



Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften *

Accuracy of Teacher Judgements and Considerations on their Improvement via Teacher Education

Constance Karing und Cordula Artelt

Zusammenfassung Der vorliegende Beitrag beschäftigt sich mit der Urteilsgenauigkeit von Lehrkräften. Nach einer kurzen Übersicht zum gegenwärtigen Forschungsstand werden Befunde aus unseren eigenen Arbeiten, die im Rahmen der BiKS-Studie entstanden sind, berichtet. Diese beziehen sich zum einen auf die Spezifität der Urteile (globale Urteile versus aufgabenspezifische Urteile) und zum anderen auf die Auswirkungen der Urteilsgenauigkeit der Lehrkräfte auf den Lernerfolg der Schülerinnen und Schüler. Der Beitrag schließt mit Implikationen für die Aus- und Weiterbildung von Lehrkräften.

Schlagwörter *diagnostische Kompetenz – Urteilsgenauigkeit – Spezifität der Urteile – Lehrpersonenaus- und -weiterbildung*

Abstract The following article addresses teacher judgement accuracy. After a brief summary of the current state of research, findings from our own work, which are based on data from the BiKS-Study, are reported. These findings focus on the specificity of teacher judgements (global judgements versus task-specific judgements) and on the effects of teacher judgement accuracy on students' learning success. The article closes with implications for initial teacher education and in-service training.

Keywords *diagnostic competence – teacher judgement accuracy – judgement specificity – initial teacher education and in-service training*

1 Einleitung

Ein wichtiger Bestandteil der beruflichen Tätigkeit von Lehrkräften ist neben dem Unterrichten, Erziehen, Innovieren und Beraten auch das Diagnostizieren und Beurteilen (Schrader, 2011; Ständiges Sekretariat der Kultusminister der Länder der Bundesrepublik Deutschland, 2004). Diagnostische Urteilsleistungen sind sowohl im Rahmen von mündlichen und schriftlichen Leistungsüberprüfungen, bei Schullaufbahnpflichtempfehlungen sowie bei der Feststellung von Lern- und Arbeitsstörungen erforderlich, als auch bei der Unterrichtsplanung und -gestaltung, wenn es darum geht, die Fähigkeiten der Schülerinnen und Schüler und die unterrichtlichen Anforderungen aufeinander abzustimmen, um eine möglichst optimale Lernumgebung zu gestalten (Helmke, Hosenfeld & Schrader, 2004; Schrader, 2011). In der pädagogisch-psychologischen Forschung wird die Genauigkeit solcher Urteilsleistungen unter dem Stichwort «Diagnostische Kompetenz» diskutiert (Artelt, 2009). Damit gemeint ist die Kompetenz der Lehrkräfte, Merkmale der Schülerinnen und Schüler, aber auch Lern- und Aufgabenanforderungen zutreffend einzuschätzen (Artelt, 2009; Artelt & Gräsel, 2009; Helmke et al., 2004). Demnach stellt die Urteilsgenauigkeit eine zentrale Komponente der diagnostischen Kompetenz dar (Kaiser, Helm, Retelsdorf, Südkamp & Möller, 2012).

2 Komponenten der Urteilsgenauigkeit

Um die Genauigkeit solcher Lehrpersonenurteile zu bestimmen, werden in der Regel die Lehrpersonenurteile über bestimmte Merkmale der Schülerinnen und Schüler (z.B. deren Leistungen) mit den tatsächlichen Ausprägungen dieser Merkmale (z.B. Leistungen der Schülerinnen und Schüler in standardisierten Schulleistungstest) verglichen. Zur Ermittlung dieser Übereinstimmung werden in der deutschsprachigen Literatur insbesondere drei Urteilstkomponenten, nämlich die Rangordnungs-, die Niveau- und die Differenzierungskomponente, verwendet, die jeweils unterschiedliche Facetten der Urteilsgenauigkeit abbilden (Karing, Matthäi & Artelt, 2011; Schrader & Helmke, 1987; Spinath, 2005). Die Rangordnungskomponente, die nach Helmke et al. (2004) das Kernstück der diagnostischen Kompetenz abbildet, beschreibt die Kompetenz der Lehrkräfte, die Rangordnung unter verschiedenen Schülerinnen und Schülern in Bezug auf ein bestimmtes Merkmal (z.B. Leistung) akkurat wiederzugeben (Helmke et al., 2004; Schrader & Helmke, 1987). Dafür werden klassenspezifische Korrelationen zwischen den Lehrpersoneneinschätzungen und den tatsächlichen Leistungen der Schülerinnen und Schüler berechnet, die theoretisch zwischen -1 und $+1$ variieren können (Helmke et al., 2004; Schrader & Helmke, 1987). Je höher die

* Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (DFG) (Kennziffer: AR 301/6-1 und AR 301/6-2) im Forschungsprojekt BiKS «Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter» (FOR 543).

