

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

## **Erfassung diagnostischer Kompetenzen Lehramtsstudierender des Fachs Chemie**

### **Einleitung**

Die akkuraten Beurteilungen der Leistungen Lernender sowie ihrer individuellen Lernvoraussetzungen und –fortschritte gehören zu den wichtigsten Aufgaben und Kompetenzen Lehrender (Kaiser & Möller, 2017, S. 56), denn solche Beurteilungen bilden die Basis für die angemessene Vorbereitung, Durchführung und Nachbereitung des Unterrichts (McElvany et al., 2009, S. 223). Somit ist es naheliegend, anzunehmen, dass sich die Fähigkeit zur akkuraten Einschätzung leistungsbezogener Merkmale positiv auf den Lernzuwachs der Lernenden auswirkt (Möller et al., 2016, S. 16). Demzufolge sind zuverlässige Analysen zur diagnostischen Kompetenz (angehender) Lehrer\*innen von großer Bedeutung.

### **Theorie**

In der empirischen Bildungsforschung ist es üblich, die diagnostische Kompetenzen anhand der Übereinstimmung eines Urteils mit der tatsächlichen Ausprägung des zu beurteilenden Merkmals zu operationalisieren (Schrader & Praetorius, 2018, S. 93). Diese grundlegende Definition der Urteilsgenauigkeit bildet unter anderem den Kern des *Heuristischen Modells der Akkuratheit diagnostischer Urteile von Lehrkräften* (Südkamp, Kaiser & Möller, 2012, S. 756; Südkamp & Praetorius, 2017, S. 34). Dieses Modell beschreibt zwei Gruppen von Variablen, welche das Urteil eines Lehrenden beeinflussen: nämlich Lehrermerkmale und Urteilsmerkmale. Darüber hinaus beschreibt das Modell zwei weitere Gruppen von Variablen, die die Schülerleistungen beeinflussen: nämlich Schülermerkmale und Testmerkmale. Allen vier Gruppen von Variablen ist gemein, dass sie die Urteilsgenauigkeit unmittelbar oder zumindest mittelbar beeinflussen.

Entsprechend der Ausführungen Schraders (1989, S. 86-89) ist die Genauigkeit eines Urteils anhand dreier statistischer Kennwerte zu bestimmen; der *Rangkomponente*, der *Niveauelemente* und der *Differenzierungskomponente*. Die *Rangkomponente* beschreibt die Übereinstimmung der durch einen Beurteiler gebildeten leistungsbezogenen Rangfolge der Lernenden mit der tatsächlichen leistungsbezogenen Rangfolge der Lernenden. Neben der Fähigkeit, die Leistungen der Schüler\*innen richtig einzuordnen, ist auch von Interesse, wie genau Lehrer\*innen das mittlere Leistungsniveau einer Lerngruppe (z.B. einer Klasse) beurteilen. Die Einschätzung des mittleren Leistungsniveaus wird mit Hilfe der *Niveauelemente* beschrieben. Die *Differenzierungskomponente* beschreibt das Verhältnis zwischen der geschätzten Streuung der Schülerleistungen und der tatsächlichen Streuung der Schülerleistungen (Schrader, 1989, S. 86-89).

Über diese drei *Komponenten der Urteilsgenauigkeit* hinaus beschreibt Schrader (1989, S. 57) unterschiedliche *Urteilsdimensionen*, die sich in Bezug auf die zu beurteilenden Merkmale unterscheiden. Entsprechend dieser Darlegungen Schraders (1989, S. 57) sind *personenbezogene* und *aufgabenbezogene Urteilsdimensionen* zu unterscheiden. *Personenbezogene Urteile* beschreiben die Beurteilungen der leistungsbezogenen Merkmale

einzelner Schülerinnen und Schüler. *Aufgabenbezogene Urteile* beziehen sich auf die Einschätzungen der Schwierigkeit einzelner Aufgaben.

