

Sascha Wittchen<sup>1</sup>  
Claus Bolte<sup>1</sup>  
Nils Machts<sup>2</sup>  
Jens Möller<sup>2</sup>

<sup>1</sup>Freie Universität Berlin  
<sup>2</sup>Christian-Albrechts-Universität zu Kiel

## Analyse chemiedidaktisch relevanter Faktoren der Leistungsdiagnostik

### Einleitung

Wie sich im Rahmen der COACTIV-Studie zeigte, hängen die diagnostischen Kompetenzen Lehrender in praktisch bedeutsamer Weise positiv mit den Leistungen ihrer Schüler\*innen zusammen (Brunner, Anders, Hachfeld & Krauss, 2011, S. 229-230). Dementsprechend ist es nicht verwunderlich, dass die KMK die Förderung diagnostischer Kompetenzen angehender Lehrer\*innen in den Standards für Lehrerbildung verankerte (KMK, 2019, S. 11-12). Unklar ist jedoch, wie es um die diagnostischen Kompetenzen (angehender) Lehrer\*innen bestellt ist und inwiefern sich der Stand der Kompetenzentwicklung zu verschiedenen Zeitpunkten der Ausbildung unterscheidet. Um diesen Fragen nachzugehen, haben wir den Simulierten Klassenraum Chemie, ein digitales Tool zur Erfassung dreier Dimensionen diagnostischer Kompetenzen (angehender) Chemie-Lehrer\*innen, entwickelt.

### Theorie

Das Ausmaß diagnostischer Kompetenzen von Lehrer\*innen wird häufig anhand der Übereinstimmung ihrer Leistungsbeurteilungen mit den durch einen Test erfassten Leistungen der zu beurteilenden Schüler\*innen erfasst (Südkamp, Kaiser & Möller, 2012, S. 756; Südkamp & Praetorius, 2017, S. 34). Dabei werden laut Schrader (1989, S. 86-89) drei Komponenten der Urteilsgenauigkeit unterschieden: die *Rangkomponente*, die *Niveauelemente* und die *Differenzierungskomponente*. Die *Rangkomponente* beschreibt, inwiefern die Beurteilungen der Schüler\*innen-Leistungen die leistungsbezogene Rangfolge der Schüler\*innen abbilden. Da Schüler\*innen, die in einer bestimmten Klasse besonders leistungsstark oder -schwach sind, in einer anderen Klasse nicht mehr als leistungsstark bzw. -schwach gelten würden, ist es ebenfalls von großem Interesse, inwiefern die Leistungsbeurteilungen das mittlere Leistungsniveau einer Klasse abbilden. Die Abbildung des mittleren Leistungsniveaus einer Klasse wird anhand der *Niveauelemente* quantifiziert. Darüber hinaus ist es, insbesondere mit Blick auf die Auswahl geeigneter Maßnahmen und Materialien zur Binnendifferenzierung, bedeutsam, inwiefern die einzelnen Leistungsbeurteilungen eines Lehrers/einer Lehrerin die Leistungsheterogenität der Klasse abbildet. Die Genauigkeit der Abbildung der Leistungsheterogenität wird mit Hilfe der *Differenzierungskomponente* beschrieben.

Neben den Komponenten der Urteilsgenauigkeit beschreibt Schrader (1989, S. 57) drei Arten von Beurteilungen, die sich im Hinblick auf den Beurteilungsgegenstand unterscheiden. Dabei handelt es sich um die Beurteilungen von Aufgabenschwierigkeiten, Beurteilungen der Leistungen einzelner Schüler\*innen in Bezug auf einzelne Aufgaben, sowie Beurteilungen der Gesamtleistungen einzelner Schüler\*innen.

Auf Basis unserer theoriegeleiteten Überlegungen zielt unsere Arbeit auf die Beantwortung folgender Forschungsfragen ab:

- *Wie akkurat beurteilen Lehramtsstudierende mit Fach Chemie die Leistungen von Schüler\*innen unter Berücksichtigung jeweils unterschiedlicher Aufgabenschwierigkeiten und der Qualität der Antworten der Schülerinnen und Schüler?*
- *Inwiefern verändert sich die Genauigkeit der jeweils unterschiedlichen Leistungsbeurteilungen, wenn dieselben Schülerinnen und Schüler (dieselbe Klasse) ein zweites Mal unterrichtet und beurteilt werden?*
- *Inwiefern unterscheiden sich die Genauigkeiten der jeweils unterschiedlichen Leistungsbeurteilungen in verschiedenen Abschnitten der ersten Ausbildungsphase?*