2 Constance Karing, Cordula Artelt

Übereinstimmung zwischen Lehrpersonenurteil und den tatsächlichen Leistungen der Schülerinnen und Schüler ist (je näher bei +1), desto akkurater fällt die Diagnose der Lehrkräfte aus (Helmke, Hosenfeld & Schrader, 2003). Dagegen beziehen sich die beiden anderen Urteilsdimensionen auf das Lehrpersonenurteil zum Niveau der Leistungen der Schülerinnen und Schüler bzw. zur Leistungsstreuung. So gibt die Niveauebene dar über Auskunft, inwieweit Lehrkräfte die Ausprägungen der Leistungen der Schülerinnen und Schüler akkurat einschätzen können, während die Differenzierungsdimension angibt, ob Lehrkräfte die Unterschiedlichkeit von Leistungen der Schülerinnen und Schüler in ihrer gesamten Bandbreite wahrnehmen oder unter- bzw. überdifferenzieren (Artelt, 2009). In Untersuchungen zur Urteilsgenauigkeit der Lehrkräfte werden vorrangig Befunde zur Rangordnungsdimension, also zu den Korrelationen zwischen Lehrpersonenurteil und den Leistungen der Schülerinnen und Schüler, berichtet. Die Mehrzahl dieser Studien deutet darauf hin, dass die Lehrkräfte im Mittel relativ angemessene Einschätzungen über Leistungen der Schülerinnen und Schüler vornehmen (z.B. Hoge & Coladarci, 1989: Median von $r = .66$; Südkamp, Kaiser & Möller, 2012: Median von $r = .53$), wobei man über die Bewertung der Angemessenheit der Einschätzungen bei einer Varianzaufklärung von 44 bzw. 28 Prozent natürlich auch geteilter Meinung sein kann. Zudem zeigen sich in diesen Untersuchungen auch beträchtliche Unterschiede zwischen den einzelnen Lehrkräften. So reicht die Spannweite von sehr guten Diagnostikerinnen und Diagnostikern bis hin zu Lehrkräften, deren Urteile so gut wie keine Übereinstimmungen mit den tatsächlichen Leistungen der Schülerinnen und Schüler aufweisen.

3 Spezifität von Urteilen

Verschiedene Forschungsgruppen stellten in ihren Untersuchungen fest, dass sich Unterschiede in der Urteilsgenauigkeit der Lehrkräfte auch in Abhängigkeit von der Spezifität der Urteile zeigen (z.B. Demaray & Elliott, 1998; Feinberg & Shapiro, 2009; Hoge & Coladarci, 1989). Der Grad der Urteilspezifität reicht dabei von Urteilen, die eine geringe Spezifität aufweisen (z.B. globale Lehrpersonenurteile zur Lesekompetenz der Schülerinnen und Schüler wie «Er/Sie kann Texte gut verstehen»), bis hin zu Urteilen, denen eine hohe Spezifität inhärent ist (z.B. aufgabenspezifische Urteile, bei denen die Lehrkraft die Leistung der Schülerinnen und Schüler für einzelne konkrete Aufgaben einschätzt). Bisherige Studien im englischsprachigen Raum fanden für den Grundschulbereich, dass aufgabenspezifische Lehrpersonenurteile in den Bereichen Lesen und Mathematik akkurater ausfielen als globale Urteile (Rangordnungsdimension, Demaray & Elliott,

1998; Feinberg & Shapiro, 2009). Jedoch wiesen diese Studien einige Besonderheiten auf. So bildete beispielsweise die eingesetzte Skala für die globalen Lehrpersonenurteile in der Studie von Demaray und Elliott (1998) nicht nur die Leistung der Schülerinnen und Schüler ab, sondern auch andere Merkmale (z.B. Verhalten, Motivation des Kindes). Da keine Trennung nach diesen verschiedenen Bereichen erfolgte, sondern für das Lehrpersonenurteil ein Gesamtscore gebildet wurde, konnte die Urteilsgenauigkeit nicht getrennt nach einzelnen Bereichen ermittelt werden. Dies kann die Höhe des Zusammenhangs zwischen dem globalen Lehrpersonenurteil und den Leistungen der Schülerinnen und Schüler beeinflussen, denn es zeigte sich bereits mehrfach, dass die Urteilsgenauigkeit der Lehrkräfte auch vom jeweiligen Urteilsgegenstand (Leistung vs. emotional-motivationale Merkmale) abhängig ist (z.B. Karing, 2009; Spinath, 2005).