Auf Basis dieser theoriegeleiteten Überlegungen, zielt unsere Arbeit auf folgende Forschungsfragen ab:

1. *Wie gut gelingt es angehenden Lehrer\*innen, die Leistungen simulierter Schüler\*innen akkurat zu beurteilen, wenn sie personenbezogene und aufgabenbezogene Beurteilungen vornehmen müssen?*
2. *Inwiefern unterscheiden sich die Genauigkeiten der Leistungsbeurteilungen, wenn unterschiedliche Schwierigkeitsniveaus von Aufgaben bei der Beurteilung der Schüler\*innen differenziert zu beachten sind?*

### **Methode**

Um unsere Forschungsfragen zu untersuchen, haben wir den *Simulierten KlassenRaum Chemie (SKR-C*; siehe Bolte et al., 2021; 2011), eine fachspezifische Version des Simulierten Klassenraums (*SKR*), weiterentwickelt. Beim *SKR* handelt es sich um ein digitales Tool zur Simulation komplexer und dynamischer Interaktionen zwischen Lehrenden und Lernenden (Südkamp, 2010; Kaiser et al., 2013). Im *SKR-C* interagieren die Versuchsteilnehmer mit 12 simulierten Schüler\*innen (sechs Jungen, sechs Mädchen), indem sie ihnen Fragen oder Aufgaben stellen. Daraufhin melden sich die simulierten Schüler\*innen in Abhängigkeit der im Voraus festgelegten Motivationsparameter. Aufgerufene Schüler\*innen antworten nun entsprechend ihrer vorab eingestellten Leistungsparameter. Nach einer gewissen Zeit (in dieser Studie nach 20 Minuten) beurteilen die Probanden die Leistungen der simulierten Schüler\*innen über alle Aufgaben hinweg sowie differenziert nach Schwierigkeitsniveaus. Auf der Basis der Beurteilungen der Proband\*innen und der voreingestellten Leistungsparameter der simulierten Schüler\*innen wird anhand der Beurteilungskomponenten nach Schrader (1989, S. 86-89) die Urteilsgenauigkeit der Proband\*innen berechnet.

Für unsere Studie haben wir einen Katalog von Aufgaben und Fragen entworfen, die für die Klassen 7 und 8 geeignet sind und sich drei Schwierigkeitsniveaus zuordnen lassen. Wir haben für jede der eigens konstruierten Aufgaben je drei Antworten entwickelt. Die drei korrespondierenden Antworten repräsentieren drei Qualitätsniveaus, die den Kategorien „*fachlich richtig*“, „*in Teilen fachlich richtig/unvollständig*“ und „*fachlich falsch*“ zuzuordnen sind. Um in Erfahrung zu bringen, wie akkurat die Probanden die Aufgabenschwierigkeiten beurteilen, haben wir unseren Versuchsteilnehmer\*innen den Katalog der Fragen/Aufgaben bereits vor dem Start des *SKR-C* vorgelegt und sie aufgefordert, die Aufgaben bezüglich ihres Anforderungsniveaus zu beurteilen.

Die Schüler\*innen des *SKR-C* wurden paarweise (jeweils zwei Jungen und zwei Mädchen) drei Leistungsniveaus zugewiesen. Entsprechend ihrer Leistungsniveaus wurden den simulierten Schüler\*innen jeweils neun Leistungsparameter zugewiesen (siehe Tabelle 1).

*Tabelle 1: Verteilung der Leistungsparameter der Schüler\*innen für verschiedene Schwierigkeitsniveaus (linke Zahl gibt die Wahrscheinlichkeit einer "fachlich richtigen" Antwort an; mittlere Zahl gibt die Wahrscheinlichkeit einer "In Teilen richtigen/unvollständigen" Antwort; rechte Zahl gibt die Wahrscheinlichkeit einer "fachlich falschen" Antwort an).*

	Leichte Aufgabe	Mittlere Aufgabe	Schwierige Aufgabe
Starke Schüler*innen	100/0/0	75/15/10	50/30/20
Mittelmäßige Schüler*innen	75/15/10	50/30/20	25/35/40
Schwache Schüler*innen	50/30/20	25/35/40	0/0/100

## Ergebnisse

Insgesamt haben 17 Lehramtsstudierende im Masterstudiengang an unserer Untersuchung teilgenommen. Die Ergebnisse unserer Studie sowie die Ergebnisse einer ausgewählten Referenzuntersuchung sind der Tabelle 2 zu entnehmen.

Tabelle 2: Ergebnisse dieser Studie differenziert nach Beurteilungskomponenten und Schwierigkeitsniveau sowie Ergebnisse einer ausgewählten Referenzuntersuchung.