### **Methode**

Zwecks der Beantwortung unserer Forschungsfragen haben wir den *Simulierten KlassenRaum* (SKR; Fielder et al. 2002; Südkamp, 2010) adaptiert und den *Simulierten KlassenRaum Chemie* (SKR-Chemie; Wittchen, Bolte Machts & Möller, 2022) entwickelt. Der SKR-Chemie erlaubt die Simulation von Unterrichtsgesprächen zwischen Chemie-Lehrer\*innen und 12 Schüler\*innen (sechs Jungen und sechs Mädchen) der Doppeljahrgangsstufe 7/8. Die Versuchsteilnehmer\*innen interagieren mit den simulierten Schüler\*innen, indem Sie ihnen Fragen und Aufgaben stellen. Dazu können sie aus einem Pool von Impulsen auswählen, die drei unterschiedliche Schwierigkeitsniveaus abbilden. Auf einen gegebenen Impuls hin melden sich einige der simulierten Schüler\*innen in Abhängigkeit ihrer festgelegten Motivationsparameter. Aufgerufene Schüler\*innen antworten in Abhängigkeit ihrer vorab festgelegten Leistungsparameter und des Schwierigkeitsniveaus der Aufgabe mit einer gewissen Wahrscheinlichkeit (siehe Wittchen, Bolte, Machts & Möller, 2022) entweder *fachlich richtig*, *in Teilen fachlich richtig/unvollständig* oder *fachlich falsch*. Die Versuchsteilnehmer\*innen werden auf die Antwort des Schülers/der Schülerin hin gebeten, diese entlang der drei oben genannten Kategorien zu beurteilen. Nach Beendigung des zwanzigminütigen Unterrichtsgesprächs werden die Versuchsteilnehmer\*innen aufgefordert, die Leistungen der einzelnen Schüler\*innen zu beurteilen. Dabei sollen sie sowohl die Gesamtleistung ungeachtet der Aufgabenschwierigkeiten als auch die Leistungen der Schüler\*innen bezüglich der drei Schwierigkeitsniveaus beurteilen. Die Leistungsbeurteilungen werden anschließend entsprechend der Beurteilungskomponenten nach Schrader (1989, S. 86-89) hinsichtlich ihrer Genauigkeiten analysiert. Als Referenz dienen hierbei die vorab eingestellten Leistungsparameter der simulierten Schüler\*innen.

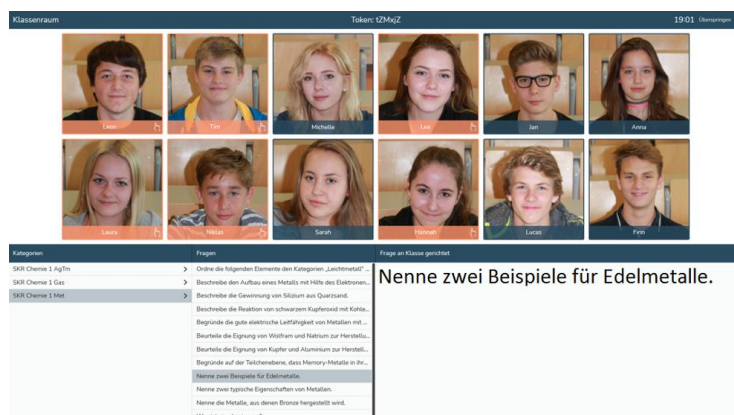


Abbildung 1: User-Interface des SKR-Chemie. Die orangene Hinterlegung der Schülernamen zeigt eine Meldung an.

## Ergebnisse

Insgesamt haben 59 Lehramtsstudierende des Faches Chemie an unserer Untersuchung teilgenommen. Davon befanden sich 35 Studierende in der Bachelor- und 24 Studierende in der Master-Phase. Die Ergebnisse unserer Studie sind der Tabelle 1 zu entnehmen.

Tabelle 1: Beurteilungskomponenten differenziert nach zu berücksichtigenden Schwierigkeitsniveaus und Durchläufen. Die jeweils erste Zeile enthält die Werte der Bachelor-Studierenden, während die jeweils zweite Zeile die Werte der Master-Studierenden enthält.

