

Zeitschrift zu Theorie und Praxis der Aus- und
Weiterbildung von Lehrerinnen und Lehrern

BEITRÄGE ZUR LEHRERBILDUNG

Pädagogische Diagnostik

31. Jahrgang Heft 2/2013

Beiträge zur Lehrerbildung. Zeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern

Organ der Schweizerischen Gesellschaft für Lehrerinnen- und Lehrerbildung (SGL)

Erscheint dreimal jährlich

Herausgeber und Redaktion

Christian Brühwiler, Pädagogische Hochschule St. Gallen, Institut Professionsforschung und Kompetenzentwicklung, Notkerstrasse 27, 9000 St. Gallen, Tel. 071 243 94 86, christian.bruehwiler@phsg.ch

Anni Heitzmann, Fachhochschule Nordwestschweiz, Pädagogische Hochschule, Institut Sekundarstufe I und II, Clarastrasse 57, 4058 Basel, Tel. 076 415 16 65, anni.heitzmann@fhnw.ch

Alois Niggli, Pädagogische Hochschule Freiburg, Dienststelle Forschung, Murtengasse 36, 1700 Freiburg, Tel. 026 305 72 55, NiggliA@edufr.ch

Christine Pauli, Universität Freiburg, Departement Erziehungswissenschaften, Rue Faucigny 2, 1700 Freiburg, Tel. 026 300 75 64, christine.pauli@unifr.ch

Kurt Reusser, Universität Zürich, Institut für Erziehungswissenschaft, Freiestrasse 36, 8032 Zürich, Tel. 044 634 27 68 (27 53), reusser@ife.uzh.ch

Annette Tettenborn, Pädagogische Hochschule Luzern, Institut für pädagogische Professionalität und Schulkultur (IPS), Töpferstrasse 10, 6004 Luzern, Tel. 041 228 70 17, annette.tettenborn@phlu.ch

Peter Tremp, Pädagogische Hochschule Zürich, Abteilung Forschung und Entwicklung, Lagerstrasse 2, 8090 Zürich, Tel. 043 305 62 30, peter.tremp@phzh.ch

Manuskripte

Manuskripte können bei einem Mitglied der Redaktion eingereicht werden. Richtlinien für die Gestaltung von Beiträgen sind auf www.bzl-online.ch verfügbar (siehe «Autoren/Autorinnen» «Manuskriptgestaltung»). Diese Richtlinien sind verbindlich und müssen beim Verfassen von Manuskripten unbedingt eingehalten werden.

Lektorat

Jonna Truniger, bzl-lektorat@bluewin.ch, www.textuell.ch

Externe Mitarbeiterinnen und Mitarbeiter

Buchbesprechungen

Jürg Rüedi, Fachhochschule Nordwestschweiz, Pädagogische Hochschule, Institut Primarstufe, Benzburweg 30, 4410 Liestal, Tel. 061 925 91 55, juerg.ruedi@fhnw.ch. Für nicht eingeforderte Rezensionsexemplare übernimmt die Redaktion keinerlei Verpflichtung.

Neuerscheinungen und Zeitschriftenspiegel

Sandro Biaggi, Pädagogische Hochschule Luzern, Frohburgstrasse 3, 6002 Luzern, Tel. 041 228 33 65, sandro.biaggi@phlu.ch

Forschung an pädagogischen Hochschulen – Kurzberichte

Stefan Denzler, Schweizerische Koordinationsstelle für Bildungsforschung (SKBF), Entfelderstrasse 61, 5000 Aarau, Tel. 062 835 23 97, www.skbf-csre.ch, stefan.denzler@skbf-csre.ch

Cartoons

Ueli Halbheer, Pädagogische Hochschule Thurgau, Unterer Schulweg 3, 8280 Kreuzlingen 2, Tel. 071 678 56 93, ulrich.halbheer@phtg.ch

Editorial

- Christian Brühwiler, Anni Heitzmann Alois Niggli, Christine Pauli,
Kurt Reusser, Annette Tettenborn, Peter Tremp 151

Schwerpunkt**Pädagogische Diagnostik**

- Friedrich-Wilhelm Schrader** Diagnostische Kompetenz von Lehrpersonen 154
- Constance Karing und Cordula Artelt** Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften 166
- Anna-Katharina Praetorius** Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter – Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis 174
- Alois Buholzer und Sandra Zulliger** Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung 186
- Guido McCombie und Damian Läge** Der Educational Profiler: Ein Instrument für die Diagnose der pädagogischen Bezugssysteme angehender Lehrpersonen 198
- Andreas Helmke und Gerlinde Lenske** Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung 214

Forum

- Carola Mantel und Bruno Leutwyler** Lehrpersonen mit Migrationshintergrund: Eine kritische Synthese der Literatur 234

Rubriken

- Forschung an pädagogischen Hochschulen – Kurzberichte** 248

Buchbesprechungen

- Alois Buholzer und Martina Brülisauer** Kommentierte Literaturliste zur pädagogischen Diagnostik 255

Drollinger-Vetter, B. (2011). Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht (Esther Brunner)	266
Engi, P. (2012). Lehrplan 21 – Instrument in einem interkantonalen Bildungsorchester (Hans Berner)	268
Gluck, M.A., Mercado, E. & Myers, C.E. (2010). Lernen und Gedächtnis. Vom Gehirn zum Verhalten (Christoph Schmid)	270
Kreis, A. (2012). Produktive Unterrichtsbesprechungen. Lernen im Dialog zwischen Mentoren und angehenden Lehrpersonen (Michael Fuchs)	272
Kuckartz, U. & Rädiker, S. (Hrsg.). (2012). Erziehungswissenschaftliche Evaluationspraxis: Beispiele – Konzepte – Methoden (Wolfgang Beywl)	274
Neuerscheinungen	276
Zeitschriftenspiegel	278

Vorschau auf künftige Schwerpunktthemen

Eine Vorschau auf die Schwerpunktthemen künftiger Hefte finden Sie auf unserer Homepage (<http://www.bzl-online.ch>). Manuskripte zu diesen Themen können bei einem Mitglied der Redaktion eingereicht werden (vgl. dazu die Richtlinien zur Manuskriptgestaltung, verfügbar auf der Homepage).

Editorial

Der Themenschwerpunkt «Pädagogische Diagnostik» ist facettenreich. Im Unterschied zu einer medizinisch-psychologischen Semantik des Begriffs zielt das vorliegende Heft auf ein Verständnis von Diagnostik, das dem pädagogischen Handeln dient. Die Ziele der pädagogischen Diagnostik können dabei höchst unterschiedlich sein und auf verschiedenen Ebenen liegen. Ist eine Beurteilung von Personen (Schülerinnen und Schülern oder Lehrpersonen) das Ziel oder dient die Diagnose der Unterstützung des Lernprozesses? Werden mit der Diagnose allgemeine Unterrichtsmerkmale charakterisiert oder spezifisch die Lernpotenziale, die Klippen und Schwierigkeiten von Aufgaben betrachtet? Wird die Diagnose zur summativen oder (förderorientierten) formativen Beurteilung und für Feedback verwendet?

Je nach Zielebene sind andere diagnostische Kompetenzen gefragt, was zu mannigfachen Spannungsfeldern führt, beispielsweise semiformelle Einschätzungen vs. Urteilsgenauigkeit, aufgabenorientierte vs. personenbezogene Diagnose, wissenschaftliches Interesse vs. Praxisrelevanz, Statusdiagnostik vs. Förderdiagnostik, psychologische vs. pädagogische Diagnostik. So dürften für die Unterrichtspraxis semiformelle bzw. beiläufige Diagnosehandlungen, die den Lernprozess der Schülerinnen und Schüler wenig stören und mit vergleichsweise wenig Zeitaufwand durchzuführen sind, eine grössere Bedeutung besitzen als die Urteilsgenauigkeit allein. Letztere ist hingegen grundlegend, wenn es um Selektionsentscheide oder um die Untersuchung wissenschaftlicher Fragestellungen geht. Im Unterschied zur psychologischen Diagnostik ist die pädagogische Diagnostik immer auf Lehr-Lern-Prozesse bezogen. Im Wesentlichen wird darunter die Fähigkeit einer Lehrperson verstanden, lern- und leistungsrelevante Merkmale zutreffend zu beurteilen sowie Aufgabenanforderungen und Strukturierungsmassnahmen akkurat einzuschätzen. Eine gute Diagnose informiert beispielsweise Lehrpersonen darüber, was die Schülerinnen und Schüler bereits verstehen und in welcher Hinsicht noch Defizite bestehen oder wo mögliche Verbesserungsmassnahmen für den Unterricht liegen könnten.

Dass der pädagogischen Diagnostik in den letzten Jahren viel Aufmerksamkeit zuteilwurde, hat nicht zuletzt mit den Tendenzen zur Individualisierung des Unterrichts, aber auch mit der Professionalisierung der Lehrerinnen- und Lehrerbildung zu tun. Soll der allgemein formulierte Grundsatz «ganzheitlich beurteilen und fördern» auf der Basis eines individualisierenden Unterrichts konkret und professionell umgesetzt werden, so sind dazu diagnostische Kompetenzen notwendig. In der Medizin beruht jede Abschätzung eines Erkrankungsrisikos oder des Verlaufs einer Krankheit auf der Grundlage von Daten früherer Patientinnen und Patienten; es sind somit statistische und probabilistische Aussagen, welche eine Diagnose stützen. Diese erhält ihren Wert aber nur, wenn sie vor dem entsprechenden Hintergrund und auf das anvisierte Ziel hin interpretiert wird. Das Gleiche gilt für die pädagogische Diagnostik: Lehrpersonen müssen hinterfragen, wann welche Diagnose sinnvoll ist, welcher Zweck damit verbunden ist und

Forschungsinstrument zur standardisierten Erfassung von Überzeugungen in vergleichenden Studien.

Der Schwerpunktteil wird abgerundet durch den Überblicksbeitrag von **Andreas Helmke und Gerlinde Lenske** zur Unterrichtsdiagnostik. Dadurch sollen Lernen und Lehren im Sinne einer empirischen Standortsbestimmung sichtbar gemacht werden. Sie stellen das Diagnoseinstrument EMU (Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung) vor, das sich an den verschiedenen Gegenstandsbe-
reichen der pädagogischen Diagnostik orientiert und mit welchem ein- und derselbe Unterrichtsgegenstand aus unterschiedlichen Perspektiven betrachtet werden kann. Mit ihrem Beitrag zielen sie auf den Kern jeder pädagogischen Diagnostik, nämlich auf die Fähigkeit, unterscheiden zu können. Es ist der differenzierte Umgang mit Diagnoseinstrumenten und Diagnoseresultaten, der letzten Endes die Qualität einer Diagnose ausmacht.

Im Forum ist der Beitrag von **Carola Mantel und Bruno Leutwyler** zu finden. Sie geben einen Überblick über den Forschungsstand zu Lehrpersonen mit Migrationshintergrund. Dabei stellen sie fest, dass erst wenig empirische Evidenz vorliege, wenn nach deren Potenzial im Umgang mit migrationsbedingter Heterogenität gefragt wird. Sie kritisieren unreflektierte Erwartungen an Lehrpersonen mit Migrationshintergrund im Sinne von «Kulturvermittlerinnen» und «Kulturvermittlern» und fordern stattdessen, dass das Potenzial im Hinblick auf die Entwicklung interkultureller Kompetenz ins Zentrum gestellt werden soll.

Am Schluss des Hefts finden sich wie üblich die verschiedenen Rubriken. Speziell hinweisen möchten wir auf die Sammelrezension von **Alois Buholzer und Martina Brülisauer** zum Schwerpunktthema «Pädagogische Diagnostik», die den anderen Buchbesprechungen vorangestellt ist.

Christian Brühwiler, Anni Heitzmann, Alois Niggli, Christine Pauli, Kurt Reusser, Annette Tettenborn, Peter Tremp

Diagnostische Kompetenz von Lehrpersonen*

Friedrich-Wilhelm Schrader

Zusammenfassung Der Beitrag gibt einen Überblick über Diagnoseleistungen und diagnostische Kompetenzen in der Lehrtätigkeit. Zu Beginn werden Grundlagen und basale Aufgabenstellungen der pädagogischen Diagnostik behandelt. Im Vordergrund steht dabei die Bedeutung der pädagogischen Diagnostik für den Lehr-Lern-Prozess. Einen weiteren Schwerpunkt bilden der Ansatz der Urteilsgenauigkeit sowie prozessorientierte Ansätze, zu denen jeweils ein knapper Überblick über ausgewählte Forschungsergebnisse gegeben wird. Am Ende wird auf den Stellenwert diagnostischer Kompetenz, Forschungsdefizite sowie Konsequenzen für die Verbesserung von Diagnoseleistungen eingegangen.

Schlagwörter diagnostische Kompetenz – Diagnosegenauigkeit – formelle und informelle Diagnosen – Urteilsprozesse – Unterricht

Teacher Diagnosis and Diagnostic Competence

Abstract The following article provides an overview of teacher diagnosis and diagnostic competence. Starting with a presentation of basic approaches and diagnostic tasks, the main focus lies on the role of educational diagnosis in instruction and with respect to student learning. Special emphasis is put on two major approaches: teachers' judgement accuracy and judgement processes. A brief review of selected results of research in these domains is given. Finally, the importance of diagnostic competence, limitations of research, and consequences regarding the improvement of teacher diagnosis are discussed.

Keywords diagnostic competence – judgement accuracy – formal and informal diagnosis – judgement processes – instruction

1 Einleitung

Diagnostische Kompetenz von Lehrpersonen bezieht sich auf die Fähigkeit, die im Lehrberuf anfallenden diagnostischen Aufgabenstellungen erfolgreich zu bewältigen, und auf die Qualität der dabei erbrachten Diagnoseleistungen (Schrader, 2011). Ziel ist es, Informationen über Lernergebnisse, Lernvoraussetzungen und Lernvorgänge von Schülerinnen und Schülern zu gewinnen, die für verschiedene pädagogische Entscheidungen (Notengebung, Versetzung, Übergangsempfehlungen, Unterrichtsplanung und -gestaltung, Schul- und Unterrichtsentwicklung) genutzt werden können. Diagno-

* Ich danke Andreas Helmke und zwei unbekanntem Gutachtern für hilfreiche Kommentare zu einer ersten Fassung dieses Beitrags.

auf welcher Datenbasis die Diagnose beruht. Je nach Antwort fällt das pädagogische Handeln dann anders aus.

Das vorliegende Heft will verschiedene Spannungsfelder im Bereich der pädagogischen Diagnostik aufnehmen und auf differenzierte Betrachtungsweisen aufmerksam machen. Der Beitrag von **Friedrich-Wilhelm Schrader** führt ins Thema ein und umreißt zunächst die Grundlagen des pädagogischen Diagnostizierens von Lehrpersonen. Anschliessend thematisiert er die Diagnosegenauigkeit und stellt Überlegungen zum Zusammenhang von Diagnosegenauigkeit, Unterrichtsqualität und Lernerfolg dar. Daraus folgert er, dass ein Verständnis von diagnostischer Kompetenz nötig sei, das sich nicht auf die blosse Diagnosegenauigkeit beschränkt.

Constance Karing und Cordula Artelt nehmen den Aspekt der Urteilsgenauigkeit auf und vertiefen ihn im Kontext der Urteilsspezifität. Sie weisen darauf hin, dass die Urteilsgenauigkeit auch vom jeweiligen Lerngegenstand abhängt, und zeigen am Vergleich von globalen und aufgabenspezifischen Lehrpersonenurteilen, dass für akkurate Urteile nicht nur diagnostische, sondern auch didaktische und fachwissenschaftliche Kompetenzen notwendig sind. Die Autorinnen sehen eine Verbesserung der Unterrichtsqualität insbesondere bei den aufgabenspezifischen Urteilen und beim Umgang mit Urteilsfehlern.

Anna-Katharina Praetorius weist auf das Spannungsfeld unterschiedlicher Beurteilungsskalen hin. Sie hebt hervor, dass Beurteilungen von Unterrichtsqualität häufig unter Verwendung hochinferenter Ratinginstrumente erfolgen. Diese erfordern Schlussfolgerungen über das sichtbare Unterrichtsgeschehen hinaus und sind deshalb oft mit Urteilsfehlern behaftet. Sie diskutiert diese Problematik am Beispiel externer Beobachtereinschätzungen, wie sie etwa in den Praktika der Lehrerinnen- und Lehrerausbildung vorkommen, und nennt Möglichkeiten zur Optimierung solcher Diagnosen.

Alois Buholzer und Sandra Zulliger untersuchen in einer Längsschnittstudie am Beispiel des Fachbereichs Deutsch, ob sich die diagnostischen Fähigkeiten angehender Lehrpersonen im Verlaufe der Ausbildung verändern. Sie verweisen dabei auch darauf, dass Diagnosekompetenz keine einheitliche Grundkompetenz darstellt und daher nicht generalisierbar ist, sondern unterschiedliche Qualitäten von diagnostischen Fähigkeiten zu unterscheiden sind.

Der Beitrag von **Guido McCombie und Damian Läge** thematisiert ebenfalls Entwicklungsmöglichkeiten von Lehramtsstudierenden. Sie stellen mit dem «Educational Profiler» ein Instrument vor, mit dem subjektive, implizit vorhandene Überzeugungen von Studierenden erfasst und offengelegt sowie Entwicklungen sichtbar gemacht werden können. Die Autoren diskutieren den Einsatz des Instruments einerseits als Entwicklungs- und Reflexionsinstrument in der Ausbildung und andererseits als mögliches

sen betreffen den Ist-Zustand und stehen in Bezug zu bestimmten Zielvorstellungen (Soll-Zustände), die sich aus Lehrzielen, curricularen Vorgaben und Bildungsstandards ableiten. Schwerpunkt der Lehrpersonendiagnostik sind Leistungen der Schülerinnen und Schüler, die in der schulischen Praxis das zentrale Zielkriterium des Unterrichts darstellen und gleichzeitig grundlegende Lernvoraussetzungen sind. Während früher Fragen der Beurteilung und Bewertung von Lernergebnissen im Vordergrund standen, richtet sich der Blick gegenwärtig stärker auf die Nutzung der Diagnostik für die Unterrichtsgestaltung, die Steuerung des Lehr-Lern-Prozesses und die Unterrichtsentwicklung. Diagnostische Kompetenz wird häufig mit der Fähigkeit gleichgesetzt, genaue diagnostische Urteile abzugeben (Schrader, 2011). Daneben rücken aber Fragen, die den Urteilsprozess und die Urteilsbildung betreffen, zunehmend in den Blickpunkt wissenschaftlichen Interesses (Artelt & Gräsel, 2009).

2 Grundlagen des pädagogischen Diagnostizierens von Lehrpersonen

Während in der pädagogischen Diagnostik wissenschaftlich fundierte Methoden des Diagnostizierens (formelle Diagnostik) im Vordergrund stehen, geht es bei der diagnostischen Kompetenz von Lehrpersonen stärker um die im Schul- und Unterrichtsalltag vorherrschenden Urteile. Diese informellen (bzw. semiformalen, vgl. Hascher, 2008) Diagnoseleistungen sind für das unterrichtliche Handeln zwar unverzichtbar, wegen ihrer Fehleranfälligkeit jedoch problematisch, sodass oft zu fragen ist, inwieweit sie durch formelle Diagnosen ergänzt oder korrigiert werden können.

Die *abschliessende Bewertung von Lernprozessen* ist Grundlage für Versetzungs- und Übertrittsentscheidungen und für die Vergabe von Berechtigungen (Abschlusszeugnisse), die oft für Selektionszwecke (Ausbildungs-, Studien- oder Arbeitsplätze) herangezogen werden und langfristige und gravierende Folgen für die Einzelne und den Einzelnen wie auch für die Gesellschaft haben (Ingenkamp & Lissmann, 2008; van Ophuysen, 2010). Deshalb sind Forderungen nach einer pädagogischen Diagnostik, die hohen Gütemassstäben (Objektivität, Reliabilität, Validität) gerecht wird, und nach einer fundierten methodischen Absicherung der Urteile gerade hier von grosser Bedeutung. Ein weiterer Schwerpunkt der Diagnostik sind *unterrichtsbezogene Diagnoseleistungen*, mit denen der Lehr-Lern-Prozess überwacht und optimiert wird (Ingenkamp & Lissmann, 2008; Nitko, 2004). Grundlegend für den Lern- und Unterrichtserfolg ist eine hinreichende Anpassung des Unterrichts an die Lernvoraussetzungen der Schülerinnen und Schüler (Adaptivität), die sowohl durch längerfristige, auf grössere Unterrichtseinheiten bezogene Massnahmen (*Makroadaptationen*) als auch durch kurzfristige Abstimmungsmassnahmen während des Unterrichts (*Mikroadaptationen*) erfolgen kann (vgl. z.B. Schrader & Helmke, 2001). Während Mikroadaptationen meist nur auf der Grundlage von diagnostischen Informationen (Beobachtungen, subjektive Einschätzungen), die während des Unterrichtens anfallen, erfolgen können, besteht bei

Makroadaptationen oftmals die Möglichkeit, gezieltere diagnostische Abklärungen (formelle Diagnostik) vorzunehmen.

Abklärung der Lernvoraussetzungen

Eine systematische Abklärung der Lernausgangslage ist vor allem bei der Übernahme einer neuen Klasse oder nach dem Wechsel in einen weiterführenden Bildungsgang sinnvoll und kann wichtige Hinweise für die Planung und Gestaltung des Unterrichts liefern (Hesse & Latzko, 2009). Schwerpunkt sind bereichsspezifische Vorkenntnisse, die sich als die besten Prädiktoren des späteren Lernerfolgs erwiesen haben (Helmke & Weinert, 1997). Vor allem in hierarchisch aufgebauten Fächern wie Mathematik handelt es sich um kumulative Lernergebnisse, bei denen einmal entstandene Lernlücken das nachfolgende Lernen (und die Lernmotivation) grundlegend beeinträchtigen können. Diagnostische Hilfsmittel sind Strukturanalysen des Vorwissens, die auf einer detaillierten, die herkömmlichen Leistungsüberprüfungen (Klassenarbeiten) ergänzenden Dokumentation von Lernverläufen basieren (Hesse & Latzko, 2009).

Überwachung des Lernfortschritts

Traditionelle sozialnorm- oder bezugsgruppenorientierte Schulleistungstests bieten die Möglichkeit, den Lernstand im Vergleich zur gesamten Altersgruppe festzustellen. Da sie meistens an Groblernzielen orientiert und nicht auf den Unterricht in einer bestimmten Klasse bezogen sind, geben sie oft nur wenige Hinweise für die Gestaltung des weiteren Lernprozesses. Kriteriumsorientierte Tests dagegen sind an Feinlernzielen orientiert und erlauben es, das Erreichen der Lehrziele an einer sachlichen Bezugsnorm zu messen (Ingenkamp & Lissmann, 2008). Für die eigentliche Lernsteuerung, bei der eine hohe Messgenauigkeit nicht im Vordergrund steht, kommen darüber hinaus von den Lehrpersonen entwickelte informelle Verfahren, offene Angaben sowie Interview- und Beobachtungsformate infrage (Shepard, 2001).

Abklärung von Lernschwierigkeiten

In einigen Bereichen – beispielhaft ist hier insbesondere die Lesekompetenz zu nennen – bieten grundlagenwissenschaftlich fundierte Testverfahren die Möglichkeit, Teilkomponenten von fachlichen Fähigkeiten zu erfassen (Hasselhorn, Schneider & Marx, 2000). Für eine breitere Ursachenanalyse, die auch den nicht kognitiven Bereich (insbesondere Lernmotivation und Lernverhalten) einbezieht, sind mittlerweile wissenschaftliche Ansätze und Messverfahren verfügbar (vgl. z.B. Hesse & Latzko, 2009), die sehr aussichtsreich erscheinen, auf die Lehrpersonen in ihrer Ausbildung aber bislang noch kaum vorbereitet werden.

Unterrichtsentwicklung und Professionalisierung

Für die Bestandsaufnahme bieten schul- und klassenübergreifende Kompetenzmessungen (Vergleichsarbeiten, andere Lernstandserhebungen) besondere Vorteile, da sie anders als viele herkömmliche Schulleistungstests eng an Kernkompetenzen bzw. Bildungsstandards orientiert sind und oft innovative Aufgabenstellungen umfassen, die

für den weiteren Unterricht wichtige Impulse liefern können (Altrichter, 2010; Helmke & Hosenfeld, 2004). Das Verständnis und die Fähigkeit zum Umgang mit solchen Informationen bilden einen zunehmend wichtigen Aspekt der diagnostischen Kompetenz. Lohnenswert erscheint es, über reine Outputdaten hinaus auch Kontext- und Prozessdaten einzubeziehen (etwa durch die Nutzung von Schüler- und kollegialem Feedback; vgl. Helmke et al., 2012).