	Optim.	Schwierigkeitsniveau				Bolte et al. (2021)	
		unabhängig	Niedrig	Mittel	Hoch	Nawi	Chemie
Rangkomponente	1	0,23	0,32	0,17	0,21	0,37	0,54
Niveauelemente	0	-0,04	-0,13	0,20	0,33	0,04	0,01
Differenzierungs-komponente	1	1,13	0,88	0,84	0,96	0,49	0,61

Die vergleichsweise niedrig ausfallenden *Rangkomponenten* deuten darauf hin, dass es unseren Probanden nur in geringem Maße gelungen ist, die korrekte leistungsbezogene Rangfolge der Schüler\*innen zu bilden. Dieser Befund deckt sich insofern mit unseren Erwartungen, als dass unsere Versuchsteilnehmer im Gegensatz zu vorangegangenen Untersuchungen mehr Beurteilungsdimensionen berücksichtigen mussten, also anspruchsvollere diagnostische Leistungen zu erbringen hatten. Die Ergebnisse zur *Niveauelementen* zeigen, dass unsere Probanden das mittlere Leistungsniveau der Schüler\*innen insgesamt eher unterschätzt haben. Die differenzierte Betrachtung der Niveaubewertungen deutet darauf hin, dass die Versuchsteilnehmer dazu neigen, mit zunehmendem Anforderungsniveau das mittlere Leistungsniveau zu überschätzen. Mit Blick auf die *Differenzierungskomponente* der schwierigkeitsunabhängigen Beurteilung der Leistung zeigt sich eine Tendenz zur Überschätzung der Leistungsstreuungen. Gleichzeitig deuten die nach Aufgabenschwierigkeit differenzierten Analysen der *Differenzierungskomponenten* auf eine leichte Unterschätzung der Streuung der Leistungen hin. Der Vergleich mit der Referenzuntersuchung von Bolte et al. (2021) lässt erkennen, dass es unseren Probanden deutlich besser gelang, die Streuung der Schüler\*innen-Leistungen akkurat zu beurteilen. Dieser Umstand lässt sich vermutlich dadurch erklären, dass den Probanden in unserer Untersuchung Aufgaben verschiedener Schwierigkeitsniveaus zur Verfügung standen; das mag eine akkurate Erfassung der Schüler\*innen-Leistungen erleichtert haben.

## Fazit & Ausblick

Der SKR-C ermöglicht einen differenzierteren Einblick in die Ausprägungen diagnostischer Kompetenzen (angehender) Chemie-Lehrer\*innen. In nachfolgenden Studien werden wir der Frage nachgehen, inwiefern Interdependenzen zwischen den Genauigkeiten *personenbezogener* und *aufgabenbezogener Beurteilungen* bestehen. Darüber hinaus wollen wir untersuchen, inwiefern sich diagnostische Kompetenzen angehender Chemie-Lehrer\*innen zu verschiedenen Ausbildungszeitpunkten (z.B. in der ersten und der zweiten Ausbildungsphase) unterscheiden. Außerdem werden wir überprüfen, inwiefern sich diagnostische Kompetenzen angehender Lehrer\*innen mit Fach Chemie verändern, wenn diese den *SKR-C* mehrmals durchlaufen.

## Literatur

- Bolte, Köppen, Möller & Südkamp (2011). Kompetenzdiagnostik im virtuellen naturwissenschaftlichen Unterricht. In: D. Höttecke (2011; Hg.). *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Zur Didaktik der Physik und Chemie. Probleme und Perspektiven.* Berlin/Münster: Lit-Verlag.
- Bolte, C., Stollin, F., Möller, J., & Südkamp, A. (2021). Analyse diagnostischer Kompetenzen von (angehenden) Chemielehrer\*innen. In Ch. Maurer, K. Rincke, & M. Hemmer (Hrsg.). *Fachliche Bildung und digitale Transformation - Fachdidaktische Forschung und Diskurse.* (S. 173-176). Regensburg: Universität Regensburg.
- Kaiser & Möller (2017). Diagnostische Kompetenz von Lehramtsstudierenden. In: Gräsel & Trempler (2017). *Entwicklung von Professionalität pädagogischen Personals*, 55 – 73. Wiesbaden: Springer.
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, 28, pp. 73–84. <http://dx.doi.org/10.1016/j.learninstruc.2013.06.001>
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für pädagogische Psychologie (Bern, Switzerland)*, 23(3-4), 223–4<223–235. <https://doi.org/10.1024/1010-0652.23.34.223>
- Möller, Jens & Machts, Nils & Retelsdorf, Jan. (2016). Diagnostische Kompetenz von Lehrkräften. Merkmale, Instrumente, Urteilsverzerrungen. *Schulmanagement*. 47. 14-17.
- 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).
- Schrader, F.-W., Praetorius, A.-K. (2018). Diagnostische Kompetenz von Eltern und Lehrern. In: D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (pp. 92 – 98), Weinheim: Beltz.
- 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).