

Südkamp, A.; Kaiser, J.; Möller, J. (2012): Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. In: *Journal of Educational Psychology* 104 (3), S. 743-762.

Stichworte

Diagnostische Kompetenz, Meta-Analyse, Beurteilungsgenauigkeit

Einleitung

In diesem Artikel werden systematisch Befunde aus Studien zur Beurteilungsgenauigkeit von Schüler*innenleistungen aufgearbeitet und zusammengefasst. Die Beurteilungsgenauigkeit wird als Hauptaspekt der diagnostischen Kompetenz bei Lehrkräften verstanden. Zur Einordnung möglicher Einschränkungen der Beurteilungsgenauigkeit wird außerdem ein heuristisches Modell von Einflussfaktoren diskutiert.

Diagnostische Kompetenz

Diagnostische Kompetenz gilt als ein breitgefächertes Konstrukt, das mehrere Teilbereiche umfasst: die Einschätzung und die Gestaltung des Unterrichts- und Prüfungsmaterials, Einschätzungen von Schüler*innenmerkmalen, -arbeiten und -verhalten sowie sowohl summative (z. B. die Beurteilungen auf Zeugnissen) als auch formative Rückmeldeformate (z. B. Feedback auf spezifische Aufgabenbearbeitung). Ein Großteil der bestehenden Forschung zur diagnostischen Kompetenz konzentriert sich auf die Genauigkeit der Beurteilungen von Schüler*innenmerkmalen und dabei insbesondere auf die Beurteilung der schulischen Leistungen.

Zentrale Maße der Beurteilungsgenauigkeit

Grundsätzlich wird die Genauigkeit der Beurteilungsgenauigkeit als Übereinstimmung der Merkmalsbeurteilung durch eine Lehrkraft mit der tatsächlichen Merkmalsausprägung betrachtet. Die tatsächliche Merkmalsausprägung wird bei Studien im Feld anhand standardisierter Tests bestimmt. Als Maß der Übereinstimmung wird üblicherweise die Korrelation

der Testleistung der Schüler*innen mit der Leistungseinschätzung durch die Lehrkraft verwendet. In Fällen, in denen die Beurteilungen und die Schüler*innenleistungen auf der gleichen Skala erfolgen, kann für jede Lehrkraft auch bestimmt werden, ob das Leistungsniveau der Klasse den Beurteilungen entspricht und ob die Streuung der Schüler*innenleistungen mit der Streuung der Beurteilungen übereinstimmt. In der zusammengefassten Studie wird nur die Korrelation als Maß der Beurteilungsgenauigkeit berücksichtigt.

Eine Meta-Analyse zur leistungsbezogenen Beurteilungsgenauigkeit

In der dargestellten Arbeit wurden 75 Studien zur Beurteilungsgenauigkeit von Lehrkräften in einer Meta-Analyse systematisch zusammengefasst. Darin wurden die in den Studien berichteten Korrelationen von leistungsbezogenen Lehrer*innenurteilen und der tatsächlichen Schüler*innenleistung gesammelt und deren statistischer Mittelwert berechnet. Unterschiede zwischen den einzelnen in den Studien berichteten Korrelationen legen weitere Analysen nahe, um herauszufinden, ob sich die Studien in ihrem Vorgehen und den untersuchten Stichproben systematisch unterschieden.

Zentraler Befund der Meta-Analyse

Das zentrale Ergebnis der Meta-Analyse ist die mittlere Beurteilungsgenauigkeit über alle zusammengefassten Studien hinweg. Sie beträgt $r = .63$. Diese Korrelation kann als hoch bewertet werden und bedeutet, dass die durch Tests gemessenen Schüler*innenleistungen etwa 40 % der Varianz in den Lehrerurteilen erklären. Gleichzeitig bedeutet dieses Ergebnis, dass in Lehrkräfteurteilen substantielle Anteile enthalten sind, die nicht durch die messbaren Schüler*innenleistungen abgedeckt werden.

Weitere Befunde der Meta-Analyse

Es zeigten sich Unterschiede in den berichteten Korrelationen zwischen den Studien. Weitere Analysen sollten Rückschlüsse darüber erlauben, ob Studien, die höhere Korrelationen berichteten, sich systematisch von Studien mit niedrigeren Korrelationen unterscheiden.