3 Empirische Forschungen zur diagnostischen Kompetenz

3.1 Diagnostische Kompetenz als Diagnosegenauigkeit

Die seit Langem bekannten Mängel der Notengebung (unzureichende Objektivität, Reliabilität, Validität; vgl. z.B. Ingenkamp & Lissmann, 2008) werden oft als Beleg für die mangelnde diagnostische Kompetenz von Lehrpersonen angeführt. Der Ansatz der Urteilsgenauigkeit beschäftigt sich systematisch mit der Frage, wie gut Lehrpersonenurteile mit den bei den Schülerinnen und Schülern gemessenen «tatsächlichen» Merkmalsausprägungen übereinstimmen. Sofern Lehrpersonenurteile und Merkmale der Schülerinnen und Schüler im gleichen Antwortformat erfasst werden (z.B. Einschätzung der von jeder Schülerin und von jedem Schüler in einem Test erreichten Punktzahl), sind alle Abweichungen zwischen beiden als Fehler zu betrachten. Bei einfachen Abweichungsmassen werden aber verschiedene Komponenten der Urteilsgenauigkeit vermischt (Cronbach, 1955), sodass üblicherweise Komponentenanalysen durchgeführt werden. In der deutschsprachigen Forschung geschieht dies allerdings fast ausschliesslich in einer vereinfachten Form, bei der eine Urteilerin oder ein Urteiler (z.B. eine Lehrperson) mehrere Personen (z.B. die Schülerinnen und Schüler der eigenen Klasse) hinsichtlich *eines* einzigen Merkmals (statt mehrerer wie bei Cronbach, 1955) beurteilt (Karst, 2012). Dabei können eine (jeweils klassenweise berechnete) *Niveauekomponente* (Über- oder Unterschätzung des mittleren Ausprägungsniveaus), eine *Differenzierungs- oder Streuungskomponente* (Über- oder Unterschätzung der Merkmalsstreuung) und eine *Rangordnungskomponente* (Korrelation zwischen Urteilen und tatsächlichen Ausprägungen) unterschieden werden (Schrader & Helmke, 1987). Die Rangordnungskomponente wird häufig als das eigentliche Mass für die diagnostische Kompetenz (diagnostische Sensitivität) angesehen, da sie von allgemeinen Urteilstendenzen (wie Milde- oder Strengetendenz) frei ist und sich auch berechnen lässt, wenn Urteile und tatsächliche Merkmalsausprägungen nicht im gleichen Antwortformat erfasst werden. Gelegentlich werden die Leistungen der Schülerinnen und Schüler auch auf einzelne Aufgaben bezogen beurteilt (gelöst/nicht gelöst) und Treffermasse berechnet (Coladarci, 1986; Karing, Matthäi & Artelt, 2011). Am häufigsten untersucht ist die Beurteilung von Leistungen der Schülerinnen und Schüler, die nicht nur eine zentrale Rolle für den Unterricht spielen, sondern sich mithilfe von Tests auch genauer und weniger verzerrungsanfällig messen lassen als nicht kognitive Merkmale der Schülerinnen und Schüler (z.B. Interesse: Karing, 2009; Fähigkeitsselbstkonzept: Praetorius, Greb, Dickhäuser & Lipowsky, 2011).

3.2 Deskriptive Ergebnisse zur Diagnosegenauigkeit

Aus Platzgründen kann hier nur ein knapper Überblick über einige zentrale Befunde zur personenbezogenen diagnostischen Kompetenz gegeben werden (zur Einschätzung von Aufgabenschwierigkeiten vgl. McElvany et al., 2009). So zeigen verschiedene Studien, dass Lehrpersonen im Durchschnitt gesehen die Rangordnung der Leistungen recht genau einschätzen (Hoge & Coladarsi, 1989). In einer auf der Basis von 75 Studien durchgeführten mehrbenenanalytisch ausgewerteten Meta-Analyse von Südkamp, Kaiser und Möller (2012) wird eine mittlere Korrelation von .63 bei einer Spannweite von -.03 bis .84 berichtet. Signifikante Moderatoren sind lediglich Testkenntnis und Übereinstimmung im Spezifitätsgrad von Urteilen und Leistungen; andere Faktoren, darunter das Fachgebiet (Sprachleistungen versus Mathematik), spielen hingegen keine Rolle. Zwischen einzelnen Lehrpersonen bzw. Klassen gibt es deutliche Unterschiede in der Höhe der Korrelation, z.B. von .02 bis .88 bei Helmke und Schrader (1987) oder von .11 bis .86 bei Hosenfeld, Helmke und Schrader (2002), und damit in der Genauigkeit der Urteile von einzelnen Lehrpersonen (zur Reliabilität und damit verbundenen Problemen vgl. Lintorf et al., 2011; Lorenz, 2011). Die Stabilität der Rangordnungskomponente bei Leistungseinschätzungen wurde von Lorenz und Artelt (2009) untersucht. Die über den Zeitraum eines halben Jahres berechneten Korrelationen (.38 für Mathematik, .51 für Textverstehen, .57 für Wortschatz; $N = 127$ Klassen) fallen ähnlich hoch aus wie die Urteilsgüte zu den einzelnen Messzeitpunkten.

Kaum etwas ist bekannt über Bedingungsfaktoren und Korrelate der Diagnosegenauigkeit wie Berufs- bzw. Lehrerfahrung oder kognitive und nicht kognitive Persönlichkeitsmerkmale (zusammenfassend z.B. Lorenz, 2011). Einer der wenigen Befunde ist das von Karing (2009) berichtete Ergebnis, dass Grundschullehrpersonen Leistungen und Fachinteresse der Schülerinnen und Schüler genauer einschätzten als Gymnasiallehrpersonen. Dagegen hängen Unterschiede in der diagnostischen Kompetenz vom Kontext ab. Dies gilt insbesondere für die Merkmalshomogenität bzw. -heterogenität, die ein urteilerleichterndes bzw. urteilerschwerendes Kontextmerkmal für die Rangordnungskomponente darstellt, allerdings auch aus rein statistischen Gründen mit der Höhe der Korrelation zusammenhängt. Masse der Diagnosegenauigkeit sind daher keine reinen Indikatoren für Personenmerkmale, sondern eher der Urteiler-Beurteilten-Relation bzw. der Beurteiler-Umwelt-Relation (Schrader, 2009).

Ergebnisse zum Urteilsniveau (Niveauelemente) werden sehr viel seltener berichtet. Insgesamt zeigt sich hier überwiegend eine Tendenz zur Überschätzung des Leistungsniveaus (zusammenfassend van Ophuysen, 2010). So berichten etwa Bates und Nettelbeck (2001) eine Tendenz zur Überschätzung der Leseleistung von 6- bis 8-jährigen Kindern, insbesondere bei leseschwachen Schülerinnen und Schülern. Dieses Befundmuster deckt sich mit Ergebnissen der Expertiseforschung, denen zufolge Expertinnen und Experten Schwierigkeiten damit bekunden, die Leistungen von Laiinnen und Laien zutreffend einzuschätzen. In einer Studie von Herppich, Wittwer, Nückles und Renkl (2013) überschätzten Tutorinnen und Tutoren, die jeweils mit einer einzigen

oder einem einzigen Lernenden zusammenarbeiteten, deren Verständnisseleistungen. Allerdings urteilten erfahrene Lehrpersonen entgegen dem allgemeinen Befund aus der Expertiseforschung genauer und konnten auch ihre eigene Diagnosefähigkeit besser einschätzen als studentische Tutorinnen und Tutoren. Bei der nur selten untersuchten Differenzierungskomponente liess sich keine einheitliche Tendenz zur Über- oder Unterschätzung feststellen (vgl. z.B. van Ophuysen, 2010). Empirische Untersuchungen (z.B. Spinath, 2005) zeigen ferner, dass die Komponenten der Diagnosegenauigkeit unabhängig sind, was die Frage aufwirft, ob bzw. in welchem Sinne man überhaupt von einem allgemeinen Konstrukt der diagnostischen Kompetenz sprechen kann.

Der Ansatz des simulierten Klassenraums, bei dem die Leistungen hypothetischer Schülerinnen und Schüler beurteilt werden, bietet die Möglichkeit, diagnostische Kompetenz unter kontrollierten Bedingungen zu untersuchen (vgl. Südkamp & Möller, 2009; Südkamp, Möller & Pohlmann, 2008). In diesem computergesteuerten System wählen die Urteilernden aus einem Menü vorgegebene Fragen aus, zu denen die von der Urteilerin oder vom Urteiler jeweils ausgewählten «Schülerinnen» und «Schüler» (gesteuert durch einen vorher eingestellten Fähigkeitsparameter und eine Zufallskomponente) eine richtige oder falsche Antwort geben, über deren Korrektheit die Probandin oder der Proband eine Rückmeldung erhält. Die nach Ende einer Unterrichtseinheit für jede Schülerin und jeden Schüler abgegebene Leistungseinschätzung (Zahl gelöster Aufgaben, Note) wird mit der tatsächlichen Leistung (Zahl der richtigen Antworten) verglichen. Bemerkenswert ist, dass Untersuchungen mit dem simulierten Klassenraum zu ähnlichen Ergebnissen führen wie im realen Klassenkontext (Rangordnungskomponente in der Höhe von .53 bis .68; leichte Tendenz zur Überschätzung des Leistungsniveaus bei einer leichten Unterschätzung der Streuung; vgl. Südkamp, Möller & Pohlmann, 2008).

3.3 Diagnosekompetenz, Unterrichtsqualität und Lernerfolg

Ob die Fähigkeit von Lehrpersonen, die Leistungen ihrer Schülerinnen und Schüler zutreffend zu beurteilen, auch für den Unterrichtserfolg eine Rolle spielt, ist bislang erst selten untersucht worden. Helmke und Schrader (1987) konnten keinen direkten Zusammenhang zwischen Diagnosegenauigkeit und Lernerfolg der Schülerinnen und Schüler im Fach Mathematik nachweisen. Hohe Diagnosekompetenz hatte aber dann einen Effekt, wenn Strukturierungshilfen und individuelle Hilfestellungen ausreichend häufig eingesetzt wurden (Katalysatorfunktion der diagnostischen Kompetenz; vgl. Helmke, 2012). In zwei neueren Untersuchungen konnten dagegen direkte Zusammenhänge mit dem Lernerfolg der Schülerinnen und Schüler aufgezeigt werden. In der Studie von Anders, Kunter, Brunner, Krauss und Baumert (2010) liess sich der Lernerfolg der Schülerinnen und Schüler in Mathematik durch die Genauigkeit der Schwierigkeitseinschätzung und die diagnostische Sensitivität vorhersagen, aber nur die Genauigkeit der Schwierigkeitseinschätzung hing auch mit der Unterrichtsqualität zusammen. Karing, Pfof und Artelt (2011) konnten für ein aufgabenspezifisches Treffermass einen Zusammenhang mit dem Lernerfolg im Lesen aufzeigen, nicht aber

für die Rangordnungskomponente. Das Treffermass sagte – erwartungsgemäss – den Lernerfolg im Lesen bei hoher, nicht aber bei niedriger Individualisierung sowie – erwartungskonträr – bei niedriger, nicht aber bei hoher Strukturierung vorher, während es bei der Vorhersage der Mathematikleistung für beide Diagnosemasse weder einfache noch differenzielle Effekte gab.

Das von Beck et al. (2008) entwickelte Konzept der adaptiven Lehrkompetenz, bei dem diagnostische Kompetenzen sowohl in der Planungs- wie auch in der Durchführungsphase des Unterrichts eng mit unterrichtsbezogenen Fähigkeiten (Klassenführung, Didaktik, Sachkompetenz) verknüpft sind, war Gegenstand vertiefender Analysen von Brühwiler (2012). Die adaptive Lehrkompetenz wurde in der Planungsphase mittels eines Vignetten-Ansatzes (Beschreibungen hypothetischer Situationen, zu denen die Probandinnen und Probanden Planungsüberlegungen äussern sollten) und in der Handlungsphase mittels eines Videotests (ein nach einem Drehbuch aufgezeichnetes Unterrichtsvideo, zu dem die Probandinnen und Probanden ihre Handlungen beschreiben und begründen sollten) erfasst. Mit der adaptiven Lehrkompetenz liess sich der Lernerfolg der Schülerinnen und Schüler vorhersagen. Gleichzeitig korrelierte sie in der Höhe von .39 mit der auf die Lernzielerreichung der Schülerinnen und Schüler bezogenen diagnostischen Urteilsfähigkeit (Rangordnungskomponente), was für die enge Verzahnung von diagnostischen und didaktischen Schritten spricht.

3.4 Diagnostische Kompetenz: Prozessorientierte Ansätze

Verschiedene Studien geben Aufschluss über die Prozesse, die diagnostischen Urteilen zugrunde liegen. Ausgehend von Befunden der Expertiseforschung (schnellere Verarbeitung, flexiblere Organisation und Nutzung von Informationen, geringere Anfälligkeit für Urteilsverzerrungen) untersuchte van Ophuysen (2006), ob sich Expertinnen und Experten (erfahrene Grundschullehrpersonen) bei der Abgabe einer Übergangsempfehlung für einen hypothetischen Schüler von Novizinnen und Novizen (Lehramtsstudierende) unterscheiden. Gezeigt werden konnte, dass Novizinnen und Novizen einer stärkeren Bestätigungstendenz (Beibehaltung des Urteils trotz inkonsistenter Informationen) unterlagen, während Expertinnen und Experten nicht nur differenzierter urteilten, sondern ihr Urteil bei inkonsistenten Zusatzinformationen auch häufiger revidierten. Untersuchungen von Krolak-Schwerdt und weiteren Kolleginnen und Kollegen (z.B. Krolak-Schwerdt & Rummer, 2005; Krolak-Schwerdt, Böhmer & Gräsel, 2009) stützen sich auf Ansätze der sozialen Urteilsforschung, bei denen zwischen einer *kategorienbasierten* (Aktivierung von Kategorien/Stereotypen) und einer *merkmalbasierten* Informationsverarbeitung (Abprüfung von Einzelmerkmalen) unterschieden wird. Krolak-Schwerdt und Rummer (2005) konnten zeigen, dass Expertinnen und Experten (Lehrpersonen) bei einer Prognoseaufgabe eine merkmalsbasierte Verarbeitung vornahmen, bei der Aufforderung, sich einen Eindruck zu bilden, dagegen kategorienbasiert vorgehen, während sich bei Laiinnen und Laien die Vorgehensweisen in beiden Fällen nicht unterschieden. Auch in einer Untersuchung von Krolak-Schwerdt, Böhmer und Gräsel (2009), in der erfahrene Lehrpersonen (Expertinnen und Exper-

ten) und Lehramtsstudierende (Novizinnen und Novizen) anhand der Fallbeschreibung eines hypothetischen Schülers entweder dessen schulische Entwicklung vorhersagen (merkmalbasierte Verarbeitung) oder sich einen Eindruck über den Schüler bilden sollten (kategorienbasierte Verarbeitung), variierten die erfahrenen Lehrpersonen ihre Urteilsbildung je nach Verarbeitungsziel und erwiesen sich damit als flexible Denkerinnen und Denker, während die Informationsverarbeitung der Laiinnen und Laien durch das Verarbeitungsziel nicht beeinflusst wurde. Krolak-Schwerdt, Böhmer und Gräsel (2013) konnten darüber hinaus zeigen, dass die Verantwortlichkeit für das Urteil (Accountability) ein Moderator für die Urteilsbildung ist. Dünnebier, Gräsel und Krolak-Schwerdt (2009) untersuchten mittels der Ankerheuristik, ob die beiden Modi der Informationsverarbeitung unterschiedlich anfällig gegenüber vereinfachenden, aber verzerrungsanfälligen heuristischen Strategien sind. Sie konnten für das Fach Deutsch, nicht aber für Mathematik aufzeigen, dass der Anker bei Expertinnen und Experten (erfahrene Lehrpersonen) unter der Bedingung einer kategorienbasierten Verarbeitung (Eindrucksbildung) einen Effekt hatte, nicht aber bei einer merkmaleitenden Verarbeitung (Schullaufbahnpfehlung). Bei den Novizinnen und Novizen (Lehramtsstudierende) zeigte sich unter beiden Bedingungen ein Ankereffekt.

4 Fazit und Ausblick

Lehrpersonendiagnostik bewegt sich in einem Spannungsfeld zwischen formeller, d.h. auf wissenschaftlich fundierten Methoden beruhender Diagnostik und informeller Diagnostik. Informelle Lehrpersonenurteile sind unverzichtbar, wenn es um die Handlungssteuerung im Unterricht geht. Eine formelle Diagnostik ist immer dann angebracht, wenn es um Entscheidungen mit schwerwiegenden und weitreichenden Konsequenzen geht. Im Unterschied dazu können lern- und unterrichtsbezogene Entscheidungen nicht in jedem Fall durch explizite Diagnosen abgesichert werden, schon gar nicht, wenn es um die Feinabstimmung des Unterrichts (Mikroadaptation) geht.

Die Urteilsgenauigkeit wird in der Forschung als Indikator für die diagnostische Kompetenz angesehen. Untersuchungen haben gezeigt, dass die Urteilsgenauigkeit direkt oder indirekt mit dem Lernerfolg der Schülerinnen und Schüler zusammenhängt. Das Problem dieser Studien ist allerdings, dass es über die Prozesse, die zwischen den diagnostischen Urteilen der Lehrpersonen und den Lernergebnissen aufseiten der Schülerinnen und Schüler vermitteln, kaum empirisch gesichertes Wissen gibt. Dafür, dass es sich bei der diagnostischen Kompetenz um ein für die Lehrtätigkeit und insbesondere für das unterrichtliche Handeln bedeutsames Merkmal handelt, das sich auch im Gesamtgefüge der Lehrpersonenmerkmale gut verorten lässt (Baumert & Kunter, 2006), sprechen gute theoretische Gründe. Auch die Forschungsübersichten von Hattie (2009, 2012) weisen darauf hin, dass regelmässiges Feedback, eine engmaschige Überwachung des Lernerfolgs und die fortlaufende Anpassung des Unterrichtsangebots durch adaptive Lernexpertinnen und Lernexperten ganz wesentlich zum Lernerfolg der Schü-

lerinnen und Schüler beitragen. Hattie sieht darin das eigentliche Erfolgsgeheimnis für einen effektiven, leistungsförderlichen Unterricht. Die mit dem Ansatz der Urteils-genauigkeit verbundenen Probleme (ungeklärte Reliabilität, fehlende Zusammenhänge mit Lehrpersonenmerkmalen, Abhängigkeit vom Urteilskontext, Unabhängigkeit verschiedener Komponenten der Urteils-genauigkeit, unklare Wirkmechanismen; vgl. Lorenz, 2011; Schrader, 2009; Spinath, 2005) werfen allerdings Fragen auf, die im Zuge weiterer Forschung zu klären sind.

Die prozessorientierten Forschungen geben Hinweise auf Faktoren, die die Genauigkeit des Urteils beeinflussen können. Sie tragen zu einem vertieften Verständnis der Unterschiede zwischen formellen und informellen Beurteilungen und der Unterschiede in der Urteilsbildung von Expertinnen und Experten einerseits und von Novizinnen und Novizen andererseits bei. Da es sich dabei um (quasi)experimentelle Studien handelt, erlauben sie auch bessere Rückschlüsse auf die kausale Interpretierbarkeit. Um die ökologische Validität zu verbessern und fundierte Rückschlüsse auf die schulische Realsituation zu ermöglichen, wäre es sinnvoll, diese grundlagenorientierten Studien stärker mit Felduntersuchungen zu verknüpfen, insbesondere auch mit solchen zur Urteils-genauigkeit (Böhmer, Gräsel, Hörstermann & Krolak-Schwerdt, 2012; Schrader, 2009).

Vor diesem Hintergrund erscheint es sinnvoll, von einem breiteren Verständnis diagnostischer Kompetenz auszugehen als von der blossen Diagnosegenauigkeit. Dass Lehrpersonenurteile genau oder zumindest frei von groben Verfälschungen und Verzerrungen sein sollten, ist zwar mehr als naheliegend. Ebenso wichtig dürfte aber auch die Bereitschaft sein, die eigenen Urteile zu reflektieren, zu revidieren und flexibel an Veränderungen anzupassen. Zu diagnostischer und pädagogischer Professionalität gehört demnach auch die Bereitschaft, sich der eigenen, die Diagnose möglicherweise beeinträchtigenden und verzerrenden Erwartungen, Urteilstendenzen und impliziten Theorien bewusst zu werden. Eine erhöhte Sensibilität für diagnostische Fragen, eine zunehmende Einsicht in die Grenzen der eigenen Urteilsfähigkeit und die zunehmende Bereitschaft, die eigenen Urteilsleistungen durch den Einsatz von Diagnoseverfahren und den Einbezug anderer Perspektiven (z.B. Selbsteinschätzungen der Schülerinnen und Schüler oder Urteile der Kolleginnen und Kollegen) regelmässig auf den Prüfstand zu stellen und zu verbessern, könnten mindestens ebenso wichtig sein wie die Urteils-genauigkeit als solche. Zu einer professionellen Lehrpersonendiagnostik gehören zudem die Fähigkeit, geeignete Strategien und Methoden der Informationsbeschaffung und -verarbeitung einzusetzen, was auch die passgenaue Auswahl und Konstruktion von Tests und anderen Verfahren zur Lernkontrolle einschliesst, sowie die Bereitschaft, sich mit den gewonnenen Daten fundiert auseinanderzusetzen und diese sachgerecht zu interpretieren (Nitko, 2004).

Trotz der bislang noch bestehenden Forschungsdefizite werden die Förderung und die Verbesserung diagnostischer Kompetenz als wichtiges Anliegen betrachtet. Ein un-

längst im Auftrag der deutschen Kultusministerkonferenz durchgeführtes Projekt setzt dabei auf eine verstärkte Aus- und Fortbildung von Lehrpersonen in vier Teilbereichen der pädagogischen Diagnostik: Individualdiagnostik, Diagnostik im Rahmen von Vergleichsarbeiten, Diagnostik im Rahmen des Bildungsmonitorings sowie Unterrichtsdiagnostik (vgl. Bos & Hovenga, 2010). Die Wissensvermittlung erfolgt dabei durch Studienbriefe und begleitende Aufgabenstellungen und Materialien. Im Bereich der Unterrichtsdiagnostik wurden ausserdem Tools entwickelt, die nicht nur darauf abzielen, Wissen zu vermitteln, sondern auch darauf, die Urteilspraxis zu verbessern (vgl. Helmke et al., 2012) und auf diese Weise viele der oben als wünschenswert genannten Kompetenzmerkmale zu erreichen. Inwieweit dieser Ansatz zum Aufbau eines fundierten diagnostischen und metakognitiven Wissens und zur Verbesserung der diagnostischen Kompetenz beiträgt, bleibt abzuwarten. Die bislang vorliegenden Erfahrungen stimmen positiv, bedürfen aber der gründlichen Abklärung und Evaluation durch wissenschaftliche Folgeuntersuchungen.

Literatur

- Altrichter, H.** (2010). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 219–254). Wiesbaden: VS Verlag für Sozialwissenschaften.
- 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. & Gräsel, C.** (2009). Gasteditorial: Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 157–160.
- Bates, C. & Nettelbeck, T.** (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21 (2), 177–187.
- Baumert, J. & Kunter, M.** (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P. et al.** (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- Böhmer, I., Gräsel, C., Hörstermann, T. & Krolak-Schwerdt, S.** (2012). Die Informationssuche bei der Erstellung der Übergangsempfehlung – Die Rolle von Fallkonsistenz und Expertise. *Unterrichtswissenschaft*, 40 (2), 140–155.
- Bos, W. & Hovenga, N.** (2010). Diagnostische Kompetenzen – besser individuell fördern. *Schule NRW*, Heft 8, 383–385.
- Brühwiler, C.** (2012). *Adaptive Lehrkompetenz und schulisches Lernen. Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler*. Unveröffentlichte Dissertation. Landau: Universität Landau, Fachbereich Psychologie.
- Coladarsi, T.** (1986). The accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78 (2), 141–146.
- Cronbach, L.J.** (1955). Processes affecting scores on «understanding of others» and «assumed similarity». *Psychological Bulletin*, 52 (3), 177–193.
- Dünnebier, K., Gräsel, C. & Krolak-Schwerdt, S.** (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 187–195.

- Hascher, T.** (2008). Diagnostische Kompetenzen von Lehrpersonen. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung* (S. 71–86). Münster: Waxmann.
- Hasselhorn, M., Schneider, W. & Marx, H.** (Hrsg.). (2000). *Diagnostik von Lese-Rechtschreibschwierigkeiten* (Jahrbuch der pädagogisch-psychologischen Diagnostik. Tests und Trends – Neue Folge, Band 1). Göttingen: Hogrefe.
- Hattie, J.** (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J.** (2012). *Visible learning for teachers. Maximizing impact on learning*. London: Routledge.
- Helmke, A.** (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. Auflage). Seelze: Klett-Kallmeyer.
- Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. & Ade-Thuraw, M.** (2012). Unterrichtsdiagnostik – Voraussetzung für die Verbesserung der Unterrichtsqualität. In S.G. Huber (Hrsg.), *Jahrbuch Schulleitung 2012. Befunde und Impulse zu den Handlungsfeldern des Schulmanagements* (S. 133–144). Köln: Carl Link.
- Helmke, A. & Hosenfeld, I.** (2004). Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven. In R.S. Jäger, A. Frey & M. Wosnitza (Hrsg.), *Lernprozesse, Lernumgebungen und Lerndiagnostik* (S. 56–75). Landau: Verlag Empirische Pädagogik.
- Helmke, A. & Schrader, F.-W.** (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3 (2), 91–98.
- Helmke, A. & Weinert, F.E.** (1997). Bedingungsfaktoren schulischer Leistungen. In F.E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Pädagogische Psychologie. Psychologie des Unterrichts und der Schule* (S. 71–176). Göttingen: Hogrefe.
- Herppich, S., Wittwer, J., Nückles, M. & Renkl, A.** (2013). Does it make a difference? Investigating the assessment accuracy of teacher tutors and student tutors. *The Journal of Experimental Education*, 81 (2), 242–260.
- Hesse, I. & Latzko, B.** (2009). *Diagnostik für Lehrkräfte*. Opladen: Barbara Budrich.
- Hoge, R.D. & Coladarci, T.** (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59 (3), 297–313.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W.** (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Lehr-Lern-Studie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (45. Beiheft der Zeitschrift für Pädagogik) (S. 65–82). Weinheim: Beltz.
- Ingenkamp, K. & Lissmann, U.** (2008). *Lehrbuch der Pädagogischen Diagnostik* (6. Auflage). Weinheim: Beltz.
- Karing, C.** (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungs- 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., Pfost, 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.
- Karst, K.** (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Münster: Waxmann.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C.** (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. Der Lehrer als «flexibler» Denker. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 175–186.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C.** (2013). The impact of accountability on teachers' assessments of student performance: a social cognitive analysis. *Social Psychology of Education*, 16 (2), 215–239.

