ergebnissen früherer Untersuchungen vergleichen, es gibt keine statistisch signifikant abweichenden Unterschiede in der Performanz der Proband*innen bzgl. der vier Komponenten (FF 2). Es konnte kein Modifikationseffekt festgestellt werden, daher ist davon auszugehen, dass das Design des Klassenraums in dieser Studie keine Auswirkungen auf die diagnostische Performanz der Proband*innen ausgeübt hat (FF 3). Im 2. Durchlauf des simulierten Klassenraums konnte in 2 Bereichen eine statistisch signifikante Verbesserung festgestellt werden, nämlich bzgl. der Rangkomponente und des globalen Abweichungsmaß; diesbezüglich ist von positiven Trainingseffekten auszugehen (FF 4). Lehramtsanwärter*innen haben keine signifikant besseren diagnostischen Urteile fallen können als weniger weit ausgebildete bzw. praxiserfahrene Lehramtsstudent*innen. Somit konnte kein (statistisch signifikanter) Erfahrungsvorteil nachgewiesen werden (FF 5).

¹ 1. Zeile: Lehramtsstudent*innen / 2. Zeile (kursiv): Lehramtsanwärter*innen – in den Referenzstudien wurden ausschließlich Lehramtsstudierende untersucht; 1. und 2. Zeile bei Südkamp u. a. (2008) repräsentiert je unterschiedliche Fach-Klassenräume, die von den Proband*innen jeweils zweimal durchlaufen wurden.

(Ausgewählte) Literatur

- Benedict, C. & Bolte, C. (2008). Erste Schritte der Analyse konzeptueller naturwissenschaftlicher Kompetenzen von Kindern im Grundschulalter – aufgezeigt am Beispiel des Teilchenkonzepts. In: Giest, H. und Wiesemann, J. (Hrsg.): *Kind und Wissenschaft – Welches Wissenschaftsverständnis hat der Sachunterricht? Probleme und Perspektiven des Sachunterrichts*, 18, Bad Heilbrunn: Klinkhardt-Verlag, 263-275.
- Bolte, C., Koeppen, G., Möller, J., & Südkamp, A. (2011). Kompetenzdiagnostik im virtuellen naturwissenschaftlichen Unterricht. In: D. Höttecke (Hg.): *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Zur Didaktik der Physik und Chemie. Probleme und Perspektiven*. Münster: Lit-Verlag. S. 146-148.
- Bolte, C., Koeppen, G., Möller, J., & Südkamp, A. (2012). Chemistry related pedagogical diagnostic competencies of pre-service chemistry teachers analyze by means of the Simulated Science Classroom. *Proceedings of the European Science Educational Research Association (ESERA)*, Lyon, France, September 2011.
http://lsg.ucy.ac.cy/esera/e_book/base/ebook/strand13/ebook-esera2011_BOLTE-13.pdf(2012-05-31)
- Erb, M. & Bolte, C. (2009). *IQ² – The Inquiry Qualification Questionnaire*. Berlin: Freie Universität Berlin – Didaktik der Chemie.
- Fiedler, A. B., Freytag, P. & Unkelbach, C. (2007). Pseudocontingencies in a simulated classroom. *Journal of Personality and Social Psychology*, 92, 655-677.
- Fiedler, K., Walther, E., Freytag, P. & Plessner, H. (2002). Judgment biases in a simulated classroom – a cognitive-environmental approach. *Organizational Behavior and Human Decision Processes*, 88, 527f.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In: Arnold, R. & Griese, C. (Hrsg.), *Schulleitung und Schulentwicklung*. Hohengehren: Schneider, 119-144.
- Kaiser, J. & Möller, J. (2017). Diagnostische Kompetenz von Lehramtsstudierenden. In Gräsel, C. & Trempler, K. (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals - Interdisziplinäre Betrachtungen, Befunde und Perspektiven*. Wiesbaden: Springer, 55-74.
- KMK (2004) = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der KMK vom 16.12.2004 – Bonn.
- KMK (2005). *Standards für die Lehrerbildung*. München: Luchterhand.
- Lintorf, K., van Ophuysen, S. & Behrmann, L. (2016). Diagnostik im Lehrerberuf. In: M. Rothland (Hrsg.), *Beruf Lehrer / Lehrerin. Ein Studienbuch*, 187-204. Münster: Waxmann Verlag
- Retelsdorf, J. & Möller, J. (2016). Diagnostik und Leistungsbeurteilung. In: J. Möller, M. Köller & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung: Schule und Unterricht – Lehren und Lernen*, Kallmeyer, 23-38
- SenBJF (2015). *Rahmenlehrplan für das Fach Chemie – Teil C*,
https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche_Fassung/Teil_C_Chemie_2015_11_10_WEB.pdf
- Südkamp, A., Möller, J. & Pohlmann, B. (2008a). Der Simulierte Klassenraum. Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22, 261-276.
- Südkamp, A., Möller, J. & Pohlmann, B. (2008b). Der Simulierte Klassenraum. Ein Instrument zur Untersuchung von diagnostischer Kompetenz. In: E.-M. Lankes: *Pädagogische Professionalität als Gegenstand empirischer Forschung*, Waxmann 2008, 87-97
- Südkamp, A., Praetorius, A.-K. (Hrsg.) (2017). Diagnostische Kompetenz von Lehrkräften. *Pädagogische Psychologie und Entwicklungspsychologie*, 94, Waxmann 2017