In der Meta-Analyse enthaltene Studien, bei denen die untersuchten Lehrkräfte darüber informiert waren, mit welchem Testverfahren (z. B. ein Test zur Messung des Leseverständnisses) ihre Beurteilungen verglichen wurden, berichteten im Mittel höhere Korrelationen als Studien, in denen Lehrkräfte Einschätzungen ihrer Schüler*innen abgaben, ohne über die Vergleichsnorm informiert zu sein. Dies ist ein Befund, der dafürspricht, dass Lehrkräfte das konkrete Verfahren zur Leistungsmessung effektiv in ihre Leistungseinschätzungen der Schüler*innen mit einbeziehen können.

Ein ähnlicher Befund ergab sich auf abstrakterer Ebene. In der Meta-Analyse enthaltene Studien, bei denen die Einschätzungen und die Testverfahren ähnlich spezifisch oder unspezifisch waren, berichteten höhere Korrelationen als Studien, bei denen z. B. ein unspezifisches Leistungsmerkmal eingeschätzt und die Einschätzung mit einem sehr spezifischen Testverfahren verglichen wurde. Schätzen Lehrkräfte z. B. die allgemeine Deutschkompetenz ihrer Schüler*innen ein, so fällt die Korrelation mit der gemessenen Leseleistung im Durchschnitt vergleichsweise niedriger aus als die Korrelation mit einem Deutschtest, der mehrere Facetten der Deutschkompetenz umfasst.

Heuristisches Modell der Beurteilungsgenauigkeit

Die vorliegende Meta-Analyse wurde um ein heuristisches Modell ergänzt, in dem mögliche Einschränkungen der Beurteilungsgenauigkeit systematisch auf Einflussfaktoren zurückgeführt werden. Diese Einflussfaktoren lassen sich auf vier Ebenen verorten. Dabei handelt es

sich um die Ebene *Lehrkräfte*, die sich systematisch in ihren Beurteilungstendenzen unterscheiden können (z. B. in der Milde oder der Strenge ihrer Urteile), die Ebene der *Schüler*innen*, die sich systematisch in anderen Merkmalen als den konkreten Leistungen unterscheiden können (z. B. Geschlecht oder Migrationshintergrund), die Ebene des *Beurteilungsformats* (z. B. die Vergabe einer Zeugnisnote oder die Schätzung der Anzahl erzielter Punkte auf einem Test) und die Ebene des *Testverfahrens* (z. B. dessen Reliabilität oder Domänenspezifität). Einige dieser Ebenen können sich gegenseitig beeinflussen, so mögen beispielsweise einige Lehrkräfte einem größeren Geschlechterstereotyp unterliegen als andere und sich deswegen unterschiedlich vom Schüler*innenmerkmal *Geschlecht* in ihren Urteilen beeinflussen lassen.

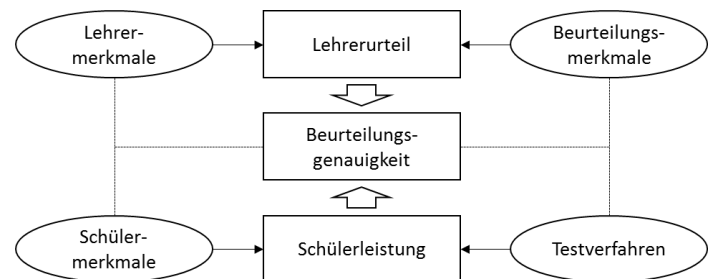


Abb. 1. Heuristisches Modell der Beurteilungsgenauigkeit

Fazit

Die Genauigkeit von Lehrkräften bei der Beurteilung von Schüler*innenleistungen kann über eine Vielzahl von Studien als hoch bewertet werden. Liegt den Lehrkräften der Standard für den Leistungsvergleich vor und ist gewährleistet, dass die Beurteilung genauso spezifisch ist wie der Vergleichsmaßstab, so fällt die Genauigkeit vergleichsweise höher aus. Dennoch ist die Beurteilungsgenauigkeit weit von optimaler Güte entfernt, was zeigt, dass Einflussfaktoren auf verschiedenen Ebenen die Beurteilungsgenauigkeit einschränken. [nm]