- Krolak-Schwerdt, S. & Rummer, R.** (2005). Der Einfluss von Expertise auf den Prozess der schulischen Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37 (4), 205–213.
- Lintorf, K., McElvany, N., Rjosk, C., Schroeder, S., Baumert, J., Schnotz, W. et al.** (2011). Zuverlässigkeit von diagnostischen Lehrerurteilen – Reliabilität verschiedener Urteilsmaße bei der Einschätzung von Aufgabenschwierigkeiten. *Unterrichtswissenschaft*, 39 (2), 102–120.
- Lorenz, C.** (2011). *Diagnostische Kompetenz von Grundschullehrkräften. Strukturelle Aspekte und Bedingungen*. Bamberg: University of Bamberg Press.
- Lorenz, C. & Artelt, C.** (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 211–222.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W. et al.** (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 25 (3), 223–235.
- Nitko, A.J.** (2004). *Educational assessment of students* (4th edition). Upper Saddle River, NJ: Pearson.
- Praetorius, A.-K., Greb, K., Dickhäuser, O. & Lipowsky, F.** (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 58 (2), 81–91.
- Schrader, F.-W.** (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 237–245.
- Schrader, F.-W.** (2011). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der 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.
- Shepard, L.A.** (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4th edition) (S. 1066–1101). Washington, DC: American Educational Research Association.
- Spinath, B.** (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer/innen und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1–2), 85–95.
- 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.
- Südkamp, A. & Möller, J.** (2009). Referenzgruppeneffekte im Simulierten Klassenraum: direkte und indirekte Einschätzungen von Schülerleistungen. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 161–174.
- Südkamp, A., Möller, J. & Pohlmann, B.** (2008). Der Simulierte Klassenraum. Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22 (3–4), 261–276.
- van Ophuysen, S.** (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnenempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38 (4), 154–161.
- van Ophuysen, S.** (2010). Professionelle pädagogisch-diagnostische Kompetenz – eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H.G. Holtappels, N. McElvany & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Band 16: Daten, Beispiele und Perspektiven* (S. 203–234). Weinheim: Juventa.

Autor

Friedrich-Wilhelm Schrader, Dr., Universität Koblenz-Landau, Campus Landau, Fachbereich Psychologie, schrader@uni-landau.de

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

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

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

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

* 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).

von mündlichen und schriftlichen Leistungsüberprüfungen, bei Schullaufbahneempfehlungen 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 Urteilskomponenten, 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 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 Urteilskomponenten auf das Lehrpersonenurteil zum Niveau der Leistungen der Schülerinnen und Schüler bzw. zur Leistungsstreuung. So gibt die Niveauelemente darüber Auskunft, inwieweit Lehrkräfte die Ausprägungen der Leistungen der Schülerinnen und Schüler akkurat einschätzen können, während die Differenzierungskomponente 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 Rangordnungskomponente, 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 & Coladarcı, 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 & Coladarcı, 1989). Der Grad der Urteilsspezifität 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 (Rangordnungskomponente, 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 For-

schungsprojekts 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 Rangordnungskomponenten 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).

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

	Lesen		Mathematik	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
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.

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 Lehrpersoneneinschätzungen unterschiedliche Wissensselemente benötigen. Während globale Urteile auf das Wissen über Personen (Fähigkeit der Schülerinnen 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 Wissensselemente 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 Urteilsgenauigkeit 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 Instrukstionsstrategien 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 Auswirkungen 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 Urteilsgenauigkeit 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 Urteilsgenauigkeit 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 Leistungsrangfolge 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 Urteilsgenauigkeit 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 Urteilsgenauigkeit wurde dabei neben der Rangord-

nungskomponente 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 Urteilsgenauigkeit 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 die-

sem 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. Bertsch-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.

- Helmke, A., Hosenfeld, I. & Schrader, F.-W.** (2003). Diagnosekompetenz 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 Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Baltmannsweiler: Schneider Verlag Hohengehren.
- Hoge, R.D. & Coladarci, 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 Urteilsgenauigkeit 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 Leistungsbe-
reich 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. Auflage) (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

Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter – Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis

Anna-Katharina Praetorius

Zusammenfassung Die Einschätzung von Unterricht durch externe Beobachterinnen und Beobachter ist seit jeher konstitutiver Bestandteil des Lehrberufs. In welchem Ausmass solche Beobachtereinschätzungen zuverlässige Aussagen über den Unterricht von Lehrkräften darstellen, wurde bislang jedoch kaum empirisch überprüft. Der vorliegende Beitrag gibt einen Überblick über Studien, die infrage stellen, inwiefern es sich bei Beobachtereinschätzungen um den «Königsweg» zur Erfassung von Unterrichtsqualität handelt. Der Beitrag schliesst mit Folgerungen für die Unterrichtspraxis. Dabei wird unter anderem die Notwendigkeit des Einsatzes mehrerer Beobachterinnen und Beobachter sowie einer Beobachtung von längeren Zeiträumen diskutiert.

Schlagwörter Unterrichtsqualität – Messung – Beobachterinnen und Beobachter – Ratings

Measuring Instructional Quality with Observer Ratings – A Critical Examination of Current Practice in Schools

Abstract Assessments of teaching by external observers have always been a constitutive part of the teaching profession. Still, the extent to which these ratings allow reliable statements about the observed teachers' instruction has rarely been empirically investigated. Against this background, the following article provides an overview of studies which challenge the assumption that observer ratings are the ideal method of measuring instructional quality. It concludes with some implications for instructional practice. Among other things, the discussion touches on the need to include several observers and to cover longer periods of observation.

Keywords instructional quality – measurement – observer – ratings

1 Einleitung

Unterrichtsbeurteilungen und -rückmeldungen sind seit jeher konstitutiver Bestandteil des Lehrberufs: Bereits in den Schulpraktika im Studium erhalten Lehramtsstudierende Rückmeldungen zu Stärken und Schwächen der von ihnen gehaltenen Unterrichtsstunden. Im Referendariat kommt zu diesen Rückmeldungen auch die (benotete) Bewertung von Unterrichtsstunden hinzu. Doch auch im Anschluss an die Ausbildung sind Lehrkräfte immer wieder mit Unterrichtseinschätzungen externer Beobachterinnen und Beobachter konfrontiert, u.a. von Schulleiterinnen und Schulleitern sowie Schulinspek-

torinnen und Schulinspektoren, aber auch bei gegenseitigen Hospitationen durch Kolleginnen und Kollegen. Diese Unterrichtsbesuche können zum einen zu Forschungszwecken erfolgen und zum anderen im Kontext der Lehrpersonenaus- und -fortbildung zu Zwecken der summativen oder formativen Evaluation eingesetzt werden. Bei summativen Einschätzungen steht die Bewertung (und Benotung) von Unterricht im Vordergrund. Dies spielt in der aktuellen Lehrpersonenausbildung eine grosse Rolle: Die Benotung von Unterrichtsstunden dient u.a. dazu, die am besten geeigneten Lehrkräfte für eine begrenzte Stellenanzahl auszuwählen. Bei formativen Einschätzungen hingegen steht die Rückmeldung zu Unterricht im Vordergrund. Solche Rückmeldungen zu den unterrichtlichen Stärken und Schwächen von Lehrkräften sollen als Basis für die Weiterentwicklung von Unterricht dienen – dies wird auch als evidenzbasierte Unterrichtsentwicklung bezeichnet (vgl. z.B. Altrichter, Messner & Posch, 2004; Helmke, 2009; Horster & Rolff, 2001). Warum eine solche evidenzbasierte Entwicklung von Unterricht wichtig ist, ist unmittelbar einsichtig: Nur wenn Lehrpersonen über den Ist-Stand informiert sind, können sie ihren Unterricht passgenau weiterentwickeln (vgl. dazu auch Helmke & Lenske, 2013 [in diesem Heft]).

Sowohl zu Forschungszwecken als auch in der Lehrpersonenaus- und -fortbildung werden zur Erfassung von Unterrichtsqualität oftmals hochinferente Ratingskalen verwendet. Bei diesen Skalen handelt es sich um eine Zusammenstellung von Items, die Schlussfolgerungen über das direkt Sichtbare hinaus erfordern. Ein Beispiel für ein solches hochinferentes Item lautet: «Mit Schülerinnen- und Schülerbeiträgen ist die Lehrkraft wertschätzend umgegangen.» Ein solcher wertschätzender Umgang kann nicht direkt aus den Verhaltensweisen der Lehrkraft abgelesen werden, sondern muss anhand diverser Indikatoren erschlossen werden (z.B. «Die Lehrkraft würdigt unvollständige/halb richtige Schülerinnen- und Schülerbeiträge, indem sie die richtigen Aspekte aufgreift»). Zur Beurteilung der Items werden dann vorgegebene Antwortmöglichkeiten (z.B. trifft nicht zu – trifft eher nicht zu – trifft eher zu – trifft zu) verwendet. Grund für den bevorzugten Einsatz solcher hochinferenter Items ist, dass diese Art von Items eine höhere Erklärungskraft für diverse Aussenkriterien (z.B. für die Leistungsentwicklung von Schülerinnen und Schülern) hat als niedriginferente Items, solche Items also, deren Beantwortung keine oder kaum Schlussfolgerungen über das Sichtbare hinaus erfordert (z.B. die Anzahl an Meldungen von Schülerinnen und Schülern) (Clausen, 2002; Sommer, 2011). Ein Nachteil von hochinferenten Items ist die Tatsache, dass bei deren Beantwortung Beurteilerfehler eine deutlich grössere Rolle spielen als bei niedriginferenten Items (vgl. auch Clausen, 2002).

In welchem Ausmass externe Beobachterinnen und Beobachter auf der Basis hochinferenter Ratinginstrumente zuverlässige Aussagen über den Unterricht von Lehrkräften treffen, wurde bislang kaum thematisiert (Hill et al., 2012; Pietsch & Tosana, 2008; Praetorius, in Druck). Dies ist erstaunlich, da die Vergabe von Stellen basierend auf Unterrichtsbeurteilungen nur dann sinnvoll ist, wenn diese Beurteilungen auch zuverlässig sind. Und auch evidenzbasierte Unterrichtsentwicklung ist lediglich in dem Ausmass

möglich, in dem die Informationen über den Ist-Stand den Ist-Stand tatsächlich widerspiegeln.

Im Folgenden wird zunächst auf die in der Literatur genannten Vorteile von externen Beobachtereinschätzungen eingegangen (Abschnitt 2). Anschliessend sollen Befunde der empirischen Unterrichtsforschung dargestellt werden, die zum Teil kritisch beleuchten, ob die Zuverlässigkeit der Unterrichtseinschätzungen externer Beobachterinnen und Beobachter tatsächlich gegeben ist (Abschnitt 3). Der Beitrag schliesst mit Schlussfolgerungen aus den berichteten Ergebnissen (Abschnitt 4).

2 Vorteile externer Beobachtereinschätzungen von Unterricht

Externe Beobachtereinschätzungen von Unterricht werden oft als «Königsweg» zur Erfassung von Unterrichtsqualität beschrieben (Helmke, 2009, S. 288; vgl. auch Clare, Valdés, Pascal & Steinberg, 2001; Pianta & Hamre, 2009; Petko, Waldis, Pauli & Reusser, 2003). Dafür gibt es diverse Gründe:

- Erstens durchlaufen Beobachterinnen und Beobachter – zumindest in der Unterrichtsforschung und im Kontext der Schulinspektion – in der Regel ein Training für die Einschätzung von Unterricht. Diese Schulung sollte es ihnen erlauben, die hohe Komplexität von Unterricht angemessen zu erfassen (Helmke, 2009; Petko et al., 2003).
- Zweitens wird argumentiert, dass Beobachterinnen und Beobachter – im Gegensatz zu Lehrkräften sowie Schülerinnen und Schülern – nicht Teil des Unterrichtsgeschehens sind und dadurch zum einen eher objektive Unterrichtseinschätzungen abgeben können und zum anderen ihre kognitiven Ressourcen vollständig für die Wahrnehmung des Unterrichts zur Verfügung haben (Rakoczy, 2008; Waldis et al., 2010).
- Drittens wird darauf verwiesen, dass Beobachterinnen und Beobachter in der Regel etliche Lehrkräfte einschätzen und dadurch in einem hohen Masse Vergleichsmöglichkeiten haben, während Lehrkräfte selbst vor allem ihren eigenen Unterricht kennen (Clausen, 2002; Rakoczy, 2008).

3 Nachteile externer Beobachtereinschätzungen von Unterricht

Aufgrund der in Abschnitt 2 genannten Vorteile werden externe Beobachtereinschätzungen oft als vergleichsweise «objektive» Möglichkeit zur Erfassung von Unterrichtsqualität beschrieben und als gegenüber anderen Methoden (vor allem Lehrkraftselbsteinschätzungen und Einschätzungen von Schülerinnen und Schülern) zu bevorzugen dargestellt (z.B. Clare et al., 2001; Helmke, 2009; Rakoczy, 2008). Studien zeigen jedoch, dass auch Beobachtungen von unbeteiligten, externen Personen nicht einfach «die Realität» und damit die tatsächliche Unterrichtsqualität abbilden (vgl. auch Clau-

sen, 2002; Hill, Charalambous & Kraft, 2012; Waldis et al., 2010). Dies liegt zum einen daran, dass auch externe Beobachterinnen und Beobachter diversen Beurteilerfehlern unterliegen (Abschnitt 3.1), zum anderen aber auch daran, dass externe Beobachtungen Restriktionen in Bezug auf die zur Verfügung stehenden Informationen unterliegen (Abschnitt 3.2).

3.1 Beurteilereffekte bei externen Beobachtereinschätzungen

Durchsucht man die Literatur in Bezug auf die Frage, wie zuverlässig bzw. genau die Einschätzungen externer Beobachterinnen und Beobachter sind, so finden sich erstaunlich wenige Studien. Die vorhandenen Untersuchungen (Clausen et al., 2003; Hill et al., 2012; Kobarg & Seidel, 2005; Matsumura, Garnier, Pascal & Valdés, 2002; Newton, 2010; Pietsch & Tosana, 2008; Praetorius, in Druck; Strong et al., 2011) zeigen, dass zwischen 0 und 41% der Variation in Ratings – und damit teilweise ein sehr hoher Anteil – auf die Beobachterinnen und Beobachter anstelle auf das zu messende Merkmal zurückzuführen sind. In der Untersuchung von Clausen et al. (2003) weist beispielsweise das Merkmal «Disziplinprobleme» einen sehr geringen Raterhaupteffekt (d.h. Unterschiede zwischen Raterinnen und Ratern in ihrer Strenge bzw. Milde) von 3% auf, das Merkmal «Klarheit» mit 41% hingegen einen sehr hohen entsprechenden Effekt.

Diesem zum Teil hohen Anteil an Ratereffekten können unterschiedlichste Ursachen zugrunde liegen. Wegen der begrenzten menschlichen Wahrnehmung unterscheiden sich Personen bereits darin, was sie beobachten (vgl. auch Lemons & Helsing, 2010). Aber auch in der Verarbeitung des Beobachteten treten deutliche Unterschiede auf (vgl. z.B. Bradburn, 2004): So werden Items aufgrund der Uneindeutigkeit von Sprache uneinheitlich interpretiert. Es werden unterschiedliche Informationen aus dem Gedächtnis abgerufen, diese Informationen dann verschieden gewichtet und kombiniert, und auch die Wahl einer geeigneten Antwortkategorie kann unterschiedlich ausfallen (vgl. auch Praetorius, in Druck). Besonders deutlich treten solche Unterschiede auf, wenn Beobachterinnen und Beobachter kein gemeinsames Training durchlaufen haben. Dies wird beispielsweise in einer Untersuchung von Praetorius, Lenske und Helmke (2010) deutlich, in der sich ungeschulte Lehramtsstudierende eine Unterrichtsstunde ansahen und anschliessend in einem Einzelinterview dazu befragt wurden. In Bezug auf ein ausgewähltes Item finden sich in Abbildung 1 die Aussagen derjenigen vier Studierenden, die bei diesem Item «stimme eher zu» gewählt haben. Wie aus den Begründungen deutlich wird, interpretieren die Studierenden den im Item enthaltenen Begriff «Diskussion» sehr unterschiedlich, beginnend bei einer Gleichsetzung mit dem Stillsein der Schülerinnen und Schüler, über deren Meldeverhalten bis hin zum Äussern der eigenen Meinung.

Um solche Beobachtereffekte so weit wie möglich zu minimieren, werden externe Beobachterinnen und Beobachter in der Unterrichtsforschung sowie im Kontext der Schulinspektion trainiert. Inwiefern diese Trainings tatsächlich dazu führen, dass die Einschätzungen in einem höheren Ausmass zuverlässig sind (= Reliabilität) und verstärkt

Item: «Die Schülerinnen und Schüler haben sich an der Diskussion im Unterricht beteiligt.»

Antwortkategorie bei allen Antworten: «stimme eher zu»

- «Die meisten schon oder waren zumindest leise und haben zugehört.»
- «Ja, nicht unbedingt jeder, aber doch sehr viele. Es gingen glaube ich die Finger nach oben, also ich stimme eher zu.»
- «Es haben sich eigentlich relativ viele Schüler dran beteiligt.»
- «Ja klar, die haben sich immer gemeldet. Diskussion, wurde es dann zum Schluss zur Diskussion – war es eine Diskussion oder war es einfach nur eine Wiedergabe? Also zum Schluss sollten sie ihre Meinung äussern, das war eine Diskussion, also stimme ich eher zu.»

Abbildung 1: Unterschiedliche Iteminterpretationen von ungeschulten Raterinnen und Ratern (Praetorius et al., 2010).

das messen, was sie messen sollen (= Validität), ist bislang kaum untersucht worden. Eine Untersuchung von Praetorius et al. (2012) weist darauf hin, dass die Validität der Einschätzungen durch entsprechende Trainings erhöht werden kann: Trainierte Raterinnen und Rater nehmen bei ihren Einschätzungen in einem stärkeren Ausmass Bezug auf Indikatoren, die laut Ratingmanual dem zu messenden Konstrukt entsprechen, und weisen zudem differenziertere Auseinandersetzungen mit den jeweiligen Items auf als untrainierte Raterinnen und Rater. In derselben Studie zeigt sich allerdings auch, dass die Reliabilität (in dieser Untersuchung gemessen über den relativen Generalisierbarkeitskoeffizienten im Rahmen der Generalisierbarkeitstheorie, vgl. Brennan, 2001) bei geschulten Raterinnen und Ratern entgegen den theoretischen Annahmen nicht höher ausfiel als bei ungeschulten Raterinnen und Ratern – und zum Teil sogar geringer ausgeprägt war als bei ungeschulten Lehramtsstudierenden (vgl. Abbildung 2). Ähnliche Ergebnisse zeigen sich auch in den wenigen weiteren Studien zu diesem Thema (z.B. Strong et al., 2011). Die Ergebnisse der Untersuchung von Praetorius et al. (2012) wei-

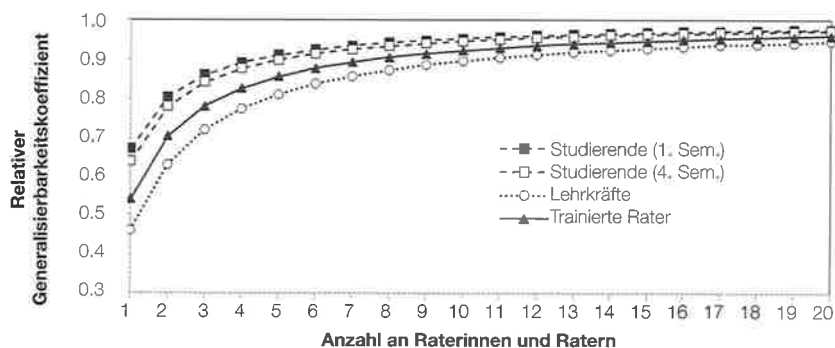


Abbildung 2: Entscheidungsstudien zur Unterrichtsdimension «Schülerorientierung» (Praetorius, in Druck, S. 125).

sen darauf hin, dass beispielsweise bei der Erfassung der Unterrichtsqualitätsdimension «Schülerorientierung» für eine Messgenauigkeit von .70 zwei Lehramtsstudierende des ersten Semesters, drei Lehramtsstudierende des vierten Semesters, drei Lehrkräfte oder zwei geschulte Raterinnen und Rater notwendig sind. Mit einer einzelnen Beobachterin oder einem einzelnen Beobachter – geschult oder ungeschult – lässt sich demnach keine hinreichende Messgenauigkeit erreichen. Möchte man individualdiagnostische Einschätzungen treffen, sind sogar Reliabilitätskoeffizienten von .90 oder höher erforderlich. Um einen solchen Koeffizienten zu erreichen, müssen deutlich mehr Raterinnen und Rater eingesetzt werden (z.B. elf ungeschulte Lehrkräfte oder acht trainierte Raterinnen und Rater).

3.2 Die eingeschränkte Informationsgrundlage externer Beobachterratings

Neben dem Auftreten von Beobachtereffekten werden Beobachterratings auch dahingehend kritisiert, dass die den Beobachterinnen und Beobachtern zur Verfügung stehende Informationsgrundlage sehr beschränkt ist. So beobachten sie Unterricht in der Regel nur im Rahmen eines sehr kurzen Zeitraums (Abschnitt 3.2.1), Informationen über den Klassenkontext und die einzelnen Schülerinnen und Schüler stehen ihnen nur begrenzt oder gar nicht zur Verfügung (Abschnitt 3.2.2) und Lehrkräfte sowie Schülerinnen und Schüler verhalten sich unter Umständen unter Beobachtung anders als im alltäglichen Unterricht (Abschnitt 3.2.3).

3.2.1 Kurzer Beobachtungszeitraum

Lehrkräfte sowie Schülerinnen und Schüler können Aussagen über Unterricht bezogen auf einen langen Zeitraum treffen. Dies ist für Beobachterinnen und Beobachter in der Regel nicht möglich, da in der Unterrichtsforschung meist ein bis zwei Unterrichtsstunden pro Lehrkraft eingeschätzt werden (Übersicht in Praetorius, in Druck). Bei kollegialen Hospitationen, Besuchen durch Schulleitungsmitglieder oder bei Unterrichtsversuchen im Rahmen des Referendariats sind es üblicherweise ebenfalls maximal ein bis zwei Stunden pro Person, die beobachtet werden. Schulinspektorinnen und Schulinspektoren wie auch Schulleitungsmitglieder besuchen zum Teil sogar nur 20 bis 30 Minuten einer Unterrichtsstunde (vgl. z.B. Hill et al., 2012; Pietsch & Tosana, 2008). Einige Autorinnen und Autoren stellen daher infrage, ob auf der Basis eines solch kurzen Zeitraums Aussagen über die Qualität des Unterrichts einer Lehrkraft im Allgemeinen überhaupt möglich sind (Berliner, 2005; Brophy, 2006; Stigler et al., 1999). Tatsächlich zeigen empirische Untersuchungen (z.B. Hill et al., 2012; Newton, 2010; Praetorius, in Druck), dass bei manchen Unterrichtsqualitätsmerkmalen die Stabilität so gering ist, dass weitaus mehr Unterrichtszeit beobachtet werden müsste, um eine stabile Einschätzung dieser Merkmale zu erhalten. So zeigte sich beispielsweise in der Untersuchung von Hill et al. (2012), dass keine der drei untersuchten Dimensionen – Reichhaltigkeit der mathematischen Inhalte, Fehler und Unpräzision seitens der Lehrkraft sowie die Partizipation der Schülerinnen und Schüler an Begründungen – beim Einsatz von zwei Raterinnen und Ratern im Rahmen einer Unterrichtsstunde zuverlässig erfasst werden

konnte. Die Untersuchung von Praetorius (in Druck) weist darauf hin, dass für eine stabile Erfassung der Unterrichtsdimension «kognitive Aktivierung» sogar neun Unterrichtsstunden pro Lehrkraft notwendig sind. Die genannte Untersuchung zeigt jedoch auch, dass es Merkmale gibt, die so stabil sind, dass eine einzelne Unterrichtsstunde für deren Einschätzung ausreicht; die Klassenführung von Lehrkräften ist hierfür ein Beispiel.

3.2.2 Fehlende Kontextinformationen

In den letzten Jahren hat sich in der Unterrichtsforschung ein Modell etabliert, das unter dem Namen «Angebots-Nutzungs-Modell» bekannt wurde (Abbildung 3; vgl. Helmke, 2009). Dieses Modell besagt, dass Unterricht immer nur ein Angebot sein kann, das von den Schülerinnen und Schülern genutzt werden muss und nur bei adäquater Nutzung zu Lerngewinnen führt. Es wird dabei davon ausgegangen, dass Unterricht immer an die jeweilige Klasse (u.a. an ihren Leistungsstand) angepasst werden sollte. Unterricht ist damit nicht nur abhängig von der jeweiligen Lehrkraft, sondern auch von den Schülerinnen und Schülern (u.a. von deren Lernpotenzial) und von vielen weiteren Merkmalen (u.a. von diversen Kontextmerkmalen). Solche zusätzlichen Informationen über die Schülerinnen und Schüler und den Kontext liegen externen Beobachterinnen und Beobachtern in der Regel jedoch nicht oder nur begrenzt vor (Clausen, 2002; Waldis et al., 2010).

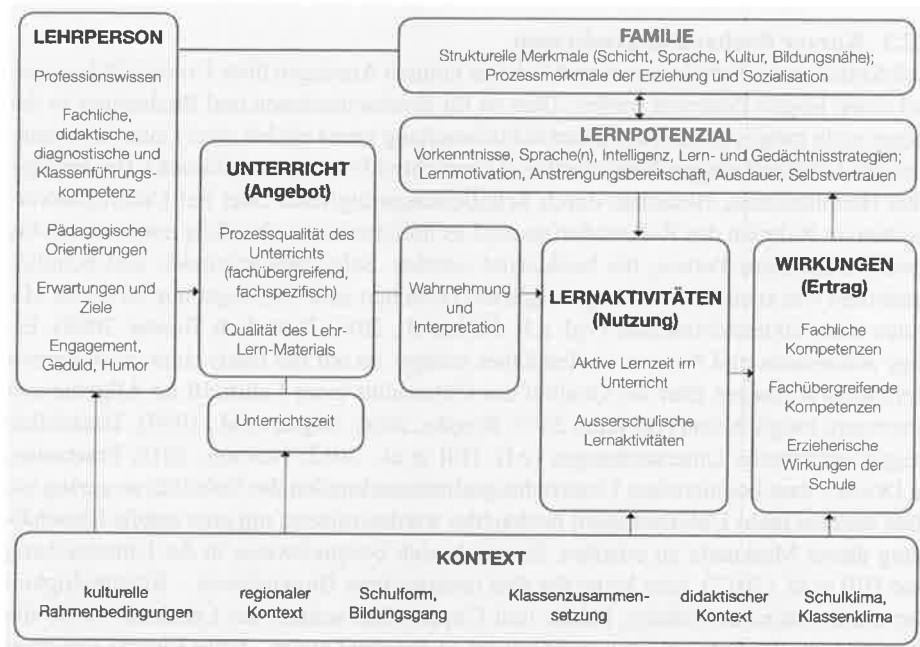


Abbildung 3: Angebots-Nutzungs-Modell des Unterrichts (Helmke, 2009, S. 73).