In unseren eigenen Arbeiten (z.B. Karing, Matthäi & Artelt, 2011), die im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Bamberger Forschungsprojekts BiKS («Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter») entstanden sind, wurde die Akkuratheit von globalen und aufgabenspezifischen Urteilen der Lehrkräfte der Sekundarstufe in den Bereichen Lesen und Mathematik untersucht und verglichen. Die Erhebung der globalen Lehrpersonenurteile erfolgte z.B. zur Lesekompetenz der Schülerinnen und Schüler anhand des Items «Er/Sie kann Texte gut verstehen», während die Lehrkräfte für die Erfassung ihrer aufgabenspezifischen Urteile einen Fragebogen vorgelegt bekamen, der einen einseitigen Prosatext und sieben Multiple-Choice-Aufgaben aus dem Lesekompetenztest umfasste, den alle an der Studie teilnehmenden Schülerinnen und Schüler bearbeitet hatten. Die Lehrkräfte wurden gebeten, für jede der sieben Aufgaben anzugeben, ob die jeweilige Schülerin bzw. der jeweilige Schüler die entsprechende Aufgabe ihrer Einschätzung nach lösen konnte oder nicht. Anhand der globalen und aufgabenspezifischen Einschätzungen konnten die Rangordnungsdimensionen berechnet werden. Es zeigte sich, dass die globalen Urteile der Sekundarschullehrkräfte in den Bereichen Lesen und Mathematik signifikant akkurater ausfielen als ihre aufgabenspezifischen Einschätzungen (vgl. Tabelle 1).

Eine globale Einschätzung scheint den Sekundarschullehrkräften demnach leichter zu fallen als die Beurteilung der Leistung der Schülerinnen und Schüler bei einer konkreten Aufgabe. Ein Grund dafür mag sein, dass die beiden Formen der Personeneinschätzungen unterschiedliche Wissensbestandteile benötigen. Während globale Urteile auf das Wissen über Personen (Fähigkeit der Schülerinnen

Tabelle 1: Akkuratheit von globalen und aufgabenspezifischen Lehrpersonenurteilen in den Bereichen Lesen und Mathematik (Rangordnungskomponente)

	Lesen		Mathematik	
	M	SD	M	SD
Globale Urteile	.34	0.30	.60	0.37
Aufgabenspezifische Urteile	.20	0.52	.46	0.40

Anmerkungen: Die dargestellten Werte sind durchschnittliche Korrelationen (Rangordnungskomponente diagnostischer Kompetenz), die mittels Fisher-Z-Transformationen berechnet wurden. Globale Urteile (Lesen): N = 64 Lehrkräfte, N = 907 Schülerinnen und Schüler; globale Urteile (Mathematik): N = 73 Lehrkräfte, N = 1006 Schülerinnen und Schüler; aufgabenspezifische Urteile (Lesen): N = 64 Lehrkräfte, N = 433 Schülerinnen und Schüler; aufgabenspezifische Urteile (Mathematik): N = 73 Lehrkräfte, N = 497 Schülerinnen und Schüler; 5. Jahrgangsstufe.

und Schüler) in Bezug auf ein allgemeines Konstrukt (z.B. Lesekompetenz) fokussieren, verlangen aufgabenspezifische Urteile sowohl Wissen über Personen als auch Wissen über Aufgabenmerkmale (z.B. Aufgabenschwierigkeit) und die zur Bearbeitung erforderlichen Strategien. Erst durch das Zusammenwirken dieser verschiedenen Wissens Elemente sind akkurate Lehrpersonenurteile möglich. Voraussetzung hierfür ist also nicht nur diagnostische, sondern auch didaktische und fachwissenschaftliche Kompetenz (Helmke et al., 2003; Karing, Matthäi & Artelt, 2011). Da Lehrkräfte hierbei die Fähigkeiten der Schülerinnen und Schüler mit den konkreten Aufgabenanforderungen abgleichen müssen, ist der kognitive Verarbeitungsaufwand bei solchen Urteilen auch höher als bei globalen Fähigkeitseinschätzungen ohne Bezug zu konkreten Einzelanforderungen (Karing, Pfof & Artelt, 2011).