	Optimum	Schwierigkeitsniveau							
		alle		leicht		moderat		schwierig	
Durchlauf	-	1	2	1	2	1	2	1	2
Rangkomponente	1	0.22	0.44	0.18	0.33	0.22	0.28	0.20	0.33
		0.27	0.34	0.16	0.30	0.19	0.28	0.22	0.23
Niveauelemente	0	0.20	0.18	0.04	0.02	0.16	0.14	0.31	0.26
		0.12	0.14	-0.06	-0.06	0.08	0.05	0.20	0.20
Differenzierungs-komponente	1	0.84	0.98	0.73	1.06	0.97	1.28	1.04	1.24
		0.83	0.97	0.80	1.02	0.90	1.03	1.05	1.09

Die Werte der *Rangkomponente* zeigen, dass es den Studierenden nach dem ersten Durchlauf kaum gelang, die leistungsbezogene Rangfolge der Schüler\*innen in ihren Beurteilungen abzubilden. Nach dem zweiten Durchlauf ist eine substanzielle Steigerung der Urteils-genauigkeiten zu verzeichnen. Die Ergebnisse deuten ferner darauf hin, dass es unseren Bachelor-Studierenden etwas besser gelang, die Leistungsrangfolge der Schüler\*innen abzubilden. Die Ergebnisse bezüglich der *Niveauelemente* lassen erkennen, dass unsere Probanden die mittlere Gesamtleistungen der Klasse deutlich überschätzen. Die nach Schwierigkeitsniveaus differenzierten Ergebnisse zeigen für beide Teilstichproben einen zunehmenden Trend zur Überschätzung des mittleren Leistungsniveaus mit zunehmender Schwierigkeit der zu berücksichtigenden Aufgaben (vgl. Wittchen, Bolte, Machts & Möller, 2022). Dieser Effekt scheint bei Master-Studierenden jedoch geringer ausgeprägt zu sein. Die Werte der *Differenzierungskomponente* zeigen, dass es den Versuchsteilnehmer\*innen recht gut gelang, die Leistungsheterogenität der Klasse in ihren Urteilen widerzuspiegeln. Auch in

diesem Aspekt der Urteilsgenauigkeit zeigt sich, dass Master-Studierende akkuratere Beurteilungen vornehmen. Insgesamt deuten die Ergebnisse (erwartungskonform) auf stärker ausgeprägte diagnostische Kompetenzen auf Seiten der Master-Studierenden hin.

### Fazit & Ausblick

Wie unsere Ergebnisse zeigen, ermöglicht der *SKR-Chemie* die differenzierte Analyse der diagnostischen Kompetenzen (angehender) Chemie-Lehrer\*innen zu verschiedenen Zeitpunkten der ersten Ausbildungsphase. In zukünftigen Studien werden wir die diagnostischen Kompetenzen von Lehramtsstudierenden und Lehramtsanwärter\*innen des Faches Chemie vergleichend untersuchen. Darüber hinaus arbeiten wir bereits an der Konzeption eines *SKR* für den integrierten naturwissenschaftlichen Anfangsunterricht der Doppeljahrgangsstufe 5/6 (*SKR-Nawi*). Schlussendlich gilt es, die Eignung des Tools zur systematischen und individuellen Förderung diagnostischer Kompetenzen während der ersten und zweiten Ausbildungsphase (quasi-)experimentell zu überprüfen.

### Literatur

- Brunner, M., Anders, Y., Hachfeld, A. & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In: M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Fiedler, Walther, E., Freytag, P., & Plessner, H. (2002). Judgment Biases in a Simulated Classroom—A Cognitive-Environmental Approach. *Organizational Behavior and Human Decision Processes*, 88(1), 527–561. <https://doi.org/10.1006/obhd.2001.2981>
- KMK – Sekretariat der ständigen Kultusministerkonferenz der Länder in der Bundesrepublik Deutschland (2019). Standards für die Lehrerbildung: Bildungswissenschaften. Abgerufen unter: [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2002/2004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2002/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf).
- Schrader, F.-W. (1989). Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts. Frankfurt am Main: Peter Lang (S. 57).
- Südkamp, A. (2010). Diagnostische Kompetenz: Zur Genauigkeit der Beurteilung von Schülerleistungen durch Lehrkräfte. Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Christian-Albrechts-Universität zu Kiel.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, (S. 743-762).
- Südkamp, A. & Praetorius, A.-K. (2017). Diagnostische Kompetenz von Lehrkräften. Münster: Waxmann. (S. 21-22).
- Wittchen, S., Bolte, C., Machts, N. & Möller, J. (2022). Erfassung diagnostischer Kompetenzen Lehramtsstudierender des Faches Chemie. In: S. Habing (Hrsg.), *Unsicherheiten als Element von naturwissenschaftsbezogenen Bildungsprozessen*. Universität Duisburg-Essen.