3.2.3 Reaktivitätseffekte

Wissen Lehrkräfte darum, dass eine bestimmte Unterrichtsstunde beobachtet werden soll, bereiten sie diese Stunde unter Umständen besonders gut vor oder setzen andere Methoden ein als in ihrem regulären Unterricht (Stigler, 1998; Stigler et al., 1999; Waldis et al., 2010). Diese Problematik wird als Reaktivitätseffekt bezeichnet: Personen verhalten sich unter Beobachtung anders, als sie es sonst tun würden. Stigler (1998, S. 141) beschreibt beobachtete Unterrichtsstunden daher als «[a] somewhat idealized version of what the teacher normally does in the classroom». Helmke (2009) hingegen verweist darauf, dass sich bestimmte Verhaltensweisen von Lehrkräften, insbesondere Routinen, kurzfristig kaum verändern lassen und daher auch beobachteter Unterricht einen guten Einblick in den regulären Unterricht bieten sollte. Beide Vermutungen, diejenige von Stigler (1998) sowie diejenige von Helmke (2009), scheinen plausibel zu sein. Welche tatsächlich besser auf Unterrichtsbeobachtungen zutrifft, ist ungeklärt: Bislang existieren keine empirischen Untersuchungen zu Reaktivitätseffekten in Bezug auf die Beobachtung von Unterricht.

4 Mangelnde Zuverlässigkeit von Beobachtereinschätzungen: Schlussfolgerungen

Externe Beobachterinnen und Beobachter werden oft als besonders geeignet zur Messung von Unterrichtsqualität beschrieben (Clare, Valdés, Pascal & Steinberg, 2001; Helmke, 2009; Pianta & Hamre, 2009). Bei näherer Betrachtung zeigen sich jedoch diverse Einschränkungen von Beobachterratings. So verfügen Beobachterinnen und Beobachter in der Regel lediglich über eine eingeschränkte Informationsgrundlage. Dies trägt dazu bei, dass ihre Einschätzungen die Unterrichtsqualität von Lehrkräften nur begrenzt abbilden können. Die in Abschnitt 3 dargestellten Studien zeigen zudem, dass externe Beobachtereinschätzungen zum Teil deutliche Ratereffekte aufweisen. Dies gilt in verstärktem Masse für ungeschulte Beobachterinnen und Beobachter wie beispielsweise Lehrkräfte.

Geht es Lehrkräften bei gegenseitigen Hospitationen lediglich darum, sich über die eigenen Vorstellungen von gutem Unterricht Gedanken zu machen und diese auszutauschen (vgl. Horster & Rolff, 2001), ist eine geringe Zuverlässigkeit der Einschätzungen nur von nachrangiger Bedeutung. Sollen die Einschätzungen hingegen als Unterrichtsdiagnose dienen und damit als Ist-Stand-Bestimmung die Grundlage für Unterrichtsentwicklung bilden, müssen hohe Maßstäbe an die Messqualität gelegt werden. Verzerrungen in den Einschätzungen können sonst im ungünstigsten Fall zu einer Verschlechterung anstelle einer Verbesserung von Unterricht führen (vgl. Praetorius et al., 2012).

Welche Schlussfolgerungen lassen sich daraus nun für die Unterrichtspraxis ziehen? Müller und Pietsch (2011) folgern aus der Messfehlerbehaftetheit der Einschätzungen

von Schulinspektorinnen und Schulinspektoren, dass die in diesem Kontext üblichen Aussagen auf Schulebene gerechtfertigt seien, Individualdiagnostik hingegen vermieden werden sollte. Für die Unterrichtspraxis ist eine solche Schlussfolgerung insofern problematisch, als Unterrichtsentwicklung und auch Bewertungen von Lehrkräften Ist-Stand-Bestimmungen auf Individualebene erfordern. Da davon auszugehen ist, dass Einschätzungen von Unterrichtsqualität aus dem Bauch heraus noch ungenauer und zudem invalide sind (Helmke et al., 2011), erscheint ein Verzicht auf Unterrichtseinschätzungen mittels standardisierter Instrumente nicht sinnvoll. Über mögliche Alternativen, die in der Unterrichtspraxis umsetzbar sind, ist bislang jedoch kaum etwas bekannt. Zieht man die berichteten Befunde heran, erweisen sich für die Umsetzung im schulischen Kontext mehrere Alternativen als gangbar – wenngleich diese jeweils mit einigem Aufwand verbunden sind:

- 1) Unterrichtsbewertungen, aber auch Rückmeldungen im Rahmen kollegialer Hospitationen beruhen gegenwärtig oft auf dem Urteil einer einzelnen Person. Diese Urteile enthalten in der Folge in hohem Masse auch Informationen über die Urteilerin oder den Urteiler und sind daher invalide. Um diese Problematik abzumildern, sollten – zumindest wenn es nicht lediglich um den Austausch von Vorstellungen über guten Unterricht unter Kolleginnen und Kollegen geht – stets mehrere Beobachterinnen und Beobachter eine Unterrichtsstunde beurteilen und deren Einschätzungen im Anschluss daran gemittelt werden, um individuelle Fehleinschätzungen zu minimieren.
- 2) Da nicht alle Unterrichtsqualitätsmerkmale innerhalb einer Unterrichtsstunde beurteilbar sind (z.B. kognitive Aktivierung, vgl. Abschnitt 3.2.1), sollten Beobachtungen zudem nach Möglichkeit auf einen längeren Zeitraum angelegt sein. Die Einschätzungen dieser Einzelbeobachtungen können danach gemittelt werden, was analog zur Mittelung mehrerer Raterinnen und Rater zu einer Minimierung des Fehleranteils in den Ratings führen sollte.
- 3) Aus diversen Forschungsbereichen (Übersicht in Hoyt & Kerns, 1999) wissen wir, dass Einschätzungen ungeschulter Beobachterinnen und Beobachter oftmals eine nur unzureichende Qualität aufweisen. Dies ist darauf zurückzuführen, dass die oben erwähnten Ursachen von Beobachtereffekten hier noch verstärkter auftreten. Insbesondere bei Unterrichtsbewertungen, aber auch in Bezug auf Unterrichtsrückmeldungen kommt einer Schulung von Beobachterinnen und Beobachtern (z.B. hospitierende Lehrkräfte oder Schulleitungsmitglieder) zur Einschätzung von Unterricht daher hohe Relevanz zu. Eine solche Schulung hat den Befunden von Praetorius et al. (2012) zufolge zwar keine positiven Effekte auf die Reliabilität der Urteile, gleichwohl aber auf deren Validität.
- 4) Neben der Optimierung der Unterrichtseinschätzungen externer Beobachterinnen und Beobachter ist auch ein Rückgriff auf Schülereinschätzungen möglich (z.B. Clausen, 2002; Kunter & Baumert, 2006; Lenske, 2012; Lüdtke, Trautwein, Kunter & Baumert, 2006; Marsh & Roche, 1997). Neben etlichen Vorteilen (u.a. der Möglichkeit eines langen Beurteilungszeitraums) werden jedoch auch in Bezug auf Einschätzungen von Schülerinnen und Schülern einige Nachteile thematisiert

(u.a. deren fehlendes methodisch-didaktisches Wissen), weswegen es sich anbietet, Schülereinschätzungen ergänzend zu und nicht anstelle von externen Beobachtereinschätzungen einzusetzen. Ein ungelöstes Problem ist dabei jedoch, dass Unterrichtseinschätzungen unterschiedlicher Perspektiven in der Regel eher geringe Übereinstimmungen aufweisen (z.B. Clausen, 2002: $-.28 \leq r \leq .45$), sodass sich im konkreten Fall die Frage stellt, auf welche Aspekte welcher Perspektive sich Lehrkräfte verlassen sollten.

«Auf den Lehrer kommt es an» (Lipowsky, 2006) – dies wurde in den letzten Jahren und Jahrzehnten vielfach empirisch bestätigt. Die im vorliegenden Beitrag dargestellten Befunde und Einschränkungen in Bezug auf die Interpretierbarkeit von Unterrichtseinschätzungen externer Beobachterinnen und Beobachter zeigen darüber hinaus: Auch auf die Messung kommt es an! Denn nur dann, wenn wir Unterrichtsqualität hinreichend genau und valide messen, können auf dieser Messung basierende wissenschaftliche Schlussfolgerungen sinnvolle Erkenntnisse bieten und unterrichtspraktische Entwicklungen zu einer Verbesserung von Unterricht führen.

Literatur

- Altrichter, H., Messner, E. & Posch, P.** (2004). *Schulen evaluieren sich selbst: Ein Leitfaden*. Seelze Velber: Kallmeyer'sche Verlagsbuchhandlung.
- Berliner, D.C.** (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56 (3), 205–213.
- Bradburn, N.M.** (2004). Understanding the question-answer process. *Statistics Canada*, 30 (1), 5–15.
- Brennan, R.L.** (2001). *Generalizability Theory*. New York: Springer.
- Brophy, J.** (2006). Observational research on generic aspects of classroom teaching. In P.A. Alexander & P.H. Winne (Hrsg.), *Handbook of educational psychology* (2. Auflage) (S. 755–780). Mahwah, NJ: Erlbaum.
- Clare, L., Valdés, R., Pascal, J. & Steinberg, J.** (2001). *Teachers' assignments as indicators of instructional quality in elementary schools* (CSE Technical Report Nr. 545). Los Angeles: National Center for Research on Evaluation.
- Clausen, M.** (2002). *Qualität von Unterricht – Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E.** (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31 (2), 122–141.
- Helmke, A.** (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. et al.** (2011). Unterrichtsdiagnostik – Voraussetzungen für die Verbesserung der Unterrichtsqualität. In A. Bartz, M. Dammann, S. Huber, C. Kloft & M. Schreiner (Hrsg.), *PraxisWissen SchulLeitung, AL 28* (Kapitel 30.71). Köln: Wolters Kluwer.
- Helmke, A. & Lenske, G.** (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge zur Lehrerbildung*, 31 (2), 214–233.
- Hill, H.C., Charalambous, C.Y. & Kraft, M.A.** (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41 (2), 56–64.
- Horster, L. & Rolff, H.-G.** (2001). *Unterrichtsentwicklung: Grundlagen einer reflektorischen Praxis* (2. Auflage). Weinheim: Beltz.

- Hoyt, W.T. & Kerns, M.-D.** (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4 (4), 403–424.
- Kobarg, M. & Seidel, T.** (2005). Coding manual – Process-oriented teaching. In T. Seidel, M. Prenzel & M. Kobarg (Hrsg.), *How to run a video study. Technical report of the IPN Video Study* (S. 108–144). Münster: Waxmann.
- Kunter, M. & Baumert, J.** (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9 (3), 231–251.
- Lemons, R.W. & Helsing, D.** (2010). High quality teaching and learning: Do we know it when we see it (and when we don't)? *Education Canada*, 48 (5), 14–18.
- Lenske, G.** (2012). *Schülerfeedback zur Unterrichtsqualität in der Grundschule: Studien zur Validität*. Unveröffentlichte Dissertation. Landau: Universität Koblenz-Landau.
- Lipowsky, F.** (2006). Auf den Lehrer kommt es an: Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (S. 47–70). Weinheim: Beltz.
- Lüdtke, O., Trautwein, U., Kunter, M. & Baumert, J.** (2006). Reliability and agreement of student ratings of the classroom environment: A reanalysis of TIMSS data. *Learning Environments Research*, 9 (3), 215–230.
- Marsh, H.W. & Roche, L.A.** (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52 (11), 1187–1197.
- Matsumura, L.C., Garnier, H.E., Pascal, J. & Valdés, R.** (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, 8 (3), 207–229.
- Müller, S. & Pietsch, M.** (2011). Was wir messen, wenn wir Unterrichtsqualität messen: Inter-Beurteilerübereinstimmung und -Reliabilität bei Unterrichtsbeobachtungen im Rahmen von Schulinspektion. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektionen in Deutschland – eine erste Zwischenbilanz* (S. 33–56). Münster: Waxmann.
- Newton, X.A.** (2010). Developing indicators of classroom practice to evaluate the impact of district mathematics reform initiative: A generalizability analysis. *Studies in Educational Evaluation*, 36 (1–2), 1–13.
- Petko, D., Waldis, M., Pauli, C. & Reusser, K.** (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik: Ansätze der TIMSS 1999 Video Studie und ihrer schweizerischen Erweiterung. *Zentralblatt für Didaktik der Mathematik*, 35 (6), 265–280.
- Pianta, R.C. & Hamre, B.K.** (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38 (2), 109–119.
- Pietsch, M. & Tosana, S.** (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität: Das Multifacetten-Rasch-Modell und die Generalisierbarkeitstheorie als Methoden der Qualitätssicherung in der externen Evaluation von Schulen. *Zeitschrift für Erziehungswissenschaft*, 11 (3), 430–452.
- Praetorius, A.-K.** (in Druck). *Eignung von hoch-inferenten Beobachterratings zur Messung von Unterrichtsqualität*. Münster: Waxmann.
- Praetorius, A.-K., Lenske, G. & Helmke, A.** (2010). *Auf der Suche nach Ursachen für Beurteilungsdifferenzen – Ist die Methode des Lauten Denkens für die Analyse unterrichtsbezogener Urteilsprozesse ertragreich?* Vortrag gehalten an der 74. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF), Jena.
- Praetorius, A.-K., Lenske, G. & Helmke, A.** (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, 22 (6), 387–400.
- Rakoczy, K.** (2008). *Motivationsunterstützung im Mathematikunterricht: Unterricht aus der Perspektive von Lernenden und Beobachtern*. Münster: Waxmann.
- Sommer, N.** (2011). Unterrichtsqualität im Urteil der externen Schulevaluation. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektionen in Deutschland – eine erste Zwischenbilanz* (S. 97–136). Münster: Waxmann.
- Stigler, J.** (1998). Video surveys: New data for the improvement of classroom instruction. In S.G. Paris & H.M. Wellman (Hrsg.), *Global prospects for education. Development, culture and schooling* (S. 129–168). Washington, DC: American Psychological Association.

- Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A.** (1999). *The TIMSS-Videotape Classroom Study* (Technical Report). Los Angeles: University of California.
- Strong, M., Gargani, J. & Hacifazlioglu, O.** (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62 (4), 367–382.
- Waldis, M., Grob, U., Pauli, C. & Reusser, K.** (2010). Der schweizerische Mathematikunterricht aus der Sicht von Schülerinnen und Schülern und in der Perspektive hochinferenter Beobachterurteile. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 171–208). Münster: Waxmann.

Autorin

Anna-Katharina Praetorius, Dr., Universität Augsburg, Lehrstuhl für Psychologie,
anna.praetorius@phil.uni-augsburg.de



Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung

Alois Buholzer und Sandra Zulliger

Zusammenfassung Die Diagnosefähigkeit von Lehrpersonen umfasst neben der Beurteilungsgüte vor allem die Fähigkeit, Diagnosen didaktisch adaptiv zu nutzen und lernfördernde Rückmeldungen daraus abzuleiten. Wie sich diagnostische Fähigkeiten bei angehenden Lehrpersonen und im Berufsalltag entwickeln, ist noch weitgehend unerforscht. Die vorliegende Längsschnittstudie mit PH-Studierenden untersucht die diagnostischen Fähigkeiten bezüglich formaler und inhaltlicher Aspekte im Fach Deutsch und diskutiert sie kritisch. Als Beurteilungsmaterial dienen videografierte Reportagen von Schülerinnen und Schülern sowie deren Vorbereitungsnotizen. Es konnte gezeigt werden, dass die Studierenden im Laufe der Ausbildung lernen, den Lernstand von Schülerinnen und Schülern differenzierter zu erfassen, und es ihnen auch immer besser gelingt, adäquate Formen der Weiterarbeit zu entwerfen. Bei anderen inhaltlichen und formalen Aspekten, z.B. der Fähigkeit zur Beurteilung des Wortschatzes oder des Textaufbaus, konnten während des Studiums hingegen keine signifikanten Veränderungen nachgewiesen werden. Die Diskussion der Ergebnisse im Kontext der Lehrerinnen- und Lehrerausbildung und früherer Forschungsarbeiten hebt die Wichtigkeit einer empirischen Überprüfung von Ausbildungszielen wie beispielsweise Diagnosekompetenz hervor und fordert weitere, vertiefte Analysen.

Schlagwörter diagnostische Fähigkeit – Kompetenzentwicklung – Lehrerinnen- und Lehrerbildung

The Development of Diagnostic Skills in Pre-Service Teachers During the Course of Their Study

Abstract Apart from the general quality of a given assessment, diagnostic skills of teachers particularly include the ability to make use of diagnoses in an instructionally adaptive way and to derive feedback which promotes the pupils' learning. However, it is still largely unexplored how diagnostic skills in pre-service teachers and in everyday teaching practice actually develop. The longitudinal study to be presented in this article investigates and critically discusses the diagnostic skills of teacher students from the University of Teacher Education Lucerne with respect to formal and content-related aspects in the subject German. Videoed school student reports and their preparatory notes served as assessment materials. The results show that students learn to assess the pupils' learning progress in a more differentiated way, and that they increasingly succeed in designing adequate forms of further work in the course of their study. No significant changes could be found concerning other content-related and formal aspects, e.g. the ability to assess vocabulary or text structure. The discussion of the findings in the context of teacher education

and previous research emphasizes the importance of an empirical examination of educational objectives such as diagnostic competence, and calls for further in-depth analyses.

Keywords diagnostic ability – competence development – teacher education

1 Einleitung

Diagnostische Fähigkeiten gehören zum *Grundinventar* des erfolgreichen Unterrichtens. Lehrpersonen beobachten kontinuierlich das Lernverhalten ihrer Schülerinnen und Schüler, ziehen Rückschlüsse auf den erworbenen Wissensstand, erschliessen mit Prüfungen, ob die anvisierten Lernziele erreicht werden, erteilen Noten und geben Prognosen für die künftige Schullaufbahn ab. Ihr diagnostisches Handeln ist vielfältig und das Spektrum möglicher Diagnoseaufgaben erstreckt sich «von den formellen Beurteilungsaufgaben (Notengebung) des Lehrers, die im Idealfall auf methodisch kontrolliert erhobenen und sorgfältig reflektiert verarbeiteten Informationen beruhen, bis zu den im Rahmen des unterrichtlichen Handelns unter Zeit- und Handlungsdruck kurzfristig zu erbringenden Urteilsleistungen» (Schrader, 1989, S. 46). Für die Erfüllung dieser breit angelegten Aufgaben sind die Lehrpersonen überwiegend auf sich allein gestellt. Es liegen zwar standardisierte Messinstrumente wie «Klassenscockpit» (vgl. www.klassenscockpit.ch) oder «Stellwerk» (vgl. www.stellwerk-check.ch) vor, welche die Möglichkeit bieten, die Leistungen der eigenen Klasse mit einer repräsentativen Stichprobe zu vergleichen. Dennoch bestimmen die Lehrpersonen weitgehend über die Prüfungsinhalte und die Art und Weise ihrer Überprüfung. Zudem entwickeln sie das Prüfinstrumentarium meistens selbst und legen auch die Beurteilungskriterien fest, anhand derer sie nach der Korrektur eine Einstufung der Leistungen ihrer Schülerinnen und Schüler vornehmen. Die einzelne Lehrperson entscheidet somit über Erfolg oder Misserfolg der Prüfung.

Lehrpersonen haben einen grossen Handlungs- und Gestaltungsspielraum bei der Diagnose und Beurteilung von Leistungen der Schülerinnen und Schüler. Dies erstaunt, werden doch diagnostische Analysen nicht nur durchgeführt, um Hinweise für die individuelle Lernförderung zu gewinnen, sondern auch, um bestimmte Berechtigungen (z.B. hinsichtlich der künftigen Schullaufbahn) zu vergeben. Da solche weitreichenden Entscheidungen letztlich auf den subjektiv gefärbten Beurteilungsprozessen und -ergebnissen der Lehrpersonen basieren, sind Ausbildung und Reflexion der eigenen diagnostischen Fähigkeiten im Laufe des Studiums für angehende Lehrpersonen umso wichtiger. Im vorliegenden Beitrag werden nach einer theoretischen Einführung Resultate einer Studie zur Entwicklung der diagnostischen Fähigkeiten während der Ausbildung vorgestellt. Im Fokus der Studie stehen dabei die Fragen, wie angehende Lehrpersonen eine Schülerleistung beurteilen, was sie der Schülerin oder dem Schüler rückmelden und wie sie mit ihr oder ihm weiterarbeiten würden.

Die empirische Unterrichtsforschung geht heute davon aus, dass es sich bei der Diagnosekompetenz *nicht um eine einheitliche Grundkompetenz* handelt, welche sich über die verschiedenen Fächer und diagnostischen Aufgaben hinweg generalisieren lässt. Diese Auffassung wird durch aktuelle Forschungsstudien bestätigt, die nur geringe Interkorrelationen zwischen den einzelnen Komponenten der diagnostischen Kompetenz nachweisen (Brunner, Anders, Hachfeld & Krauss, 2011; Spinath, 2005). Spinath (2005) konnte z.B. in ihrer Studie keine Korrelationen zwischen den Akkuratheitskomponenten innerhalb eines Merkmals und über die Merkmale hinweg feststellen und geht damit von einer hohen Merkmalspezifität aus. Sie legt nahe, den Begriff der diagnostischen Kompetenz nicht mehr zu verwenden und ihn durch «diagnostische Fähigkeiten» zu ersetzen. In neueren Arbeiten wird diese Anregung aufgenommen. So wird in der COACTIV-Studie der Begriff der diagnostischen Fähigkeiten genannt, die dem fachdidaktischen und dem pädagogisch-psychologischen Wissen zugeordnet werden (Anders, Kunter, Brunner, Krauss & Baumert, 2010; Brunner et al., 2011). Diagnostische Fähigkeiten sind somit spezifisch und stark mit den Fachinhalten gekoppelt, was im Bereich der Ausbildung von Lehrpersonen für eine enge Anbindung an fachwissenschaftliche und fachdidaktische Inhalte spricht.

2 Diagnostizieren als Grundlage für individuelle Lernförderung und Lernberatung

Die Urteilsgüte steht häufig im Zentrum der Diskussionen um die diagnostischen Fähigkeiten von Lehrpersonen. Diagnosen spielen im Hinblick auf das Unterrichtsgeschehen jedoch eine weit umfassendere Rolle. Sie sind eingebunden in den Planungsprozess von individuellen Förderangeboten oder wichtig für die Differenzierung des Unterrichts. Diagnosen werden so als Teil von didaktischen Überlegungen verstanden, um die Lernangebote adaptiv an die Schülerinnen und Schüler anzupassen (Horstkemper, 2006; Buholzer, 2006). Unterstrichen wird die enge Verbindung von diagnostischen und didaktischen Überlegungen durch das Ergebnis von Weinert und Helmke (1987), wonach Lehrpersonendiagnosen erst dann wirksam werden, wenn sie mit unterrichtlichen Massnahmen (starke Strukturierung des Unterrichtsangebots) zusammenhängen. Teil davon ist eine angemessene Rückmeldung der diagnostischen Erkenntnisse an die betreffende Schülerin oder den betreffenden Schüler. Insbesondere die Kommunikation von diagnostischen Urteilen kann für Lehrpersonen jedoch eine grosse Herausforderung darstellen, «weil immer die Gefahr besteht, dass die Lernmotivation eines Schülers oder einer Schülerin, die eine wesentliche Ressource ihres Bildungsprozesses darstellt, durch unpassende Äusserungen untergraben wird» (Herzog, 2013, S. 6). Gleichzeitig ist bekannt, dass systematische Rückmeldungen zum Lernstand und zur Lernentwicklung eines der wirksamsten Instrumente zur Förderung von Lernen darstellen. Dazu gehören sowohl die Analyse des Lernstands zusammen mit einer einzelnen Schülerin oder einem einzelnen Schüler als auch Überlegungen dazu, wie das Lernen verbessert werden könnte und welche Schritte anstehen, damit die Einzelne,

der Einzelne oder die Klasse als Ganze weiter vorankommen. Diagnosen bilden so die Grundlage für Lernberatungen, um über Fehler zu sprechen und die aufgezeigten Ressourcen weiter zu vertiefen. Ressourcen bilden einen wichtigen Teil von Diagnosen: nicht nur aus Motivationsgründen, sondern auch als Anknüpfungspunkt für ein nachfolgendes Lernangebot.

2.1 Unterschiedliche Qualitäten von diagnostischen Fähigkeiten

Diagnostische Fähigkeiten können bei Lehrpersonen *mehr oder weniger gut ausgebildet* sein; dies legen verschiedene Studien zur Genauigkeit von Urteilsleistungen nahe. Die Genauigkeit von Urteilen wird empirisch häufig durch einen Vergleich der Vorhersagewerte von Lehrpersonen mit den tatsächlich gemessenen Leistungen der Schülerinnen und Schüler erfasst. Aus diesen Studien können zwei Hauptergebnisse herausgelesen werden: Erstens gelingt den Lehrerinnen und Lehrern der soziale Vergleich recht gut. Dies zeigt sich an den treffgenauen Leistungsrangfolgen, die Lehrpersonen von ihren Klassen erstellen können. Zweitens geben die Lehrpersonen in Bezug auf die Beurteilung der absoluten Leistung einer Schülerin oder eines Schülers eher ungenaue Schätzungen ab (Schrader & Helmke, 1990; Spinath, 2004). Vergleicht man beispielsweise die erteilten Noten mit objektiven Testergebnissen, so stellte Ingenkamp bereits 1971 fest, dass für objektiv gleiche Rechenleistungen unterschiedliche Noten vergeben werden. Auch in neueren Studien erweisen sich Notensetzungen (auch zwischen verschiedenen Klassen) als nicht vergleichbar. In der MARKUS-Studie konnten mässige Korrelationen von $r = .30$ zwischen den Klassen aufgezeigt werden (Jäger, 2009). In der Meta-Analyse von Hoge und Coladarci (1989) zeigt sich eine mittlere Korrelation von $r = .66$ zwischen den Lehrpersoneneinschätzungen und den tatsächlichen Leistungen der Schülerinnen und Schüler. Auf der Basis von nach Hoge und Coladarci (1989) publizierten Studien haben Südkamp, Kaiser und Möller (2012) eine weitere Meta-Analyse durchgeführt. Sie fanden ebenfalls eine mittlere Korrelation von $r = .63$ zwischen den Lehrpersonenuurteilen und den Leistungen der Schülerinnen und Schüler in Standardtests.