4 Relevanz der Urteils genauigkeit der Lehrkräfte für den schulischen Lernerfolg

Wie bereits oben angesprochen, sind akkurate Lehrpersonenurteile für die adaptive Unterrichtsgestaltung und den Lernerfolg der Schülerinnen und Schüler bedeutsam. Dem liegt die Annahme zugrunde, dass Unterrichtsmaßnahmen dann optimal wirken, wenn eine hinreichende Passung zwischen den Lernvoraussetzungen der Schülerinnen und Schüler und den gestellten Anforderungen im Unterricht vorliegt (Helmke et al., 2004; Schrader, 2010). Eine notwendige Voraussetzung dafür ist aber, dass Lehrkräfte sowohl lern- und leistungsrelevante Merkmale der Schülerinnen und Schüler als auch das Schwierigkeitspotenzial des Unterrichtsmaterials akkurat einschätzen (Anders et al., 2010) sowie die Schwierigkeit der eingesetzten Instruk tionsstrategien und -methoden angemessen beurteilen können. Demnach ist für einen adaptiven und effektiven Unterricht eine enge Verknüpfung von diagnostischen Informationen und didaktischen Unterrichtsmaßnahmen erforderlich (Ingenkamp, 1992; Schrader & Helmke, 2001). Die wenigen Studien, die sich den Auswir-

kungen diagnostischer Kompetenz auf die Leistungsentwicklung von Schülerinnen und Schülern widmen, sind vorrangig im mathematischen Bereich angesiedelt und beziehen sich überwiegend auf die Rangordnungskomponente. Schrader (1989; vgl. Schrader & Helmke, 1987) stellte fest, dass die Urteils genauigkeit der Hauptschullehrkräfte keinen direkten Einfluss auf den Lernerfolg der Schülerinnen und Schüler im Fach Mathematik hatte, sondern sich nur dann positiv auf die Leistungsentwicklung der Klasse auswirkte, wenn im Unterricht gleichzeitig viele Strukturierungs- und Unterstützungsmaßnahmen eingesetzt wurden. Demgegenüber zeigte sich in den Arbeiten von Lehmann et al. (2000) sowie Brunner et al. (2011), dass sich die Urteils genauigkeit der Sekundarschullehrkräfte auf den Lernerfolg der Schülerinnen und Schüler im Fach Mathematik auswirkte. So wiesen z.B. die Schülerinnen und Schüler in der Studie von Brunner et al. (2011) eine höhere Mathematikleistung auf, wenn ihre Lehrkräfte die Leistungsran gfolge der Schülerinnen und Schüler (Rangordnungskomponente) akkurater einschätzen konnten. In unseren eigenen Arbeiten (z.B. Karing, Pfof & Artelt, 2011), die im Rahmen der BiKS-Studie durchgeführt wurden, untersuchten wir den Zusammenhang zwischen der Urteils genauigkeit von Sekundarschullehrkräften und der Leistungsentwicklung der Schülerinnen und Schüler in den Bereichen Lesen und Mathematik von der fünften zur sechsten Jahrgangsstufe. Basis hierfür waren Indikatoren diagnostischer Kompetenz, die sich auf der Grundlage der aufgabenspezifischen Einschätzungen der Lehrkräfte bilden ließen. Als Maß für die Urteils genauigkeit wurde dabei neben der Rangordnungskomponente auch der sogenannte aufgabenspezifische Treffer herangezogen. Im Gegensatz zur Rangordnungskomponente beinhaltet der aufgabenspezifische Treffer einen exakten Abgleich von Lehrpersoneneinschätzung und Leistung der Schülerinnen und Schüler auf der Ebene einzelner Aufgaben. Die Besonderheit dieses Indikators der Urteils genauigkeit liegt also darin, dass dabei beachtet wird, ob die von der Lehrkraft als gelöst (bzw. nicht gelöst) angegebenen Aufgaben