2.2 Diagnostische Fertigkeiten und Fähigkeiten erwerben

Diagnostische Kompetenzen bestehen aus Fertigkeiten und Fähigkeiten, die in der Ausbildung vermittelt und später im Schulalltag verfeinert und ausgebaut werden. Wie andere professionelle Kompetenzen auch werden sie durch jahrelange praktische Erfahrung und deren theoriegeleitete Reflexion ausgebildet. Die Wissensgrundlage für den Aufbau von Kompetenzen wird in der Ausbildung gelegt «und in der Form von Reflexions- und Gestaltungswissen in die Praxis mitgenommen, sodass auf der Basis verallgemeinerungsfähigen wissenschaftlichen Wissens über individuelles Fallverstehen eine (dann zunehmend) professionelle Praxisbewältigung angegangen werden kann» (Lersch, 2006, S. 177). Es kann daher davon ausgegangen werden, dass der Grossteil der diagnostischen Kompetenzen im Rahmen der Berufstätigkeit und nicht in der Grundausbildung erworben wird (Hascher, 2008). Wie sich die einzelnen Kompetenzbereiche im Laufe der Berufsbiografie ausbilden, ist jedoch unklar. In dieser Hin-

sicht besteht sowohl theoretischer als auch empirischer Forschungsbedarf (Baumert & Kunter, 2006). Kaum untersucht ist auch die Entwicklung der Kompetenzen im Laufe der Ausbildung zur Lehrperson (Baer, Dörr, Guldemann, Kocher, Larcher, Müller & Wyss, 2008).

Diagnostische Kompetenz erschöpft sich nicht darin, einzelne Diagnoseinstrumente zu kennen und diese anzuwenden. Die diagnostizierende Lehrperson benötigt Metawissen, um angepasste Methoden und Instrumente auswählen, die Ergebnisse einordnen und schliesslich richtig interpretieren zu können. Das Wissen über Entwicklungsverläufe und Störungsbilder sowie deren Ursachen, Kenntnisse über Möglichkeiten der Intervention und Prävention wie auch das Wissen über die personellen und organisatorischen Zuständigkeiten sind auf das Engste miteinander verknüpft (Hascher, 2008). Hascher (2003) schlägt vor, dass Studierende nebst eigenen Lernerfahrungen im Unterricht, in Werkstätten und in Projekten insbesondere in der berufspraktischen Tätigkeit Diagnosen erstellen und diese zusammen mit der Praktikumsleiterin oder dem Praktikumsleiter evaluieren sollten. Wichtig scheint dabei die Reflexion der subjektiven Wahrnehmungs- und Beurteilungsprozesse zu sein: Woraus erschliesst sich, ob eine Schülerin den Auftrag verstanden statt? Woran erkennt man, dass ein Schüler bei einer Stillarbeit nicht mehr weiterkommt?

Für die Ausbildung werden von Oser (1997) in Bezug auf die pädagogische Diagnostik zwei Standardgruppen beschrieben. Unter dem ersten Standard «Schülerunterstützende Beobachtung und Diagnose» werden Ausbildungsziele genannt, welche den Umgang mit kritischen Entwicklungen und auftretenden Problemen betreffen. Der Umgang bezieht sich dabei sowohl auf einzelne Schülerinnen und Schüler wie auch auf das Geschehen im Klassenzimmer. Die zweite Standardgruppe umfasst die Ausbildungsziele zum Bereich «Leistungsmessung». Die Curricula von pädagogischen Hochschulen in der Schweiz orientieren sich unter anderem an diesen Standards von Oser (1997) und konkretisieren ihre Ziele in Kompetenzbeschreibungen. Diese Beschreibungen, die festhalten, was die PH-Studierenden zu unterschiedlichen Zeitpunkten und am Ende ihrer Ausbildung können und wissen sollen, bilden für die Ausgestaltung der verschiedenen Lehrangebote (Module) wie auch für die Gesamtausbildung einen wichtigen Orientierungsrahmen. Die Zielvorstellungen sind dahin gehend formuliert, dass deren Umsetzung nicht nur deklariert, sondern auch empirisch überprüft werden kann (Alleman-Ghionda & Terhart, 2006).

An der PH Luzern beispielsweise tragen Ausbildungsmodule der Fachdidaktik zusammen mit den Modulen der Bildungs- und Sozialwissenschaften dazu bei, bei den angehenden Lehrpersonen diagnostische Fähigkeiten aufzubauen. Im Referenzrahmen der Ausbildung wird die Diagnose- und Beurteilungskompetenz in Anlehnung an Ingenkamp und Lissmann (2008) folgendermassen umschrieben: «Um Lernprozesse möglichst optimal anzuregen und zu begleiten und die Prozesse und Leistungen der Schülerinnen und Schüler zu beurteilen, müssen Lehrpersonen das Verständnis ihrer

Schülerinnen und Schüler erkunden, ihre Lern- und Denkwege verstehen und Lernergebnisse analysieren» (PH Luzern, 2012, S. 12). Die diagnostische Kompetenz wird als Voraussetzung für die Adaptivität des Unterrichts angesehen (Beck, Baer, Guldemann, Bischoff, Brühwiler, Müller, Niedermann, Rogalla & Vogt, 2008), da die Fähigkeit, den Lernstand und die Leistungsfähigkeit der Schülerinnen und Schüler unter Berücksichtigung unterschiedlicher Bedingungen einschätzen zu können, eine Voraussetzung für die Auswahl von Aufgaben und einer angepassten Form der Unterstützung darstellt. Ob es aber in der Lehrerinnen- und Lehrerbildung gelingt, die oben diskutierten Ziele und Standards zu erreichen, wurde bisher wenig überprüft. Die nachfolgend vorgestellte Längsschnittuntersuchung (2003 bis 2007) geht der Frage nach, wie sich die diagnostischen Fähigkeiten von angehenden Lehrpersonen im Fach Deutsch während ihres Studiums verändern.

3 Studie zur Entwicklung von diagnostischen Fähigkeiten von PH-Studierenden

Wie vorangehend ausgeführt, gibt es nur wenige gesicherte empirische Erkenntnisse dazu, wie sich die diagnostischen Fähigkeiten von Lehrpersonen entwickeln. Zudem handelt es sich bei den diagnostischen Fähigkeiten nicht um ein einheitliches Konstrukt. Eine umfassende Erhebung der diagnostischen Fähigkeiten ist deshalb nicht möglich. In der vorzustellenden Studie werden sie als Fähigkeit aufgefasst, eine mündliche Deutschaufgabe unter Beachtung der Ziele der Aufgabenstellung angemessen einzuschätzen, auf dieser Grundlage eine Rückmeldung zu geben sowie einen Vorschlag zur Weiterarbeit zu unterbreiten. Die Ergebnisse beziehen sich somit nur auf das Fach Deutsch und auf die Arbeitsergebnisse einzelner Schülerinnen oder Schüler.

3.1 Design der Längsschnittstudie

Im Rahmen der Qualitätsüberprüfung der Lehre an der damals neu gegründeten PHZ Luzern¹ wurde von 2003 bis 2007 eine Längsschnittstudie durchgeführt. Die diagnostischen Fähigkeiten der Studierenden im Fach Deutsch wurden am Anfang (2003), in der Mitte (2005) und am Ende ihres Studiums (2006/2007) mit der immer gleichen experimentellen Methode erfasst. Die Studie begann gleichzeitig mit der Aufnahme des Lehrbetriebs an der PHZ Luzern im Jahr 2003. Für die Bestimmung der Fortschritte während der Ausbildung wurde mit einem echten Längsschnitt gerechnet. Für den Vergleich von Anfang und Mitte des Studiums konnte mit 60 Studierenden der Sekundarstufe I gerechnet werden und für den Vergleich von Anfang und Ende des Studiums mit 28 Studierenden. Im Folgenden werden die Resultate für die Studierenden der Sekundarstufe I detailliert aufgezeigt. Ebenfalls untersucht, aber in diesem Beitrag nicht weiter berücksichtigt wurden auch die Studierenden aus den Studiengängen Kindergarten/Unterstufe sowie Primarstufe.

¹ Am 1. August 2013 wurde aus der PHZ Luzern die Pädagogische Hochschule Luzern (PHLU).

Zur Erfassung ihrer diagnostischen Fähigkeiten wurde den Studierenden eine Videoaufnahme einer Schülerin oder eines Schülers gezeigt, die oder der eine selbst verfasste zweiminütige Radioreportage zu einem Strandbad vorstellt. Die den Schülerinnen und Schülern vorgelegte Aufgabe, die den Orientierungsarbeiten zum Fach Deutsch entnommen wurde, überprüft zum mündlichen Sprachgebrauch das Lernziel «Über Ereignisse und Beobachtungen in der Standardsprache berichten». Den Studierenden wurden am Anfang ihres Studiums, d.h. zum ersten Messzeitpunkt, zwei Radioreportagen – eine von Schülerin/Schüler A und eine von Schülerin/Schüler B – gezeigt. Zu Arbeitsergebnis A (und zu einem weiteren Arbeitsergebnis, das jedoch nicht ausgewertet wurde) wurde bei den Studierenden in der Mitte des Studiums, d.h. zum zweiten Messzeitpunkt, eine (erneute) Beurteilung eingeholt. Arbeitsergebnis B wiederum wurde den Studierenden (zusammen mit einem weiteren, nicht ausgewerteten Arbeitsergebnis) am Ende des Studiums nochmals vorgelegt. Durch die Variation der Arbeitsergebnisse sollte der Wiedererkennungseffekt über die drei Messzeitpunkte hinweg verringert werden; trotzdem dürfte die Vergleichbarkeit durch das zweimalige Vorlegen des gleichen Arbeitsergebnisses gewährleistet sein.

Zur gezeigten Radioreportage wurde den Studierenden ein Fragebogen² mit vier offenen Fragen ausgeteilt. Als Zusatzinformationen enthielt der Fragebogen den Arbeitsauftrag an die Schülerinnen und Schüler mitsamt den Lernzielen, die Beurteilungskriterien sowie die Notizen der betreffenden Schülerin oder des betreffenden Schülers. Die Studierenden wurden danach gefragt, was sie auf der Basis der Notizen der Schülerin bzw. des Schülers und der beobachteten Videosequenz in Bezug auf den Lernstand feststellen, was sie rückmelden und wie sie mit ihr oder ihm weiterarbeiten würden.

Als Vergleichsmaßstab zu den Angaben der Studierenden wurde ein Expertenurteil erstellt. Die Expertenrunde bestand aus je einer Fachperson der Zielstufe und der Fachdidaktik Deutsch sowie zwei Fachpersonen der Bildungs- und Sozialwissenschaften. Die Expertinnen und Experten beurteilten das Arbeitsergebnis unter den gleichen Bedingungen wie die Studierenden. Die einzelnen Experteneinschätzungen wurden zu einem Gesamturteil zusammengefasst, welches als Grundlage für die Erstellung eines Kategoriensystems mit inhaltlichen und formalen Kategorien diente. Die inhaltlichen Kategorien nehmen Bezug auf die Beurteilung des verwendeten Wortschatzes, des Aufbaus des Arbeitsergebnisses, der Stimmführung und der Aussprache sowie auf den Inhalt der Weiterarbeit. Mit den formalen Kategorien wird erfasst, ob der Lernstand differenziert dargestellt wurde, die Beurteilungen sich auf Tatsachenbeschreibungen beziehen und die Rückmeldungen sowie die vorgeschlagene Weiterarbeit einen Bezug zur Aufgabenstellung aufweisen. Bei den inhaltlichen Kategorien variiert die Anzahl der erfassten Kriterien je nach beurteiltem Arbeitsergebnis. Mithilfe des Kategoriensystems wurde eine quantitative Inhaltsanalyse durchgeführt. Die offenen Antworten der Studierenden wurden dahin gehend ausgewertet, ob der jeweilige im Kategoriensystem festgehaltene Aspekt ganz, teilweise oder gar nicht vorhanden war. Dafür wurde eine dreistufige

² Der Fragebogen kann beim Autor und bei der Autorin dieses Beitrags bezogen werden.

Skala verwendet (0 = nicht vorhanden; 1 = teilweise vorhanden; 2 = vorhanden). Zur Bestimmung der Fortschritte wurden mit dem Wilcoxon-Test die jeweiligen Werte der einzelnen Messzeitpunkte miteinander verglichen. Das Signifikanzniveau wurde nach Bonferroni-Holm korrigiert und das Ausgangsniveau für ein signifikantes Ergebnis auf 0.05 festgelegt.

3.2 Fortschritte bei den einzelnen diagnostischen Fähigkeiten

Durch die Auswertung der Ergebnisse kann insbesondere aufgezeigt werden, welchen Aspekten die Studierenden bei der Beurteilung des Arbeitsergebnisses Beachtung geschenkt haben und welchen nicht bzw. auf welche Bereiche sie während ihres Studiums sensibilisiert wurden. In Tabelle 1 werden die Werte der Studierenden der Sekundarstufe I und die signifikanten Fortschritte in den einzelnen Kategorien über die drei Messzeitpunkte hinweg aufgeführt.

Betrachtet man das Gesamtergebnis aller Kategorien, so zeigen sich bei den Studierenden der Sekundarstufe I keine signifikanten Unterschiede zwischen dem ersten und

Tabelle 1: Studierende der Sekundarstufe I: Veränderungen über die drei Messzeitpunkte hinweg

		Mittelwerte und Standardabweichungen									
		t ₁		t ₂		t ₃		t ₁ - t ₃		B	
		A	B	A	B	A	B	A	B		
M	SD	M	SD	M	SD	M	SD	B			
Formale Kategorien	Differenzierung Lernstand	1.2	0.5	0.9	0.5	1.5	0.4	1.3	0.3	$p = 0.004$	
	Interpretation Lernstand	0.8	0.3	0.8	0.4	1.1	0.4	1.1	0.4	n.s.	
	Fokus Rückmeldung	1.6	0.5	1.2	0.7	1.3	0.6	1.6	0.6	n.s.	
	Fokus Weiterarbeit	1.2	0.5	1.2	0.4	1.0	0.4	1.4	0.4	n.s.	
Formale Kategorien (gesamt)		1.2	0.3	1.0	0.3	1.2	0.3	1.4	0.3	$p = 0.001$	
Inhaltliche Kategorien	Beurteilung Wortschatz	0.6		0.7	0.3	0.6		0.6		n.s.	
	Beurteilung Aufbau	0.9	0.8	0.7	0.5	1.0	0.7	0.8	0.6	n.s.	
	Beurteilung Stimmführung	0.8	0.4	0.6	0.4	0.8	0.4	0.9	0.3	n.s.	
	Beurteilung Aussprache	0.4	0.4	0.1	0.2	0.6	0.4	0.4	0.4	$p = 0.003$	
	Angemessenheit Weiterarbeit	0.4	0.5	0.2	0.3	0.4	0.3		0.5	$p = 0.000$	
Inhaltliche Kategorien (gesamt)		0.6	0.3	0.5	0.2	0.7	0.3	0.7	0.2	$p = 0.001$	
Gesamt		0.9	0.2	0.7	0.2	0.9	0.2	1.0	0.2	$p = 0.000$	

Anmerkungen: M = Mittelwert; SD = Standardabweichung.

Beurteilung von Schülerin/Schüler A: $n = 60$; Beurteilung von Schülerin/Schüler B: $n = 28$.

Die Mittelwerte können zwischen 0 und 2 variieren (0 = nicht vorhanden; 1 = teilweise vorhanden; 2 = vorhanden).

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben, signifikante Unterschiede ($p < 0.05$) nach der Bonferroni-Holm-Korrektur.

dem zweiten, jedoch zwischen dem ersten und dem dritten Messzeitpunkt. Die Gesamtmittelwerte der drei Messzeitpunkte variieren lediglich zwischen 0.7 und 1.0 und streuen nur sehr geringfügig um den jeweiligen Mittelwert. Die Mittelwerte der inhaltlichen Kategorien liegen im Vergleich zu den formalen Kategorien wesentlich tiefer und somit weiter weg vom Expertenurteil. Da auf eine Kontrollgruppe verzichtet wurde, sind die unterschiedlichen Werte (und Entwicklungen) zwischen den Kategorien von besonderem Interesse für die Auswertung.

Bei den formalen Aspekten der Beurteilung zeigen sich deutliche Fortschritte der Studierenden bei der differenzierten Darstellung des Lernstands («Differenzierung Lernstand»). Eine differenzierte Darstellung des Lernstands liegt dann vor, wenn mindestens fünf inhaltlich verschiedene Aspekte präzise beschrieben werden (was und wie etwas ist) und der Situationsbezug (z.B. Videoaufnahme) berücksichtigt wurde. Sowohl zwischen dem Anfang und der Mitte als auch zwischen dem Anfang und dem Ende des Studiums machen die Studierenden hier hochsignifikante Fortschritte in der Beurteilung. Die Studierenden haben somit gelernt, genauer hinzuschauen, differenziertere Einschätzungen vorzunehmen und den Lernstand präziser zu beurteilen. Inwieweit die Studierenden ihre Beurteilungen belegt und interpretiert haben, wurde unter der formalen Kategorie «Interpretation Lernstand» zusammengefasst. Die Studierenden der Sekundarstufe I weisen bereits in der Mitte des Studiums einen Fortschritt auf, nicht jedoch am Ende des Studiums. Mit der Kategorie «Fokus Rückmeldung» wurde erfasst, ob sich die Rückmeldung, die die Studierenden der betreffenden Schülerin oder dem betreffenden Schüler geben würden, auf die Aufgabenstellung bezieht. Bei dieser Kategorie ist zwischen dem Anfang und der Mitte des Studiums ein signifikanter Rückgang von 1.6 auf 1.3 zu beobachten, während sich zwischen dem Anfang und dem Ende des Studiums keine Unterschiede zeigen. Diese Abnahme in der Mitte des Studiums ist anhand der vorliegenden Daten schwer zu erklären. Es scheint, dass sich die Studierenden in der Mitte ihrer Ausbildung freiere Rückmeldungen zutrauen und erst auf das Ende des Studiums hin lernen, ihre Rückmeldungen auf die Aufgabenstellung zu fokussieren.

Bei den einzelnen inhaltlichen Kategorien zeigen sich keine signifikanten Fortschritte (mit Ausnahme der Beurteilung der Aussprache und der Angemessenheit der Weiterarbeit). Obwohl die Ausgangswerte relativ tief liegen und somit eine Annäherung an das Expertenurteil möglich wäre, wurden keine Fortschritte erzielt. Ein möglicher Grund dafür könnte darin bestehen, dass sich die Studierenden im Rahmen der Ausbildung nicht mit diesen Lerninhalten befasst haben oder dass das Gelernte nicht auf die vorgelegte Aufgabenstellung transferiert werden konnte. Ein weiterer Grund für das Ausbleiben von Fortschritten im inhaltlichen Bereich könnte in der Ausgestaltung des Forschungsdesigns liegen. Beim Expertenurteil haben Personen aus verschiedenen Fachbereichen mitgewirkt, wodurch insbesondere bei den inhaltlichen Kategorien eine Annäherung an das kumulierte Expertenurteil für die Studierenden schwieriger wird.

In Bezug auf die Vorschläge zur Weiterarbeit kann festgehalten werden, dass unter formalen Gesichtspunkten bei den Rückmeldungen zur Weiterarbeit («Fokus Weiterarbeit») während der Ausbildung keine Fortschritte erzielt wurden. Hier wurde beurteilt, ob die Vorschläge zur Weiterarbeit innerhalb der Aufgabenstellung liegen. Unter inhaltlichen Gesichtspunkten hingegen sind deutliche Fortschritte zwischen dem ersten und dem letzten Messzeitpunkt zu verzeichnen. Das heisst, es gelingt den Studierenden immer besser, aus den diagnostischen Urteilen eine adäquate Weiterarbeit abzuleiten, d.h. in diesem Fall, eine nochmalige Diagnose oder eine Wiederholung der Übung auf der Grundlage eines Analogbeispiels, eines Trainings einer Teilfertigkeit und/oder der Veränderung der Lernsituation vorzuschlagen.

4 Schlussfolgerungen

Aus den Studienergebnissen geht hervor, dass die Studierenden deutliche Fortschritte bei der differenzierten Darstellung des Lernstands insgesamt und insbesondere in Bezug auf die Einschätzung der Aussprache erzielt haben. Die Studierenden sind darüber hinaus zunehmend in der Lage, aus der Beurteilung Schlüsse für die weitere Arbeit zu ziehen. Dieser Punkt ist nicht unwesentlich, weiss man doch aus der empirischen Unterrichtsforschung, dass die Fähigkeit von Lehrpersonen, ein Schülermerkmal treffend zu beurteilen, bei den Schülerinnen und Schülern allein noch nicht zu einem Lernfortschritt führt, sondern erst in Verbindung mit einer daran anschliessenden Strukturierung des Angebots (Weinert & Helmke, 1987). Es wird aber auch ersichtlich, dass sich – in Übereinstimmung mit anderen Ergebnissen von Studien zur Diagnosekompetenz (z.B. Spinath, 2005) – nicht alle Facetten der Diagnosekompetenz in gleichem Masse weiterentwickeln. So zeigen sich bei den inhaltlichen Aspekten Wortschatz, Aufbau des Arbeitsergebnisses und Stimmführung während des Studiums keine signifikanten Fortschritte in der Beurteilung. Offenbar setzen sich Beurteilungsprozesse aus unterschiedlichen Subprozessen zusammen, die im Laufe der Ausbildung unterschiedlich stark aufgegriffen und entwickelt werden. Die unterschiedlichen Annäherungen könnten ausserdem auf die Tatsache zurückgeführt werden, dass es sich beim Vergleichsmassstab um eine normative Setzung, gebildet durch die Einschätzungen der Expertinnen und Experten, handelt.

Mit dieser – im Rahmen der Qualitätsüberprüfung durchgeführten – Untersuchung konnten zwar erste Hinweise zum Aufbau von diagnostischen Fähigkeiten gewonnen werden, für Ableitungen von Implikationen für die Aus- und Weiterbildung ist es jedoch momentan noch zu früh – dafür sind weitere Studien erforderlich. Wir wissen erst wenig darüber, wie die diagnostischen Urteile zustande gekommen sind und welche Kognitionen im Beurteilungsprozess leitend waren. Wie Studien belegen, hängen diagnostische Urteile auch mit Kontextfaktoren zusammen; beispielsweise fällt es Lehrpersonen in Klassen mit ausgeprägter Heterogenität leichter, akkurate Diagnosen zu stellen (Karing, 2009). Auch der Zweck der Beurteilung unterstützt oder mildert

Urteilsverzerrungen (Krolak-Schwerdt, Böhmer & Gräsel, 2009). Diesen Kontextbedingungen bzw. der ökologischen Validität müsste bei einer revidierten Studiendurchführung mehr Beachtung geschenkt werden. Dazu gehört auch eine differenziertere Erfassung der Voraussetzungen bei den Studierenden (z.B. Erfahrungshintergrund und Vorwissen). Zudem müsste mit einem Kontrollgruppendesign überprüft werden, wie sich die Kompetenzentwicklung der PH-Studierenden im Vergleich zu Studierenden ohne entsprechende Fachinputs entwickelt. Nur so können kausale Rückschlüsse auf Lerninhalte und Studienaufbau gezogen werden.

Literatur

- Allemann-Ghionda, C. & Terhart, E. (2006). Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf. *Zeitschrift für Pädagogik*, 51. Beiheft, 7–11.
- 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.
- Baer, M., Dörr, G., Guldemann, T., Kocher, M., Larcher, S., Müller, P. & Wyss, C. (2008). Wirkt Lehrerbildung? – Kompetenzaufbau und Standarderreicherung in der berufswissenschaftlichen Ausbildung an drei Pädagogischen Hochschulen in der Schweiz und in Deutschland. *Empirische Pädagogik*, 22 (3), 259–273.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M. & Vogt, F. (2008). *Adaptive Lehrkompetenz. Analyse von Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- 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.
- Buholzer, A. (2006). *Förderdiagnostisches Sehen, Denken und Handeln*. Donauwörth: Auer.
- Hascher, T. (2003). Diagnose als Voraussetzung für gelingende Lernprozesse. *Journal für Lehrerinnen- und Lehrerbildung*, 3 (2), 25–30.
- Hascher, T. (2008). Diagnostische Kompetenzen im Lehrberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung* (S. 71–86). Münster: Waxmann.
- Herzog, W. (2013). Beurteilungskultur. *Profi-L*, Heft 1, 4–6.
- Hoge, R.D. & Coladarci, T. (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research*, 59 (3), 297–313.
- Horstkemper, M. (2006). Fördern heisst diagnostizieren. Pädagogische Diagnostik als wichtige Voraussetzung für individuellen Lernerfolg. *Friedrich Jahresheft XXIV: Diagnostizieren und Fördern*, 4–7.
- Ingenkamp, K. (Hrsg.). (1971). *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte*. Weinheim: Beltz.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jäger, R.S. (2009). Diagnostische Kompetenz und Urteilsbildung als Element von Lehrprofessionalität. In O. Zlatkin-Troitschanskaia (Hrsg.), *Lehrprofessionalität: Bedingungen, Genese, Wirkungen und ihre Messung* (S. 105–116). Weinheim: Beltz.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 197–209.

- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C.** (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 175–186.
- Lersch, R.** (2006). Lehrerbildung im Urteil der Auszubildenden. Eine empirische Studie zu beiden Phasen der Lehrerausbildung. *Zeitschrift für Pädagogik*, 51. Beiheft, 164–181.
- Oser, F.** (1997). Standards in der Lehrerbildung. Teil I: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerbildung*, 15 (1), 26–37.
- PH Luzern.** (2012). *Referenzrahmen der PH Luzern*. Luzern: PH Luzern.
- Schrader, F.-W.** (1989). *Diagnostische Kompetenz von Lehrern und ihre Bedeutsamkeit für die Gestaltung des Unterrichts*. Frankfurt am Main: Lang.
- Schrader, F.-W. & Helmke, A.** (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22 (4), 312–324.
- Spinath, B.** (2004). Diagnostische Kompetenzen von Lehrerinnen und Lehrern. *Lernende Schule*, Nr. 26, 16–17.
- 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.
- 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.
- Weinert, F.E. & Helmke, A.** (1987). Schulleistungen – Leistungen der Schule oder der Kinder? *Bild der Wissenschaft*, 24, 62–73.

Autor und Autorin

Alois Buholzer, Prof. Dr., PH Luzern, Institut für Schule und Heterogenität (ISH),
alois.buholzer@phlu.ch
Sandra Zulliger, lic. phil., PH Luzern, Institut für Schule und Heterogenität (ISH),
sandra.zulliger@phlu.ch



Der Educational Profiler: Ein Instrument für die Diagnose der pädagogischen Bezugssysteme angehender Lehrpersonen

Guido McCombie und Damian Läge

Zusammenfassung Lehramtsstudierende treten mit einem System an impliziten Überzeugungen in die Ausbildung ein, welches man als ihr pädagogisches Bezugssystem bezeichnen kann. Dieses kann die Aufnahme und die Verarbeitung von Ausbildungsinhalten beeinflussen, da es als Filter wirkt. Der vorliegende Beitrag beschreibt ein Instrument zur Diagnose von solchen impliziten pädagogischen Bezugssystemen im Rahmen der Lehrerinnen- und Lehrerbildung. Es handelt sich dabei um den Educational Profiler (E-Profiler), der an der Universität Zürich entwickelt wurde. Er erhebt die subjektiven Überzeugungen jeder und jedes Studierenden auf standardisierte Weise und in Relation zueinander. Empirische Daten zeigen, dass der E-Profiler einerseits als Forschungsinstrument für wissenschaftliche Fragestellungen und andererseits als individuelles Reflexionsinstrument in der Lehrerinnen- und Lehrerbildung von Nutzen sein kann.

Schlagwörter pädagogische Überzeugungen – pädagogisches Bezugssystem – Lehrerinnen- und Lehrerbildung

Educational Profiler: A Tool for the Diagnosis of Educational Reference Systems of Future Teachers

Abstract Students enter teacher education with a set of implicit beliefs, which can be regarded as a personal educational reference system. This reference system can influence perception and processing of new information in the course of teacher education, because it acts as a kind of filter. The following article describes a tool for diagnosing such implicit educational reference systems in the context of teacher education. It was developed at the University of Zurich and it is called Educational Profiler (E-Profiler). It provides a standardised and relational assessment of students' subjective beliefs. Empirical data show that E-Profiler can be a useful tool for addressing both scientific research questions and the process of individual reflection in teacher education.

Keywords educational beliefs – educational reference system – teacher education

1 Einleitung

Berufsbezogene Überzeugungen spielen eine wichtige Rolle für die Qualität des Lehrhandelns, «indem sie die Auswahl von Zielen und Handlungsplänen, die Wahrnehmung und Deutung von Situationen, ... beeinflussen und steuern» (Reusser, Pauli & Elmer, 2011, S. 478; vgl. auch Calderhead, 1996; Woolfolk Hoy, Davis & Pape,

2006). Es muss deshalb das erklärte Ziel der Lehrerinnen- und Lehrerbildung sein, auf diese Überzeugungen Einfluss nehmen zu können. Die Lehrerinnen- und Lehrerbildung sieht sich dabei mit der Herausforderung konfrontiert, dass die angehenden Lehrpersonen mit einem breiten Set an berufsbezogenen Überzeugungen in die Ausbildung eintreten (Blömeke, 2005; Fang, 1996; Kane, Sandretto & Heath, 2002; Pajares, 1992). Wenngleich verschiedene Studien darauf hinweisen, dass sich berufsbezogene Überzeugungen im Verlauf der Ausbildung zur Lehrperson durchaus verändern können (Dann, Cloetta, Müller-Fohrbrodt & Helmreich, 1978; Tanner, 1993; Blömeke, Müller, Felbrich & Kaiser, 2008; Biedermann, Brühwiler & Krattenmacher, 2012), gelten Überzeugungen generell als veränderungsresistent (Nespor, 1987; Pajares, 1992). Dies hat damit zu tun, dass die eigenen Überzeugungen bei der Wahrnehmung und Verarbeitung von Informationen als Bezugssystem agieren, auf dessen Grundlage konkret wahrgenommene Ereignisse interpretiert und bewertet werden.

Lerntheoretisch können die Überzeugungen von angehenden Lehrpersonen auch als Chance angesehen werden. Überzeugungen sind dann «Schemata» zu relevanten Konstrukten der Lehrerinnen- und Lehrerbildung und erhalten so den Charakter von Vorwissen, welches für den Lernprozess ein entscheidendes Element darstellt (Ausubel, Novak & Hanesian, 1978; Dochy, 1992; Schraw, 2006). Eine direkte Bezugnahme auf die Überzeugungen der Studierenden gelingt jedoch selten. Das Hauptproblem ist dabei, dass das eigene System von subjektiven Überzeugungen nur schwer zugänglich ist, da es sich um implizite Konstrukte handelt, welche nur selten explizit reflektiert werden. Sie wirken vielmehr im Hintergrund (Bischof, 1996; Pajares, 1992). Subjektive berufsbezogene Überzeugungen lassen sich deshalb als ein pädagogisches Bezugssystem beschreiben (McCombie & Läge, 2012). Das eigene Bezugssystem umfasst die subjektiven Ideen, Anschauungen und Einstellungen, von deren Richtigkeit man überzeugt ist. Der Begriff des Bezugssystems beruht auf den Überlegungen von Bischof (1996, 2009) und hat seine Wurzeln in der aus der Gestaltpsychologie bekannten Terminologie von Figur und Grund (Müsseler, 2007). Bischof erweitert den Begriff des Hintergrunds und nennt ihn «Medium», was stärker den umschliessenden, einbettenden Charakter des Hintergrunds betont. Die Figur weist eine konkrete Gestalt auf. Das Medium hingegen «fällt kaum auf. Es ist amorph und eben daher auch anonym» (Bischof, 1996, S. 127). Das pädagogische Bezugssystem ist ein klassisches Beispiel für ein Medium. Das Bezugssystem dient dazu, Handlungsoptionen zu bewerten und einzuordnen. Es liefert uns das Koordinatensystem, die nötige Orientierungshilfe, damit wir uns zurechtfinden können (Bischof, 1996, S. 76).

Beschreibt man die berufsbezogenen Überzeugungen von Studierenden als pädagogisches Bezugssystem, so wird ersichtlich, dass die Überzeugungen nicht isoliert nebeneinander bestehen, sondern zusammen ein implizites System aus systematischen Zusammenhängen bilden. Gleichzeitig hat die Ausprägung der einzelnen Komponenten aber stark idiosynkratischen und zwischen Personen breit variierenden Charakter. Um mit oder an den Überzeugungen von Personen arbeiten zu können, müssen

sie als Erstes bewusst gemacht werden, denn Erfolg versprechend sind Lehrprozesse dann, wenn sie «aktiv an die vorhandenen beliefs anknüpfen ...» (Blömeke, 2005, S. 8). Erst die Diagnose von subjektiven Überzeugungen erlaubt es, diese zu erkennen und sich bewusst damit auseinanderzusetzen. Im Folgenden wird nun ein Instrument vorgestellt, welches für sich beansprucht, eine solche Diagnose valide und reliabel leisten zu können. Es ist der an der Universität Zürich entwickelte Educational Profiler (McCombie & Läge, 2012).

2 Der Educational Profiler

Der Educational Profiler (E-Profiler) erfasst die subjektiven Überzeugungen in einem System von acht inhaltlichen Bereichen mit je zwei Facetten (vgl. Tabelle 1; für die genaue Definition der einzelnen Facetten und eine Übersicht über die Items vgl. McCombie & Läge, 2012).

Tabelle 1: Die acht Bereiche mit je zwei Facetten des pädagogischen Bezugssystems

Bereich 1: Lehren und Lernen Facette 1a: Konstruktion Facette 1b: Instruktion	Bereich 5: Schulischer Auftrag 1 Facette 5a: Qualifikation Facette 5b: Sozialisation
Bereich 2: Erziehungsstil Facette 2a: Selbstbestimmung Facette 2b: Fremdbestimmung	Bereich 6: Schulischer Auftrag 2 Facette 6a: Selektion Facette 6b: Integration
Bereich 3: Formbarkeit Facette 3a: Umwelt Facette 3b: Anlage	Bereich 7: Leistung Facette 7a: Fordern Facette 7b: Fördern
Bereich 4: Verhaltensklärung Facette 4a: Situational Facette 4b: Dispositional	Bereich 8: Beziehung Facette 8a: Distanz Facette 8b: Nähe

Insgesamt werden die Überzeugungen also auf 16 Skalen erfasst. Die Darstellung erfolgt in einem Spinnendiagramm, wobei die Lage der einzelnen Facetten der empirisch ermittelten Ähnlichkeit der einzelnen Facetten entspricht (McCombie & Läge, 2012). Als optimale Struktur erweist sich eine Circumplex-Lösung. Diese Kreis- oder Ringstruktur ist dasjenige Modell, welches die empirisch ermittelten Ähnlichkeitsbeziehungen zwischen den Facetten optimal abbildet. Ähnliche Facetten liegen nahe beieinander, gegensätzliche Facetten liegen einander gegenüber (vgl. Abbildung 1).

Die Hälfte der Facetten (jene aus den Bereichen «Erziehungsstil», «Formbarkeit», «Verhaltensklärung» und «Beziehung») sind in der methodischen Anlage als klar bipolare Facetten definiert. Bei den restlichen vier Bereichen wurde demgegenüber davon aus-

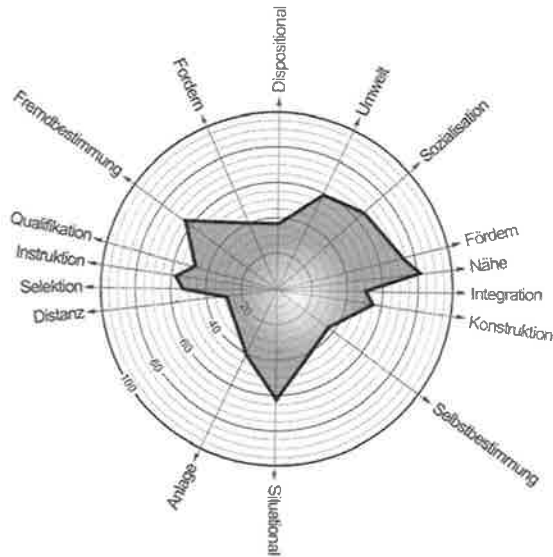


Abbildung 1: Anordnung der Facetten in der Form eines Spinnendiagramms. Die Nähe der einzelnen Facetten spiegelt deren empirisch ermittelte Ähnlichkeit wider. Eingezeichnet ist ein Beispielprofil. Jede Facette hat eine mögliche Ausprägung zwischen 0 und 100 Punkten. Je höher der Wert, desto stärker ist die Zustimmung zur Facette.

gegangen, dass sich die jeweiligen Facetten gegenseitig nicht ausschliessen, d.h. dass es theoretisch durchaus möglich ist, in beiden Facetten dieser Bereiche ähnlich hohe oder tiefe Werte zu erzielen (McCombie & Läge, 2012). Die Bereiche 1 («Lehren und Lernen») und 6 («Schulischer Auftrag 2») erweisen sich empirisch aber trotzdem als antagonistisch. Die Bereiche 5 («Schulischer Auftrag 1») und 7 («Leistung») hingegen teilen sich nicht in zwei gegensätzliche Pole auf.

Letztlich ergeben sich zwei gegensätzliche Bündel und es bietet sich an, dasjenige auf dem linken Halbkreis als «konservativ» und jenes auf dem rechten Halbkreis als «progressiv» zu bezeichnen. «Konservativ» und «progressiv» sind dabei als wertneutrale Labels zu verstehen; alternativ würden sich auch Begriffe wie «traditionalistisch, realistisch» versus «modern, idealistisch» anbieten (Fend, 2001, 2008). Die Facetten «Selbstbestimmung», «Sozialisaton», «Fördern» und «Umwelt» kommen auf dem rechten Halbkreis zu liegen, weshalb sie den progressiven Facetten zugerechnet werden. Die Facetten «Fremdbestimmung» und «Anlage» komplementieren die konservativen Facetten. Die Facetten «Situational» und «Dispositional» hingegen lassen sich keinem der beiden Bündel zuordnen.

Eine solche relationale Darstellung der einzelnen Facetten wird erst möglich, weil die Einschätzungen der einzelnen Facetten nicht isoliert, sondern in Beziehung zueinander erfasst werden. Für die Erstellung eines eigenen pädagogischen Profils müssen Personen webbasiert verschiedene pädagogische Aussagen (Items) auf einer Idealskala (Ryf & Läge, 2007) zwischen zwei Polen («ich stimme stark zu» – «ich stimme nicht zu») platzieren (vgl. Abbildung 2).

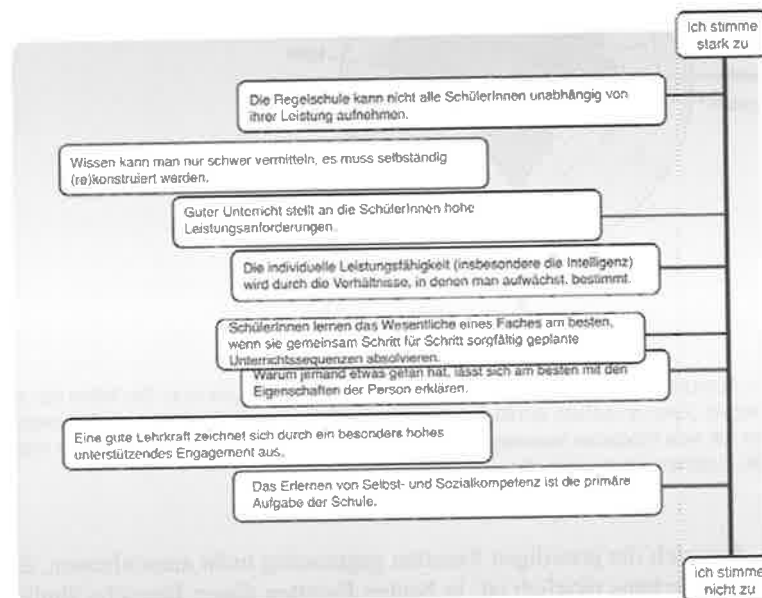


Abbildung 2: Idealskala mit den Polen «ich stimme stark zu» und «ich stimme nicht zu», zwischen welchen die verschiedenen pädagogischen Aussagen platziert werden können.

Insgesamt müssen 48 Aussagen eingeschätzt werden, welche in sechs Sets mit je acht Aussagen präsentiert werden. Die Personen werden also immer mit acht verschiedenen Aussagen passend zu den 16 Facetten (es tauchen nie zwei Items der gleichen Facette im gleichen Set auf) konfrontiert und müssen diese auf der Idealskala mittels Drag-and-drop am Computer in eine Rangfolge bringen. Dabei ist es ihnen überlassen, wie sie die Breite der zur Verfügung stehenden Idealskala nutzen. Die Idealskala wird ohne metrische Angaben präsentiert; effektiv entspricht sie aber einer Skala von 0 bis 100. Am Schluss der Erhebung erhalten die Personen ihr persönliches pädagogisches Bezugssystem in Form eines Spinnendiagramms, welches abgespeichert werden kann. Ebenfalls heruntergeladen werden können die Definitionen der einzelnen Facetten und eine Interpretationsanleitung für das eigene Profil. Die Erhebung ist online unter www.e-profiler.ch frei zugänglich.

Über die individuelle Auswertung hinaus lassen sich die verschiedenen pädagogischen Profile der einzelnen Personen einer Gruppe mittels Nonmetrischer Multidimensionaler Skalierung (NMDS) visualisieren und vergleichen (für einen Überblick zur Multidimensionalen Skalierung vgl. Borg, Groenen & Mair, 2010; für genauere Informationen zur NMDS vgl. McCombie & Läge, 2012). Bei einer NMDS werden die Profile anhand der Korrelationen über alle Paare miteinander verglichen. Daraus ergibt sich eine Proximitätsmatrix, welche die Basis dafür bildet, die Profile der verschiedenen Personen als Punkte in einem Raum abzubilden. Ziel dieses Verfahrens ist eine möglichst genaue Abbildung der Proximitäten durch Distanzen in einem geometrischen Raum. Dies stellt ein mathematisches Problem dar, für dessen Lösung ein aufwendiger Algorithmus¹ verwendet wird. Auf diese Weise entsteht eine pädagogische Landkarte (vgl. Abbildung 3). Für diese Karte gilt: Je ähnlicher die pädagogischen Profile von zwei Personen sind, desto näher beieinander kommen sie als Punkte in der Karte zu liegen. Eine Darstellung in einer NMDS-Karte erlaubt es, zu erkennen, wie homogen oder heterogen sich die Mitglieder einer Gruppe (z.B. ein Jahrgang von Studierenden, eine Modulgruppe oder auch Lehrpersonen eines Schulhauses) bezüglich ihrer Profile im Raum positionieren.

Mittels eines Property-Fitting-Verfahrens (Oberholzer, Egloff, Ryf & Läge, 2008) lässt sich der Raum auch semantisch beschreiben, d.h. es wird ersichtlich, welche Bereiche der Karte für welche Facetten stehen. Dazu werden mittels linearer Regression jene Facetten in die Karte gelegt, welche die Varianz zwischen den verschiedenen Profilen zu erklären vermögen. So wird ersichtlich, wie sich die verschiedenen Personen in der Karte inhaltlich gruppieren. Die Position der einzelnen Punkte der Karte auf einer Achse steht dann immer im rechten Winkel zu dieser Achse (vgl. Abbildung 3). Die anhand des Property-Fitting-Verfahrens in die Karte gelegten Facetten verlaufen durch den Schwerpunkt der Karte. Dieser bezeichnet den Durchschnittswert aller Personen in der Karte für die jeweilige Facette. Auf diese Weise lässt sich erkennen, welche Personen über oder unter dem Durchschnittswert der Facette liegen.

Der E-Profiler erstellt so auf individueller Ebene eine valide und reliable Diagnose der pädagogischen Überzeugungen von Studierenden. Die Reliabilität und die Validität wurden im Rahmen der Entwicklung des E-Profilers überprüft. Bezüglich der Reliabilität zeigt sich, dass, wenn Personen ihr Profil zweimal erstellen, diese beiden Profile miteinander übereinstimmen. Valide ist das Instrument in dem Sinne, dass das eigene Profil von den Personen als zu ihnen passend empfunden wird. Das heisst, wenn man Personen auffordert, ihr Profil in einer Auswahl von verschiedenen Profilen zu identifizieren, dann gelingt ihnen dies überzufällig (für die genaue Überprüfung der Validität und der Reliabilität des E-Profilers vgl. McCombie & Läge, 2012). Die impliziten Überzeugungen werden mit dem E-Profiler standardisiert erfasst und somit sichtbar und

¹ Die hier verwendete NMDS basiert auf dem ROBUSCAL-Algorithmus, welcher an der Universität Zürich entwickelt wurde (vgl. dazu Läge, 2001; Läge, Ryf, Daub, Jäger & Bosia, 2005; Läge, Ryf, Daub & Jäger, 2008).

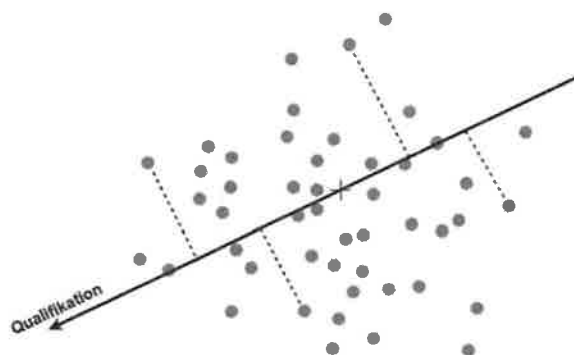


Abbildung 3: Beispielkarte einer NMDS-Karte mit einer mittels Property-Fitting hineingelegten Achse («Qualifikation»). Die Position der Punkte zur Achse steht jeweils im rechten Winkel. Lesebeispiel: Personen (Punkte), die entlang der Achse links vom Schwerpunkt liegen, weisen in der Facette «Qualifikation» einen höheren Wert auf als der Durchschnittswert aller Personen (Punkte) in dieser Facette.

vergleichbar. Die Analyse mittels NMDS erlaubt es darüber hinaus, mehrere Personen einer Gruppe zueinander in Beziehung zu setzen. Dies ermöglicht ein Gesamtbild der Gruppe, ohne dass der Blick auf das einzelne Profil verloren geht.

3 Empirische Daten

Exemplarisch werden zur Verdeutlichung des E-Profilers im Folgenden zwei empirische Untersuchungen geschildert. Es handelt sich dabei einerseits um den Vergleich der pädagogischen Bezugssysteme von Studierenden im ersten Semester und jenen von Studierenden im letzten Semester der Ausbildung an zwei pädagogischen Hochschulen und andererseits werden qualitative Daten berichtet, welche aus der Verwendung des E-Profilers im Rahmen der Ausbildung stammen.

3.1 Der Vergleich der pädagogischen Bezugssysteme der Studierenden an pädagogischen Hochschulen zu Beginn und am Ende der Ausbildung

Für den Vergleich von Studienanfängerinnen bzw. Studienanfängern und Studierenden im letzten Semester der Lehrpersonenausbildung haben Studierende von zwei pädagogischen Hochschulen (Zürich und St. Gallen) mit dem E-Profiler ihr pädagogisches Profil erstellt. Es wurde jeweils der ganze Jahrgang des ersten und des letzten Semesters via E-Mail angeschrieben und zur Teilnahme eingeladen. Total haben 169 Studienanfängerinnen und Studienanfänger sowie 152 Studierende im letzten Semester ihr Profil erstellt.

Als Erstes wurden mittels eines t-Tests für unabhängige Stichproben die Mittelwerte für die einzelnen Facetten der beiden Gruppen miteinander verglichen (vgl. Tabelle 2). Es zeigen sich dabei in den bipolaren Facetten «Selbstbestimmung – Fremdbestimmung» und «Nähe – Distanz» hochsignifikante Unterschiede. Weiter bestehen signifikante Unterschiede in den konservativen Facetten «Instruktion», «Qualifikation», «Selektion» und «Fordern» und in den progressiven Facetten «Sozialisation» und «Konstruktion». Es lässt sich beobachten, dass die Mittelwerte in den konservativen Facetten bei den Studierenden am Ende der Ausbildung tiefer und bei den progressiven Facetten höher sind als zu Beginn des Studiums. Die Ausnahme bildet dabei die Facette «Fordern», bei welcher die Studienanfängerinnen und Studienanfänger einen tieferen Wert aufweisen als die Studierenden im Abschlusssemester. Weiter zeigt sich, dass die Unterschiede in den nicht signifikanten Facetten jeweils nur gering sind. Die grössten Effektstärken finden sich in den Facetten «Instruktion – Konstruktion» und «Fremdbestimmung – Selbstbestimmung», die restlichen Effektstärken sind als mittel bis klein einzustufen (vgl. Tabelle 2).

Tabelle 2: Vergleich der Mittelwerte aller Facetten der Studierenden im ersten Semester und der Studierenden im Abschlusssemester

Facette	1. Semester (n = 169)		Abschluss- semester (n = 152)		Signifikanz	Effektstärke
	MW	SD	MW	SD		
Instruktion	57	13.9	46	13.2	**	0.82
Fremdbestimmung	56	15.8	46	14.0	**	0.67
Qualifikation	48	17.4	40	16.7	**	0.47
Distanz	29	12.1	24	11.2	**	0.43
Selektion	48	15.8	43	18.2	**	0.29
Dispositional	37	13.9	35	13.8		
Anlage	41	12.7	40	12.1		
Integration	50	18.9	54	19.5		
Fördern	77	12.3	78	12.6		
Situational	63	13.9	65	13.8		
Umwelt	59	12.7	60	12.0		
Fordern	40	14.5	47	15.6	**	-0.43
Nähe	71	12.1	76	11.2	**	-0.51
Sozialisation	58	17.3	68	16.5	**	-0.59
Selbstbestimmung	44	15.7	54	14.0	**	-0.67
Konstruktion	54	14.5	64	15.4	**	-0.67

* $p < .05$; ** $p < .01$

Die Abbildung der Profile in einer NMDS-Karte bestätigt und erweitert dieses Bild. Es fällt als Erstes auf, dass die Studienanfängerinnen und Studienanfänger sich breiter im Raum verteilen, also in ihren Profilen heterogener sind (vgl. Abbildung 4). Die Studierenden im Abschlusssemester gruppieren sich mehrheitlich im rechten oberen Bereich der Karte. Die Verteilung der Studierenden im ersten Semester hingegen streut deutlich stärker gegen links unten in der Karte.

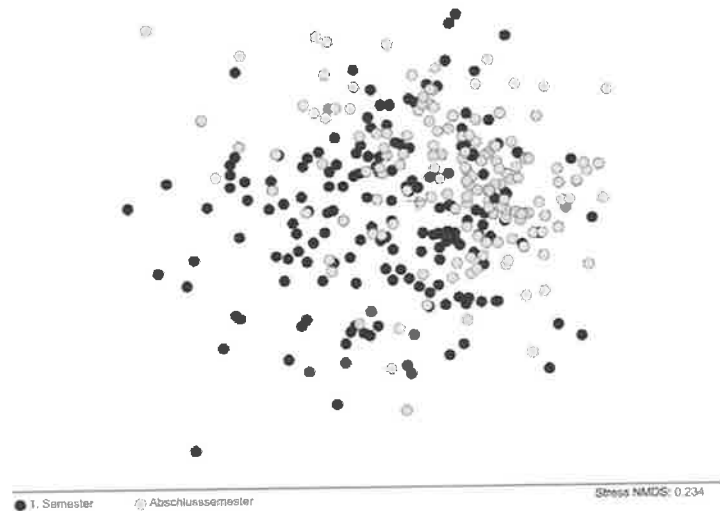


Abbildung 4: NMDS-Karte der Studierenden im Abschlusssemester (grau) und der Studierenden im ersten Semester (schwarz).

Legt man nun mit Property-Fitting die zwischen den Profilen in der Karte differenzierenden Facetten als Achsen in die Karte, so zeigt sich, dass dieser linke untere Bereich der Karte mit den konservativen Facetten («Fremdbestimmung», «Distanz», «Instruktion» und «Qualifikation») assoziiert werden kann (vgl. Abbildung 5), während der rechte obere Bereich für die progressiven Facetten «Sozialisation», «Selbstbestimmung», «Nähe» und «Konstruktion» steht. In der Diagonale von links oben nach rechts unten differieren die Profile hingegen entlang der Facetten «Integration» und «Selektion», wobei hier keine unterschiedliche Aufteilung nach Studierenden im ersten Semester und jenen im letzten Semester zu erkennen ist.

Es zeigt sich also, dass zwischen der Zeit als Studienanfängerin oder Studienanfänger und dem Ende der Ausbildung eine Verschiebung hin zu den progressiveren Facetten stattfindet. Die Studierenden sind am Ende ihrer Ausbildung als Gruppe zudem homogener, da sie eine Clusterdichte² von 0.872 aufweisen. Die Clusterdichte der Stu-

² Die Clusterdichte in der Karte entspricht der mittleren Distanz innerhalb eines Clusters geteilt durch die mittlere Distanz aller Punkte in der Karte. Ein Wert von <1 weist deshalb darauf hin, dass die Punkte des Clusters enger zusammenliegen als zufällige Punkte in der Karte.

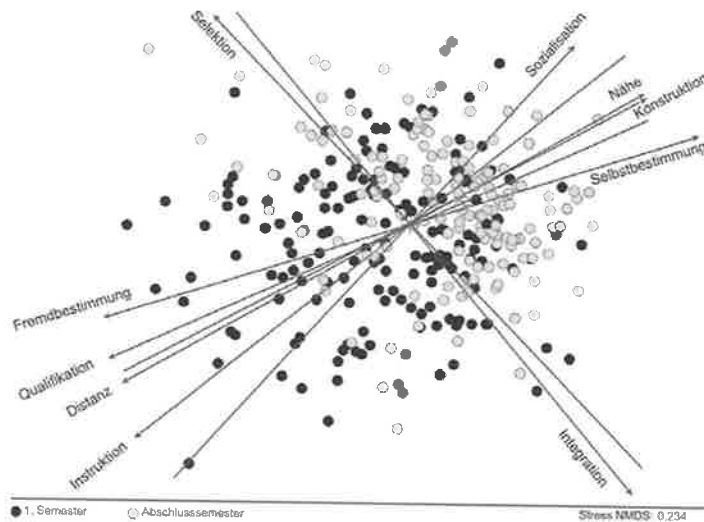


Abbildung 5: NMDS-Karte der Studierenden im Abschlusssemester (grau) und der Studierenden im ersten Semester (schwarz) mit Property-Fitting.

dierenden im ersten Semester beträgt demgegenüber 1.0366. Für die Facetten «Selektion» und «Integration» lässt sich beobachten, dass eine Trennung von Studierenden im ersten Semester und im letzten Semester nicht gelingt und die Streuung in beiden Gruppen in etwa gleich hoch bleibt.

Dieser kleine Einblick in die empirischen Daten zeigt, dass zwischen Studierenden im ersten und jenen im letzten Semester der Lehrpersonenausbildung in der Ausprägung ihrer pädagogischen Bezugssysteme signifikante Unterschiede bestehen. Die Studierenden scheinen im Verlauf ihrer Ausbildung progressiver zu werden. Für die Lehrerinnen- und Lehrerbildung ist dies ein erfreulicher Befund, da ein Einfluss auf die pädagogischen Grundhaltungen der Studierenden zu bestehen scheint und dies darüber hinaus in die intendierte Richtung. So scheint ein eher instruktionales Verständnis von Lehr- und Lernprozessen einem stärker konstruktivistischen zu weichen und die Selbstbestimmung der Schülerinnen und Schüler wird am Ende der Ausbildung höher gewichtet, während der Fremdbestimmungswert den Selbstbestimmungswert zu Beginn des Studiums noch übersteigt.

Der Blick auf die absoluten Werte weist aber auch darauf hin, dass sich die Verschiebungen nur im mittleren Bereich der Skalen bewegen und man keine extremen Veränderungen beobachten kann. Die Darstellung in der NMDS-Karte macht ebenfalls auf diesen Punkt aufmerksam. Obwohl die Clusterdichte für die Studierenden im Abschlusssemester auf eine homogene Gruppe hinweist, sind die Gruppen weit davon ent-

fernt, zwei getrennte Cluster zu bilden. Es gibt viele Überschneidungen und in beiden Gruppen sind sowohl konservative wie auch progressive Profile zu finden.

Weiter lassen die vorliegenden Querschnittsdaten keine Rückschlüsse über tatsächliche Entwicklungsverläufe zu. Ob sich das pädagogische Bezugssystem im Einzelfall durch die Ausbildung effektiv verändert, lässt sich nicht sagen. Die vorliegenden Tendenzen sind deshalb in einem echten Längsschnitt zu überprüfen.

3.2 Verwendung des E-Profilers in der Ausbildung

In einem Modul im siebten Semester der Sekundarlehrpersonenausbildung an der Pädagogischen Hochschule St. Gallen (PHSG) wurde der E-Profiler als Teil eines Leistungsnachweises eingesetzt. Die Studierenden mussten ihr eigenes pädagogisches Profil zu Beginn des Semesters erstellen und bis zum Ende des Semesters schriftlich reflektieren. Genauere Ausführungen zum Instrument und zur Interpretation des eigenen Profils erfolgten im Rahmen des Moduls. Die Studierenden mussten vier Leitfragen zu ihrem Profil beantworten: 1) Was löst das Profil bei Ihnen aus? 2) Gibt es einzelne Facettenwerte, welche Sie als eher zu hoch oder zu tief einstufen? 3) Wie stehen Sie zum Durchschnittsprofil des Jahrgangs? 4) Welche persönlichen Überzeugungen (Facettenwerte) werden Sie in der Praxis leicht verfolgen können und bei welchen wird Ihre Überzeugung wohl eher getestet?

Um der Frage nachzugehen, wie die Studierenden das eigene pädagogische Profil wahrnehmen und einschätzen, wurden nur die ersten beiden Fragen berücksichtigt. Von insgesamt 56 Studierenden haben 46 die ersten beiden Fragen explizit beantwortet, weshalb nur die Antworten dieser Studierenden ausgewertet wurden. Die qualitative, zusammenfassende Inhaltsanalyse (Bortz & Döring, 2006; Mayring, 2010) zeigt, dass die grosse Mehrheit der Studierenden ihr Profil als passend empfindet, d.h. sie nehmen es als repräsentativ für ihre eigenen pädagogischen Überzeugungen wahr. Nur gerade drei Studierende waren mit ihrem Profil grundsätzlich nicht einverstanden. Bei diesen Studierenden zeigt sich gleichzeitig eine allgemeine Skepsis gegenüber dem Verfahren. Auch wenn die meisten Studierenden ihr Profil als stimmig wahrnehmen, so sind einige beim ersten Betrachten doch überrascht oder gar etwas erschrocken. Sie empfinden das eigene Profil erst auf den zweiten Blick als mehrheitlich stimmig. So führt eine Studentin aus:

Auf den ersten Blick empfand ich mein Profil sehr einseitig. Ich war zu Beginn fast ein wenig schockiert über dieses Ungleichgewicht in meiner pädagogischen Einstellung. Als ich es dann etwas genauer zu studieren begann, merkte ich rasch, dass es schon ziemlich der Tatsache entspricht und meine tatsächlichen pädagogischen Ansichten mehr oder weniger wahrheitsgetreu darstellt.

Es wird ebenfalls angemerkt, dass die Erstellung des Profils nicht einfach war und man sich schwergetan hat bei der Platzierung der verschiedenen Aussagen. Trotz der Schwierigkeiten erfahren diese Studierenden ihr Profil aber als überraschend passend. Folgende Aussage einer Studentin verdeutlicht dies:

Obwohl ich zum Teil das Ausfüllen der Umfrage als sehr schwierig empfand, mich nicht eindeutig für eine Reihenfolge entscheiden konnte und manchmal auch nicht ganz sicher war, dass ich die Aussagen richtig verstehe, finde ich, dass die Grafik meinen Charakter als Lehrperson sehr gut widerspiegelt.

Aber auch Studierende, welche ihr eigenes Profil wenig überrascht, betonen den Nutzen, den das Profil für sie haben kann. Die folgende Aussage eines Studenten zeigt dies sehr schön:

Ich fand es interessant, mein eigenes pädagogisches Profil zu erstellen und es in Form eines Spiders vor mir zu haben. Natürlich wusste ich schon vorher, wie meine Einstellungen in gewissen Bereichen ungefähr sind, aber es war doch nochmals etwas anderes, diese recht umfangreiche Darstellung zu studieren.

Die meisten Studierenden finden jedoch auch Facettenwerte, mit denen sie nicht übereinstimmen, weil sie ihnen zu hoch oder zu tief erscheinen. Insbesondere in den Facetten «Selbstbestimmung – Fremdbestimmung» (13 Nennungen), «Instruktion – Konstruktion» (11 Nennungen), «Integration – Selektion» (11 Nennungen) und «Nähe – Distanz» (10 Nennungen) werden die erzielten Werte infrage gestellt. Aber auch «Qualifikation – Sozialisation» (8 Nennungen) und «Fördern – Fördern» (6 Nennungen) werden vermehrt genannt. Dabei lässt sich erkennen, dass man sich scheinbar an gewissen Normvorstellungen orientiert, d.h. dass man das Gefühl hat, dass die eigenen Werte gewissen Ansprüchen nicht entsprechen. Folgende Aussage weist in diese Richtung:

Auffallend sind die Facetten Integration/Selektion, Fremdbestimmung/Selbstbestimmung, Sozialisation/Qualifikation sowie die eher tiefere Einstufung bezüglich Fördern und Fordern. Mir scheint, als würden diese Werte nicht unbedingt eine ideale Lehrperson darstellen (wie sie von der Gesellschaft/Wirtschaft erwartet wird), weshalb ich versucht wäre, diese Werte etwas zu «neutralisieren».

Die Analyse der Aussagen der Studierenden zu ihren Profilen zeigt, dass das eigene Profil grundsätzlich als passend empfunden wird, es also gelingt, die individuellen pädagogischen Bezugssysteme valide zu diagnostizieren. Dabei ist aber zu beachten, dass die meisten Studierenden doch zumindest in einigen Facetten mit den erzielten Werten nicht einverstanden sind. Gerade eine gewisse Einseitigkeit in bestimmten Facetten zwingt die Studierenden zum Nachdenken. Dabei wird zum Teil auch ersichtlich, dass man sich bei bestimmten Werten an einem impliziten Ideal orientiert. So sollte eine Lehrperson anscheinend auch genügend fordern und eine gewisse Distanz zu den Schülerinnen und Schülern einnehmen. Auch scheint Selbstbestimmung positiv bewertet zu sein, da sich einige Studierende für die in ihren Augen zu tiefen Werte auf dieser Facette rechtfertigen.

Allgemein lässt sich zusammenfassen, dass die Erstellung eines eigenen Profils zwar nur wenige Überraschungen bereithält, gleichwohl aber zur Auseinandersetzung mit den eigenen Überzeugungen anregt und gerade auch im Vergleich mit den Profilen anderer Personen ein sinnvolles Reflexionsinstrument darstellen kann.

4 Diskussion

Der Einblick in die empirischen Daten zeigt das Potenzial des E-Profilers auf. Auf der individuellen Ebene macht der E-Profiler die zuvor nur implizit vorhandenen Überzeugungen für die Person selbst, aber auch für andere Personen sichtbar und somit diskutierbar. Als diagnostisches Instrument leistet der E-Profiler drei wesentliche Dinge. Erstens trägt er dazu bei, Transparenz zu schaffen. Das eigene pädagogische Bezugssystem wird besser erkennbar und fassbar. Zweitens wird durch die vorgegebenen Begrifflichkeiten eine gemeinsame Sprache geschaffen, die die Kommunikation über die subjektiven Bezugssysteme ermöglicht. Und drittens können die Profile miteinander verglichen werden, indem sie auf einer mathematisch-quantitativen Ebene über Korrelationen und die Darstellung in einer NMDS-Karte miteinander in Beziehung gesetzt werden. So gelingt es, die subjektiven Bezugssysteme durch eine standardisierte Erfassung zu vergleichen und sich damit auseinanderzusetzen. Die Auswertung auf der Gruppenebene (via Durchschnittsprofil und NMDS-Karte) ermöglicht es der Einzelnen und dem Einzelnen, sich innerhalb einer Gruppe zu positionieren und das eigene Profil sozial zu validieren.

Es gilt dabei jedoch zu beachten, dass es schwierig ist, die Validität des E-Profilers absolut zu belegen. Ein pädagogisches Bezugssystem ist nicht abschliessend explizierbar; es sind mehrere Wege möglich, um zu einem sinnvollen Set von Facetten zur Beschreibung des Bezugssystems zu kommen. Letztlich handelt es sich um den Versuch, anhand der ausgewählten Bereiche möglichst das ganze Blickfeld abzudecken und keinen blinden Fleck zuzulassen (für die Herleitung der Facetten vgl. McCombie & Läge, 2012). Theoretisch lehnt sich der E-Profiler zwar an ältere Untersuchungen an, welche dem Versuch, ein pädagogisches Bezugssystem zu messen, am nächsten kommen (Cook, Leeds & Callis, 1951; Dann et al., 1978; Tanner, 1993), geht aber inhaltlich darüber hinaus. Gleichzeitig unterscheidet er sich von aktuellen Untersuchungen, die stärker konkrete Überzeugungen, insbesondere die epistemologischen Überzeugungen erfassen (z.B. Biedermann, Brühwiler & Krattenmacher, 2012; Blömeke et al., 2008; Hartinger, Kleickmann & Hawelka, 2006; Leuchter, 2009; Renkl & Stern, 1994; Staub & Stern, 2002; Voss, Kleickmann, Kunter & Hachfeld, 2011). Tatsächlich liegen keine anderen Instrumente vor, anhand welcher sich die Kriteriumsvalidität des E-Profilers belegen liesse. Für die Überprüfung der Validität musste deshalb ein anderer Weg gefunden werden. Dazu wurde Personen im Rahmen der Testkonstruktion eine Auswahl an Profilen in einer NMDS-Karte dargeboten. Die Personen mussten dann dasjenige Profil auswählen, welches sie für ihr eigenes hielten. Es zeigte sich dabei, dass eine grosse Mehrheit ihr eigenes Profil identifizieren konnte (McCombie & Läge, 2012). Dies macht deutlich, dass der E-Profiler die pädagogischen Bezugssysteme insofern valide misst, als die subjektiven Bezugssysteme so erfasst werden, wie sie von den Personen wahrgenommen werden. Die Befragung von Studierenden in der vorliegenden Untersuchung verdeutlicht dies: Die Studierenden empfinden ihre eigenen Profile als stimmig.

Es zeigt sich, dass der E-Profiler für die Lehrerinnen- und Lehrerbildung ein wertvolles Entwicklungs- und Reflexionsinstrument darstellen kann. Das eigene Profil und das Durchschnittsprofil der Gruppe können den Studierenden als Ansatzpunkte dafür dienen, die eigenen Überzeugungen zu reflektieren. Insbesondere der Vergleich mit anderen Studierenden kann dabei helfen, die eigene Position kritisch zu hinterfragen. Zu diesem Zweck kann auch die Abbildung der Profile in einer pädagogischen Landkarte hilfreich sein. Diese Darstellung lässt die Studierenden erkennen, dass sie sich in ihren Profilen zum Teil deutlich unterscheiden, was die Relativität der eigenen Position aufzeigen kann. Neben dem Nutzen für die persönliche Reflexion ist es auch denkbar, dass Dozierende bei der Vermittlung theoretischer Inhalte direkter auf vorhandene subjektive Überzeugungen der Studierenden Bezug nehmen und den Studierenden auf diese Weise die Verknüpfung von subjektiver und wissenschaftlicher Theorie besser gelingt. Natürlich reicht eine Diagnose der subjektiven Überzeugungen noch nicht aus, um diese auch zu verändern oder zu entwickeln, doch ein erster wichtiger Schritt wird durch die Diagnose der Überzeugungen erreicht. Denn nur an explizierten Überzeugungen kann auch gearbeitet werden.

Über seine Funktion als Entwicklungsinstrument hinaus ist der E-Profiler auch als Forschungsinstrument interessant, da er es erlaubt, die subjektiven Bezugssysteme von grösseren Stichproben standardisiert zu erheben. So wird es möglich, die pädagogischen Bezugssysteme von verschiedenen Gruppen zu vergleichen und beispielsweise zu sehen, ob sich die Überzeugungen von Studierenden zu Beginn und am Ende der Ausbildung unterscheiden. Aber auch Vergleiche von angehenden und berufstätigen Lehrpersonen oder der Vergleich verschiedener Ausbildungsgänge oder Schulstufen sind denkbar. Solche quantitativen Untersuchungen geben einen Einblick in die subjektiven Überzeugungen der Studierenden und können als Anhaltspunkte für die Lehrerinnen- und Lehrerbildung dienen. So erlaubt zum Beispiel der hier vorgestellte Vergleich von Studienanfängerinnen bzw. Studienanfängern und Studierenden am Ende der Ausbildung eine Aussage darüber, mit welchen Vorstellungen Studierende in die Ausbildung eintreten, was wiederum wichtige Ansatzpunkte für die Vermittlung von Lerninhalten liefern kann. Es zeigt sich weiter, dass sich die pädagogischen Bezugssysteme im Verlauf der Ausbildung in wesentlichen Facetten zu verändern scheinen und die Studierenden am Ende ihrer Ausbildung tendenziell über progressivere Bezugssysteme verfügen. Dieser zentrale Befund deckt sich mit älteren Forschungsergebnissen (Dann et al., 1978; Tanner, 1993) und steht auch im Einklang mit neueren Studien (z.B. Blömeke et al., 2008). Dabei sind aber die Unterschiede zwischen den Studierenden zu Beginn und am Ende des Studiums absolut gesehen eher klein und es stellt sich die Frage, ob die Ausbildung im gewünschten Masse Einfluss nehmen kann oder dies verstärkt tun sollte. Ob sich die pädagogischen Bezugssysteme über die Ausbildungszeit hinweg tatsächlich verändern, lässt sich anhand der vorliegenden Querschnittsdaten nicht sagen; dies müsste in einer echten Längsschnittstudie verifiziert werden.

Zum Schluss lässt sich festhalten, dass der E-Profiler ein Instrument darstellt, welches zur Unterstützung von erfolgreichen Lehr- und Lernprozessen in der Lehrerinnen- und

Lehrerbildung von Nutzen sein kann, da er es erlaubt, die subjektiven Überzeugungen von Lehrpersonen aktiv und direkt einzubeziehen. Dadurch wird für Dozierende wie auch für Studierende ein Anknüpfen an den subjektiven Überzeugungen möglich, was für den Lehr- und Lernprozess nur förderlich sein kann.

Literatur

- Ausubel, D.P., Novak, J.D. & Hanesian, H. (1978). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.
- Biedermann, H., Brühwiler, C. & Krattenmacher, S. (2012). Lernangebote in der Lehrerausbildung und Überzeugungen zum Lehren und Lernen. *Zeitschrift für Pädagogik*, 58 (4), 460–475.
- Bischof, N. (1996). *Das Kraftfeld der Mythen*. München: Piper.
- Bischof, N. (2009). *Psychologie. Ein Grundkurs für Anspruchsvolle*. Stuttgart: Kohlhammer.
- Blömeke, S. (2005). *Lehrerausbildung – Lehrerhandeln – Schülerleistungen: Perspektiven nationaler und internationaler empirischer Bildungsforschung* (Antrittsvorlesung, 10. Dezember 2003). Online verfügbar unter: <http://edoc.hu-berlin.de/humboldt-vl/139/bloemeke-sigrid-3/PDF/bloemeke.pdf> (03.12.2012).
- Blömeke, S., Müller, S., Felbrich, A. & Kaiser, G. (2008). Epistemologische Überzeugungen zur Mathematik. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer* (S. 219–246). Münster: Waxmann.
- Borg, I., Groenen, P.J.F. & Mair, P. (2010). *Multidimensionale Skalierung*. München: Rainer Hampp.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Heidelberg: Springer.
- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D.C. Berliner & R.C. Calfee (Hrsg.), *Handbook of educational psychology* (S. 709–725). New York: Macmillan.
- Cook, W.W., Leeds, C.H. & Callis, R. (1951). *Minnesota Teacher Attitude Inventory*. New York: The Psychological Corporation.
- Dann, H.-D., Cloetta, B., Müller-Fohrbrodt, G. & Helmreich, R. (1978). *Umweltbedingungen innovativer Kompetenz*. Stuttgart: Klett-Cotta.
- Dochy, R.S. (1992). *Assessment of prior knowledge as a determinant for future learning: The use of prior knowledge state tests and knowledge profiles*. London: Kingsley.
- Fang, Z. (1996). A Review of Research on Teacher Beliefs and Practice. *Educational Research*, 38 (1), 47–65.
- Fend, H. (2001). *Qualität im Bildungswesen*. Weinheim: Juventa.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartinger, A., Kleickmann, T. & Hawelka, B. (2006). Der Einfluss von Lehrervorstellungen zum Lernen und Lehren auf die Gestaltung des Unterrichts und auf motivationale Schülervariablen. *Zeitschrift für Erziehungswissenschaft*, 9 (1), 110–126.
- Kane, R., Sandretto, S. & Heath, C. (2002). Telling Half the Story: A Critical Review of Research on the Teaching Beliefs and Practices of University Academics. *Review of Educational Research*, 72 (2), 177–228.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift. Zürich: Philosophische Fakultät der Universität Zürich.
- Läge, D., Ryf, S., Daub, S. & Jäger, C. (2008). *Robustness in nonmetric multidimensional scaling* (AKZ-Forschungsbericht Nr. 59). Zürich: Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie.
- Läge, D., Ryf, S., Daub, S., Jäger, C. & Bosia, L. (2005). *Die Behandlung ausreisserbehafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag* (AKZ-Forschungsbericht Nr. 21). Zürich: Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie.
- Leuchter, M. (2009). *Die Rolle der Lehrperson bei der Aufgabenbearbeitung. Unterrichtsbezogene Kognitionen von Lehrpersonen*. Münster: Waxmann.

- Mayring, P.** (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Murck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 601–613). Wiesbaden: VS Verlag für Sozialwissenschaften.
- McCombie, G. & Läge, D.** (2012). *Educational Profiler. Konstruktion eines Messverfahrens für den Vergleich von pädagogischen Bezugssystemen mittels Idealskala und Nonmetrischer Multidimensionaler Skalierung*. Zürich: Verlag Forschung & Lehre. Online verfügbar unter: http://www.mccombieconsulting.com/multimedia/E-Profiler_ebook2012.pdf (09.08.2013).
- Müsseler, J.** (2007). *Allgemeine Psychologie*. Berlin: Spektrum.
- Nespor, J.** (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19 (4), 317–328.
- Oberholzer, R., Egloff, S., Ryf, S. & Läge, D.** (2008). *Datenauswertung im Bereich der Skalierung. Benutzerhandbuch*. Zürich: Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie. Online verfügbar unter: www.prodax.ch/ProDaX-Handbuch.pdf (09.08.2013).
- Pajares, M.F.** (1992). Teachers' beliefs and educational research. *Review of Educational Research*, 62 (3), 307–332.
- Renkl, A. & Stern, E.** (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und schulischen Lerngelegenheiten für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für Pädagogische Psychologie*, 8 (1), 27–39.
- Reusser, K., Pauli, C. & Elmer, E.** (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 478–495). Münster: Waxmann.
- Ryf, S. & Läge, D.** (2007). *Die Idealskala – ein Messwerkzeug zur optimierten Erhebung von Präferenzen* (AKZ-Forschungsbericht Nr. 57). Zürich: Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie.
- Schraw, G.** (2006). Knowledge: Structures and processes. In P.A. Alexander & P.H. Winne (Hrsg.), *Handbook of educational psychology* (S. 245–263). Mahwah, NJ: Erlbaum.
- Staub, F.C. & Stern, E.** (2002). The Nature of Teachers' Pedagogical Content Beliefs Matters for Students' Achievement Gains: Quasi-Experimental Evidence From Elementary Mathematics. *Journal of Educational Psychology*, 94 (2), 344–355.
- Tanner, H.** (1993). *Einstellungsänderungen während der Lehrerausbildung und Berufseinführung*. Weinheim: Deutscher Studienverlag.
- Voss, T., Kleickmann, T., Kunter, M. & Hachfeld, A.** (2011). Überzeugungen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften* (S. 235–258). Münster: Waxmann.
- Woolfolk Hoy, A., Davis, H. & Pape, S.J.** (2006). Teacher knowledge and beliefs. In P.A. Alexander & P.H. Winne (Hrsg.), *Handbook of educational psychology* (S. 715–737). Mahwah, NJ: Erlbaum.