auch tatsächlich dieselben Aufgaben sind, die die Schülerin oder der Schüler löste (bzw. nicht löste). Die wichtigsten Ergebnisse unserer Analysen waren, dass der aufgabenspezifische Treffer, nicht aber die Rangordnungskomponente, positiv mit der Entwicklung der Lesekompetenz der Schülerinnen und Schüler zusammenhing. Bemerkenswert ist dabei, dass dieser Zusammenhang durch Unterrichtsmerkmale moderiert wurde. So zeigte sich eine verstärkte Lesekompetenzentwicklung aufseiten der Schülerinnen und Schüler nur dann, wenn die Lehrkräfte eine hohe Trefferquote aufwiesen und gleichzeitig im Unterricht einen hohen Anteil an individualisierenden Maßnahmen einsetzten. Für den mathematischen Bereich konnten diese Zusammenhänge allerdings nicht nachgewiesen werden. Die Ergebnisse zur Lesekompetenzentwicklung verdeutlichen, dass es nicht allein diagnostischer Kompetenz bedarf, sondern dass es notwendig ist, diagnostisches Wissen in entsprechende Unterrichtsmaßnahmen umzusetzen, damit die Schülerinnen und Schüler in ihrer Kompetenzentwicklung davon profitieren können.

5 Implikationen für die Aus- und Weiterbildung

Obwohl akkuraten Lehrpersonenurteilen eine entscheidende Bedeutung für eine effektive Unterrichtsgestaltung sowie für den Lernerfolg der Schülerinnen und Schüler beigemessen wird, wird dieses Thema selten im Rahmen der Lehrpersonenausbildung und -weiterbildung behandelt (Brunner et al., 2011; Helmke et al., 2004). Angesichts unseres Befundes zu den globalen und aufgabenspezifischen Lehrpersonenurteilen sehen wir einen möglichen Ansatzpunkt zur Verbesserung der Urteilsgenauigkeit insbesondere bei den aufgabenspezifischen Lehrpersonenurteilen. Wie bereits ausgeführt, verlangen aufgabenspezifische Urteile sowohl Personen- als auch Aufgaben- und Strategiewissen, also eine Kombination von diagnostischem, fachwissenschaftlichem und fachdidaktischem Wissen (vgl. Helmke et al., 2003). Daher könnte ein direkter Förderansatz in der Stärkung dieser Wissenskomponenten liegen. Wichtig wären in diesem Zusammenhang vor allem vermehrte institutionelle Angebote, z.B. während des Studiums, die zum Aufbau dieser Wissensinhalte bedeutsam beitragen könnten (vgl. auch Brunner et al., 2006). Gleichwohl sollten auch die globalen Lehrpersonenurteile nicht außer Acht gelassen werden. Zwar fielen diese in unseren Analysen akkurater aus als die aufgabenspezifischen, dennoch ergaben sich auch hier nur schwache bis moderate Ausprägungen und deutliche Unterschiede zwischen den einzelnen Lehrkräften. Vor dem Hintergrund, dass die Genauigkeit dieser Urteile davon abhängig ist, was die Lehrkraft unter diesem Merkmal bzw. Konstrukt versteht und inwieweit sie dafür tatsächlich valide Indikatoren verwendet (Schrader,

2009, 2011), bieten sich hier ebenfalls verschiedene Ansatzpunkte zur Verbesserung der Urteilsgenauigkeit an. So können beispielsweise nach Schrader und Helmke (2001) objektive Informationen über Merkmale der Schülerinnen und Schüler (z.B. Leistung), wie sie standardisierte diagnostische Verfahren und Tests bieten, eine wertvolle Hilfe für das Lehrpersonenurteil darstellen. Dies setzt allerdings voraus, dass Lehrkräfte Kenntnisse über entsprechende Verfahren erlangen und diese gegebenenfalls auch anwenden können. Ein weiteres (in unseren Arbeiten jedoch nicht spezifisch untersuchtes) Merkmal einer soliden Wissensgrundlage diagnostischer Urteilsbildung ist das Wissen der Lehrkräfte über mögliche Urteilsfehler und -tendenzen, die die Genauigkeit ihrer Urteile beeinflussen (Helmke, 2009; Wilbers, 2012). Einige bekannte Urteilsfehler sind beispielsweise der Haloeffekt oder der logische Fehler. Beim Haloeffekt wird aufgrund weniger (nicht valider) Indikatoren (z.B. Aussehen, Sprache) auf globale Merkmale der Schülerinnen und Schüler geschlossen, während beim logischen Fehler von der Ausprägung eines bestimmten Merkmals auf ein anderes Merkmal der Schülerin bzw. des Schülers geschlossen wird, ohne dass dies empirisch begründet wäre (z.B. Rechtschreibprobleme als Indiz für geringe Intelligenz, vgl. Helmke, 2009). In der Aus- und Weiterbildung von Lehrkräften sollten daher zusätzlich auch Urteilsfehler und -tendenzen thematisiert werden.

Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57 (3), 175–193.
- Artelt, C. (2009). Diagnostische Urteile von Lehrkräften im Bereich der Lesekompetenz. In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität: Bildungsaufgabe und Forschungsfeld* (S. 125–136). Weinheim: Juventa.
- Artelt, C. & Gräsel, C. (2009). Gasteditorial: Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 157–160.
- 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* (S. 215–234). Münster: Waxmann.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T. et al. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9 (4), 521–544.
- Demaray, M.K. & Elliott, S.N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13 (1), 8–24.

- Feinberg, A.B. & Shapiro, E.S.** (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102 (6), 453–462.
- Helmke, A.** (2009). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern*. Seelze: Kallmeyer. 173 Genauigkeit von Lehrpersonenurteilen
- Helmke, A., Hosenfeld, I. & Schrader, F.-W.** (2003). Diagnostikkompetenz in Ausbildung und Beruf entwickeln. *Karlsruher pädagogische Beiträge*, Heft 55, 15–34.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W.** (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnostikkompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Baltmannsweiler: Schneider Verlag Hohengehren.
- Hoge, R.D. & Coladarsi, T.** (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, 59 (3), 297–313.
- Ingenkamp, K.** (1992). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Kaiser, J., Helm, F., Retelsdorf, J., Südkamp, A. & Möller, J.** (2012). Zum Zusammenhang von Intelligenz und Urteils-genauigkeit bei der Beurteilung von Schülerleistungen im Simulierten Klassenraum. *Zeitschrift für Pädagogische Psychologie*, 26 (4), 251–261.
- Karing, C.** (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 197–209.
- Karing, C., Matthäi, J. & Artelt, C.** (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25 (3), 159–172.
- Karing, C., Pfof, M. & Artelt, C.** (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for Educational Research Online*, 3 (2), 119–147.
- Lehmann, R.H., Peek, R., Gänsfuß, R., Lutkat, S., Mücke, S. & Barth, I.** (2000). *Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik (QuaSUM)*. Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (MBJS).
- Schrader, F.-W.** (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt am Main: Peter Lang.
- Schrader, F.-W.** (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 237–245.
- Schrader, F.-W.** (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl. age) (S. 102–108). Weinheim: Beltz.
- Schrader, F.-W.** (2011). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch zur Forschung zum Lehrerberuf* (S. 683–698). Münster: Waxmann.
- Schrader, F.-W. & Helmke, A.** (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1 (1), 27–52.
- Schrader, F.-W. & Helmke, A.** (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.
- Spinath, B.** (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1–2), 85–95.
- Ständiges Sekretariat der Kultusminister der Länder der Bundesrepublik Deutschland.** (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16.12.2004. Online verfügbar unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf (28.08.2013).
- Südkamp, A., Kaiser, J. & Möller, J.** (2012). Accuracy of teachers' judgments of students' academic achievement: A Meta-Analysis. *Journal of Educational Psychology*, 104 (3), 743–762.
- Wilbers, K.** (2012). *Wirtschaftsunterricht gestalten. Toolbox: Eine traditionelle und handlungsorientierte Didaktik für kaufmännische Bildungsgänge*. Berlin: epubli GmbH.

Autorinnen

Constance Karing, Dr., Friedrich-Schiller-Universität Jena, Institut für Psychologie, Abteilung für Forschungssynthese, Intervention und Evaluation, constance.karing@uni-jena.de.

Cordula Artelt, Prof. Dr., Otto-Friedrich-Universität Bamberg, Fakultät Humanwissenschaften, Lehrstuhl für Empirische Bildungsforschung, cordula.artelt@uni-bamberg.de.

erschienen in: *Beiträge zur Lehrerbildung*, 31 (2), 2013, S. 166 – 173
www.bzl-online.ch/archivdownload/artikel/BZL_2013_2_166-173.pdf