Autoren

- Guido McCombie**, Prof. Dr., Pädagogische Hochschule St. Gallen, guido.mccombie@phsg.ch
Damian Läge, Prof. Dr., Universität Zürich, Psychologisches Institut, Allgemeine Psychologie (Kognition), Arbeitsgruppe Angewandte Kognitionspsychologie, d.laenge@psychologie.uzh.ch

Unterrichtsdiagnostik als Voraussetzung für Unterrichts-entwicklung*

Andreas Helmke und Gerlinde Lenske

Zusammenfassung Wenn man Hatties Forderung, das Lernen sichtbar zu machen, auf die Prozesse des Lehrens erweitert, dann ergibt sich die Notwendigkeit einer empirischen Standortbestimmung des Unterrichts. Hierfür wurde das Konzept der Unterrichtsdiagnostik entwickelt, dessen Potenzial am Beispiel des Diagnosewerkzeugs EMU (*Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung*) veranschaulicht wird. Im Kern geht es dabei um die datenbasierte und kriteriengeleitete Beschreibung des Unterrichts und um den Abgleich unterschiedlicher Perspektiven: Selbsteinschätzung der unterrichtenden Lehrperson, Feedback der hospitierenden Kollegin oder des hospitierenden Kollegen und Schülerfeedback. Die Grundgedanken und die dafür entwickelten Instrumente und Prozeduren lassen sich sowohl für die Lehrpersonenausbildung als auch für die Lehrpersonenfortbildung und -professionalisierung nutzen.

Schlagwörter Unterrichtsdiagnostik – Unterrichtsentwicklung – Feedback – Unterrichtsqualität – diagnostische Kompetenz – Unterrichtsbeobachtung – Perspektivenwechsel

Diagnosis of Classroom Instruction from Different Perspectives as a Prerequisite for Improving Teaching and Learning

Abstract Starting out from Hattie's focus on making student learning visible, we argue that this is also crucial to teaching processes. For this purpose, we have developed a system of questionnaires, software and other materials which helps teachers to gather an empirical data base in order to monitor and improve their instruction. Based on modern research on teaching effectiveness, and on a concrete lesson, the core of the instrument consists of tools for a systematic comparison of different perspectives: the teacher's self-report, rating of an observing colleague, and students' feedback.

Keywords quality of instruction – diagnostic competence – feedback – classroom observation – change of perspectives

1 Einführung

Ob John Hattie mit seiner monumentalen Studie «Visible Learning» (Hattie, 2009), die in der Zwischenzeit auch in deutscher Sprache erschienen ist (Hattie, 2013), tatsächlich «teaching's holy grail» – also den Heiligen Gral – gefunden hat, wie vom «Times Educational Supplement» berichtet, mag dahingestellt bleiben. Auf jeden Fall

* Wir danken Friedrich-Wilhelm Schrader für sehr konstruktive Korrekturvorschläge zu diesem Text.

hat diese Studie etwas bewirkt, was weder die grossen Large-Scale-Studien (wie PISA, IGLU, DESI und TIMSS) noch zahlreiche andere wissenschaftliche Bücher, Berichte, Foren, Appelle und Schriften geschafft haben: Den Blick der Öffentlichkeit, insbesondere auch der Bildungspolitik und der Lehrpersonenaus- und -fortbildung, wieder auf das Kerngeschäft der Schule, nämlich auf den Unterricht zu lenken und damit auf die wichtigsten Protagonistinnen und Protagonisten in diesem Geschäft, die Lehrerinnen und Lehrer.

Eines der grundlegenden Postulate der Hattie-Studie besteht in der Notwendigkeit, das pädagogische Tun auf eine *empirische Grundlage* zu stellen, d.h. das Handeln auf Daten (empirische Evidenz) zu stützen. Hatties Hauptaugenmerk ist darauf gerichtet, das Lernen sichtbar zu machen, indem vielfältige Situationen geschaffen werden, in denen Lehrpersonen Einblick in den laufenden Lernprozess ihrer Schülerinnen und Schüler nehmen, einschliesslich möglicher Lernhindernisse, Missverständnisse und Fehler. Ein zentraler Faktor ist dabei das *Feedback*, also die Rückmeldung über den aktuellen Stand. Bemerkenswerterweise sieht Hattie Feedback jedoch nicht als Einbahnstrasse, obwohl er ursprünglich eine andere Auffassung vertreten hatte:

Der Fehler, den ich machte, war, in Feedback etwas zu sehen, was die Lehrpersonen den Lernenden geben. Das haben sie meist nicht getan, auch wenn sie behaupteten, dass sie es die ganze Zeit tun würden. Das meiste Feedback, das sie gaben, war sozialer und verhaltensbezogener Art. Erst als ich entdeckt habe, dass Feedback besonders wirksam ist, wenn es der Lehrperson von den Lernenden gegeben wird, begann ich es besser zu verstehen. Wenn Lehrpersonen Feedback von den Lernenden einfordern – oder zumindest offen sind gegenüber dem, was Lernende wissen, was sie verstehen, wo sie Fehler machen, wo sie falsche Vorstellungen haben, wo es ihnen an Engagement mangelt – dann können Lehren und Lernen miteinander synchronisiert werden und wirksam sein. Feedback an die Lehrperson hilft, das Lernen sichtbar zu machen. (Hattie, 2013, S. 206)

Denkt man diesen Ansatz weiter, dann ist es nur konsequent, in analoger Weise zur Sichtbarmachung des *Lernens* auch das *Lehren*, also den Unterricht, sichtbar zu machen. Diese Ausweitung spricht Hattie übrigens explizit an: «Der wichtigste Aspekt besteht darin, im Klassenzimmer Situationen zu schaffen, in denen die Lehrpersonen mehr Feedback über ihren Unterrichtsstil erhalten können» (Hattie, 2013, S. 15). Hierzu bedarf es geeigneter Werkzeuge und Methoden – und genau darum geht es in diesem Beitrag.

2 Unterrichtsdiagnostik

2.1 Theoretische Einbettung

Diagnostische Kompetenzen werden von der Pädagogischen Psychologie (siehe Schrader 2011, 2013 [in diesem Heft]) wie von der Bildungspolitik seit jeher für wichtig, ja unabdingbar gehalten. Für Weinert (Weinert & Helmke, 1996) ist diagnostische Kompetenz – neben didaktischer Kompetenz, fachlicher Kompetenz und Klassenführungskompetenz – eine der vier Kernkompetenzen des Lehrberufs. Ähnlich hoch

wird die diagnostische Kompetenz in der Bildungspolitik eingeschätzt (vgl. die Professionsstandards der Kultusministerkonferenz, 2004); hier lauten die vier zentralen Kompetenzbereiche Unterrichten, Erziehen, Beurteilen und Innovieren. Allerdings ist mit «Diagnostik» in diesem Zusammenhang meist Individualdiagnostik gemeint. Wie Abbildung 1 zeigt, kann sich Diagnostik jedoch auf sehr unterschiedliche Sachverhalte und Gegenstände beziehen: von Schülerleistungen und -kompetenzen über andere lernrelevante Schülermerkmale, beobachtbares Lern- und Aufmerksamkeitsverhalten (z.B. Helmke & Renkl, 1992) bis hin zu Merkmalen der Lehrpersonenpersönlichkeit (Herlt & Schaarschmitt, 2007) und Lehrpersonenprofessionalität (z.B. in der Studie COACTIV, Kunter et al., 2011). Darüber hinaus kann Diagnostik auch auf die Qualität der unterrichtlichen Prozesse bezogen sein.

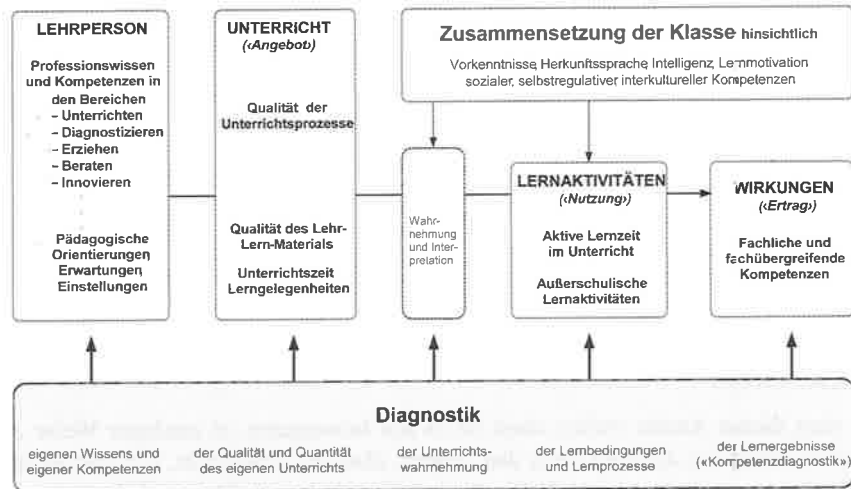


Abbildung 1: Gegenstandsbereiche der pädagogischen Diagnostik (Quelle: www.unterrichtsdiagnostik.info).

Folgt man dem Konzept einer Evidenzbasierung pädagogischer Praxis, wie es – keineswegs nur, aber besonders prononciert – von Hattie vertreten wird, dann setzt eine gezielte Weiterentwicklung des Unterrichts eine geeignete Standortbestimmung, also «Daten», voraus. Hierzu gibt es in der Forschung und in der Schulpraxis vielfältiges Material, das in zunehmendem Masse auch im Internet zugänglich ist. Für die gezielte Nutzung empirischer Informationen über Unterricht entwickelte das Landauer Forschungsteam (vgl. www.unterrichtsdiagnostik.info/media/files/Autorenteam.pdf) das Konzept «Unterrichtsdiagnostik». Während man bei «Diagnostik» im alltäglichen Sprachgebrauch häufig zunächst an die Medizin denkt, ist doch festzuhalten, dass dieses aus dem Griechischen stammende Wort ursprünglich keinen bereichsspezifischen Charakter hatte, sondern nur die Erforschung eines Sachverhaltes meint mit dem Ziel, be-

obachtete Merkmale einem Klassifikationssystem zuzuordnen; wer dies kann, ist ein *diagnostikos* («zum Unterscheiden begabt»).

Ähnlich wie beim Konzept des «reflective practitioner» bei Schön (1983), des «forschenden Lehrers» bei Altrichter und Posch (2007) und der «unterrichtsintegrierten Selbstevaluation» (Beywl, 2013) geht es auch bei der Unterrichtsdiagnostik darum, sich in forschender Absicht dem eigenen Unterricht zu nähern. Zusätzlich beinhaltet das im Folgenden vorgestellte Konzept der Unterrichtsdiagnostik den datenbasierten, systematischen und expliziten Abgleich unterschiedlicher Perspektiven unter Verwendung kriteriengeleiteter Werkzeuge (vgl. Abbildung 2). Dies entspricht der Sichtweise von Hattie (2013, S. 281): «Begegnungen unter Lehrpersonen: Hier diskutieren, bewerten und planen sie ihren Unterricht im Licht der Feedback-Evidenz: über den Erfolg und die weiteren Wirkungen ihrer Lehrstrategien und Konzepte, über Fortschritt und angemessene Herausforderungen. Dies ist nicht (nur) kritische Reflexion, sondern kritische Reflexion *im Licht der Evidenz, also im Licht empirischer Belege* zu ihrem Unterricht.»



Abbildung 2: Datenbasiertes Feedbackgespräch im Anschluss an eine Unterrichtsstunde (Quelle: www.unterrichtsdiagnostik.info/video/).

2.2 Werkzeuge und Prozeduren

Die Unterrichtsdiagnostik lässt eine grosse Bandbreite von Methoden, Werkzeugen und Perspektiven zu (für eine ausführliche Übersicht vgl. Helmke, Helmke & Pham, 2012). So kann die Diagnostik des Unterrichts

- überwiegend *quantitativ* (mit Einschätzskalen, Checklisten, Fragebögen) oder aber auch *qualitativ* erfolgen, wie z.B. beim Werkzeug «Fokus Unterricht» (Brosziewski & Maeder, 2007);
- auf einen *längeren Zeitraum* gerichtet sein, also auf Schülerseite kumulative Urteile erfordern (wie im Falle der Large-Scale-Studien und Lernstandserhebungen vom Typ PISA, TIMSS oder IGLU, aber auch wie in den meisten Schülerfragebögen der Schulinspektionen und Qualitätsagenturen), oder nur eine *konkrete Unterrichtsstunde* (oder eine Lerneinheit) betreffen;

- aus unterschiedlichen *Perspektiven* erfolgen, d.h. Urteile der unterrichtenden Lehrperson, der Schülerschaft, anwesender Gäste (Hospitantinnen und Hospitanten) und im Falle der Videoanalyse auch Urteile einer Gruppe miteinbeziehen;
- sich auf *fachübergreifende* («generische») Aspekte der Unterrichtsqualität oder auf *fachspezifische* (fachdidaktische, fachwissenschaftliche) Merkmale beziehen.

3 Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung

Die Unterrichtsforschung hat gezeigt, dass Urteile über Unterricht aus verschiedenen Perspektiven oft nicht übereinstimmen und jeweils unterschiedliche blinde Flecken aufweisen. Deshalb wird gefordert, bei der Unterrichtsdiagnostik ausdrücklich unterschiedliche Perspektiven einzubeziehen und verstärkt auf Perspektivenabweichungen zu fokussieren (vgl. Clausen, 2002). Genau dieser Forderung kommt das in Landau entwickelte Werkzeug EMU («Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung») nach. Es wurde im Auftrag der deutschen Kultusministerkonferenz (KMK) für die Praxis der Lehrpersonenaus- und -fortbildung entwickelt und umfasst Fragebögen, Texte, Handreichungen, Folien, Videos authentischer Unterrichtsstunden sowie Software für die Auswertung und Visualisierung von Unterrichtsbeobachtungen (Helmke et al., 2013). Das gesamte Material ist im Internet über die Website www.unterrichtsdiagnostik.info zugänglich, wo es uneingeschränkt, kostenlos und ohne irgendwelche Anmeldungen, Registrierungen, Pass- oder Kennwörter heruntergeladen und genutzt werden kann. Es richtet sich primär an Schulen, daneben aber auch an die Lehrpersonenaus- und -fortbildung. Häufig kommt der Anstoß dazu von den Schulen selbst; immer öfter empfiehlt aber auch die Schulaufsicht – im Rahmen von Zielvereinbarungen nach einer externen Evaluation –, EMU als Verfahren der Selbstevaluation zu verwenden. EMU hat sich in der Schulpraxis wie auch in der Lehrpersonenaus- und -fortbildung bewährt und ist in mehreren deutschen Bundesländern bereits Teil des regulären Aus- und Fortbildungsangebots im Lehramtsstudium. Darüber hinaus wird EMU auch von berufstätigen Lehrpersonen, Schulleitungen und Schulaufsichtsbeamtinnen und -beamten genutzt (so etwa in Bayern, Baden-Württemberg, Nordrhein-Westfalen, Sachsen-Anhalt und Rheinland-Pfalz sowie im Kanton Zürich).

Im Folgenden sollen die wichtigsten Features von EMU beschrieben werden (für weiterführende Hinweise und Details zum praktischen Vorgehen sowie zu den Gelingensbedingungen und Stolpersteinen in der Praxis vgl. www.unterrichtsdiagnostik.info sowie Helmke, Helmke & Schrader, 2012; Helmke et al., 2011).

3.1 Gegenstandsbereich

EMU orientiert sich, was die Auswahl von Qualitätsbereichen anbelangt, an der internationalen Unterrichtsforschung (Marzano, Pickering & Pollock, 2005; Hattie, 2009;

Helmke, 2012) und sieht eine Reihe von Qualitätsbereichen vor (in Klammern dahinter: die Anzahl der Items).

- *Basisbereiche*: Effiziente Klassenführung (5), lernförderliches Klima und Motivierung (6), Klarheit und Strukturiertheit (5) sowie Aktivierung (6) als Merkmale *prozessualer Unterrichtsqualität*. Hinzu kommt ein *Bilanzbereich* (5), d.h. eine Einschätzung der Unterrichtsstunde in emotionaler (Wohlfühlen), motivationaler (Interessantheit) und kognitiver Hinsicht (Lernertrag, Passung).
- *Zusatzbereiche*: Das oben genannte Fundamentum kann um Zusatzbereiche («Additum») ergänzt werden. Hierfür stehen gegenwärtig Itempools zu folgenden Bereichen zur Verfügung: kognitive Aktivierung (42), Umgang mit Vielfalt/Individualisierung (20), kooperatives Lernen/Gruppenarbeit (40), Lehrersprache (10), fachliche/fachdidaktische Qualität (3) sowie Kompetenzorientierung/Orientierung an den Bildungsstandards (alle derzeit geltenden Bildungsstandards).
- Zusätzlich können Instrumente zum Unterricht der externen Evaluation/Schulinspektion des Bundeslandes eingesetzt werden, selbst entwickelte Items zu Bereichen, die der Schule wichtig sind, oder es können Beobachtungsaufträge individuell vereinbart werden.

EMU ist zwar im Kern ein *quantitativer* Ansatz (Einschätzbögen), umfasst jedoch auch *qualitative* Elemente, beispielsweise (a) eine Checkliste zum Thema «Umgang mit Vielfalt» als Grundlage dafür, über Individualisierung und Differenzierung ins Gespräch zu kommen, und (b) Anregungen für konkrete Beobachtungsaufträge, etwa zum Verhalten ausgewählter Schülerinnen und Schüler.

3.2 Abgleich von Perspektiven

Den Kern von EMU stellt der Abgleich unterschiedlicher Perspektiven auf ein und denselben Urteilsgegenstand dar; dies ist üblicherweise eine konkrete Unterrichtsstunde. Dadurch bietet EMU Anlässe, miteinander über Konsens sowie Dissens – und die Gründe dafür – ins Gespräch zu kommen. Die jeweils identischen Formate und äquivalenten Formulierungen gewährleisten, dass die Urteile der unterrichtenden Lehrperson, des hospitierenden Gastes und der Schülerinnen und Schüler auf den gleichen Gegenstand gerichtet sind und gefundene Unterschiede zwischen den drei Ansätzen somit nicht auf die Unterschiedlichkeit der eingesetzten Werkzeuge, sondern auf unterschiedliche Sichtweisen zurückzuführen sind. Erst diese wechselseitige Anschlussfähigkeit macht einen Abgleich von Perspektiven möglich. Abbildung 3 veranschaulicht den Sachverhalt der Triangulation in Gestalt unterschiedlicher Mittelwertsprofile (rechte Seite) und zeigt darüber hinaus die Streuung der Antworten innerhalb der betreffenden Klasse (Stabdiagramme).

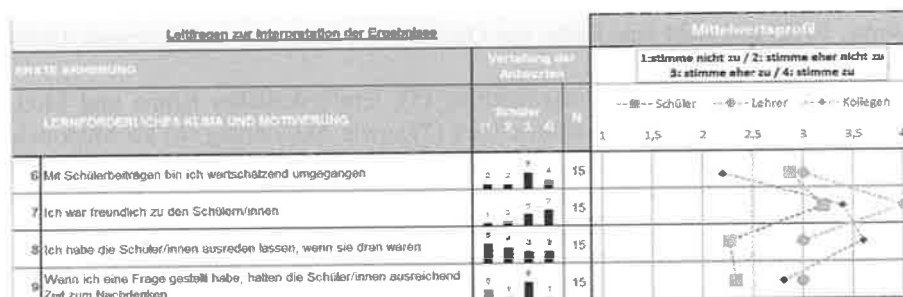


Abbildung 3: Triangulation: Unterricht aus drei Perspektiven (Ausschnitt aus einer Visualisierung, die die EMU-Software nach Eingabe der Rohdaten «auf Knopfdruck» erzeugt).

Zur Veranschaulichung der Architektur von EMU soll nachfolgend ein Beispielitem aus drei Perspektiven aufgeführt werden (4-stufiges Antwortformat: stimme nicht zu – stimme eher nicht zu – stimme eher zu – stimme zu):

- Unterrichtende Lehrerin: «Wenn ich eine Frage gestellt habe, hatten die Schüler/innen ausreichend Zeit zum Nachdenken.»
- Hospitierende Lehrperson: «Wenn die Kollegin eine Frage gestellt hat, hatten die Schüler/innen ausreichend Zeit zum Nachdenken.»
- Schülerfragebogen: «Wenn die Lehrerin in dieser Unterrichtsstunde eine Frage gestellt hat, hatte ich ausreichend Zeit zum Nachdenken.»

3.3 Modularisierung

EMU liegt ein Baukastenprinzip zugrunde: Je nach Bedarf und Situationseinschätzung lassen sich sehr unterschiedliche Anwendungsmöglichkeiten in der Art eines Menüs zusammenstellen. So kann man

- das Instrument zunächst einmal nur für sich selbst, als eine Art «Logbuch» verwenden.
- einen *Überblick* über alle fünf Bereiche gewinnen, aber auch nur *einen* Bereich auswählen.
- das Instrument nur *punktuell* («Momentaufnahme») oder zur Erfassung von *Veränderungen* als Ergebnis der Unterrichtsentwicklung mehrfach einsetzen.
- die eigene Sicht nur mit *einer* anderen Perspektive statt mit beiden (Kolleginnen/Kollegen, Klasse) abgleichen.

Diese Modularisierung (Auswahl von Perspektiven, Inhalten, Referenzzeitraum) ist nach den Erfahrungen aus der Pilotierungsphase für die Akzeptanz in der Praxis ausserordentlich wichtig. Dazu kommt die Möglichkeit, den angebotenen Itempool von EMU je nach Schwerpunkt des Schulprofils zu ergänzen und ihn nach persönlichen Interessen und Bedürfnissen zu individualisieren. In der Lehrerinnen- und Lehrerausbildung im Rahmen von Studienseminaren wird der im Internet angebotene fachübergreifende Teil von EMU beispielsweise durch fachspezifische Aspekte ergänzt. Dies ist gleichbe-

deutend mit einer Koppelung von generischen Merkmalen der Unterrichtsqualität mit fachdidaktischen und fachwissenschaftlichen Aspekten. Aufgrund dieses bedeutsamen Aspekts wurden – obwohl der Schwerpunkt von EMU klar auf fachübergreifenden Aspekten liegt – an mehreren Stellen Brücken zu fachspezifischer Reflexion geschlagen, zum Beispiel:

- Orientierung an den Bildungsstandards (D, CH, A);
- kollegiale Reflexion zur Erreichung der fachlichen Lernziele der Unterrichtsstunde;
- Impulse für die fachdidaktische Reflexion.

Schliesslich wurde dem Wunsch insbesondere berufsbildender Schulen Rechnung getragen, die dem Schülerfeedback häufig nicht nur eine einzelne Unterrichtsstunde, sondern eine komplette Unterrichtseinheit zugrunde legen möchten; hierfür können sie die EMU-Vorlage entsprechend dem in ihrem Bundesland bzw. Kanton verwendeten Vokabular adaptieren.

3.4 Praxisgerechte Formate

Ursprünglich war von der KMK als Auftraggeberin ein «Studienbrief» als Produkt erwartet worden, und einen solchen hatten wir, in Gestalt einer über 120 Seiten langen Broschüre, damals auch konzipiert und erstellt. Die darauffolgende einjährige Pilotierungsphase in Schulen, Studienseminaren und Universitäten sowie die Evaluation von EMU (Ade-Thurow, 2011) zeigten jedoch, dass die Akzeptanz vor allem in der Schulpraxis zu wünschen übrig liess: Die beteiligten Schulleitungen und Lehrpersonen äusserten in der Mehrzahl ganz offen, sie seien nicht bereit, sich mit einem derart langen Text auseinanderzusetzen. Als Resultat dieser Pilotierung entstanden im Dialog mit Vertreterinnen und Vertretern von Schulen handliche und praxisgerechte Formate (alles verfügbar unter www.unterrichtsdiagnostik.info), beispielsweise

- eine einseitige Kurzinformation (gut zu lesen in der grossen Pause),
- eine zehnsseitige Broschüre (gut zu lesen in einer Freistunde),
- Fragebögen,
- vertiefende Texte, abrufbar über einen Hyperlink in der oben genannten Broschüre,
- eine Powerpoint-Präsentation für die Vorstellung von EMU in den Schulen.

3.5 Veränderungssensitivität

EMU bietet die Möglichkeit einer Messwiederholung (z.B. in der gleichen Klasse, im gleichen Fach und mit äquivalentem Studententypus), um durch eine Follow-up-Erhebung Veränderungen visualisieren zu können. Wenn Unterrichtsentwicklung erfolgreich war, dann sollte sich das auch in den Daten niederschlagen, beispielsweise in einer veränderten Schülerwahrnehmung des Unterrichts. Abbildung 4 zeigt den Vergleich «Vorher – Nachher» für eine Lehrperson, die EMU in ihrer Klasse eingesetzt hatte und vom Ergebnis der Erstmessung überrascht war, weil ihr dies nie zuvor in den Sinn gekommen wäre: Mehr Schülerinnen und Schüler, als ihr lieb waren, attestierten ihr, zu

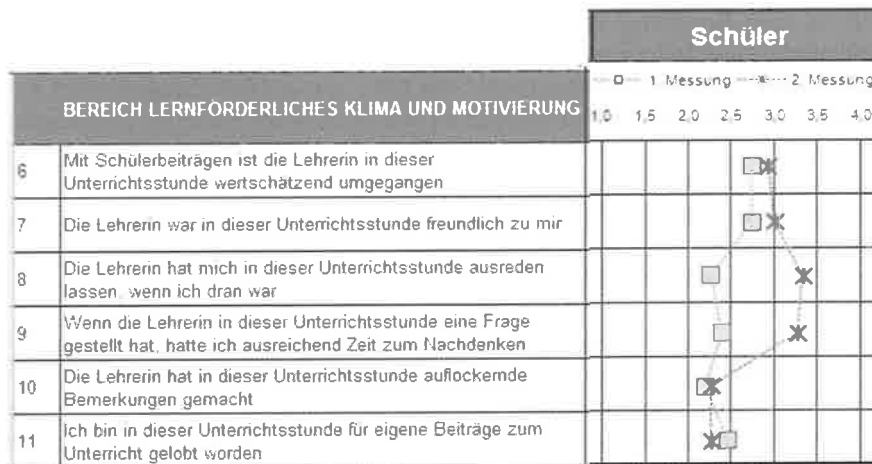


Abbildung 4: Vorher-Nachher-Messung mit EMU (Ausschnitt).

wenig Zeit zum Überlegen gehabt zu haben und häufig unterbrochen worden zu sein. Daraufhin arbeitete diese Lehrperson an ihrem Interaktionsverhalten und führte mithilfe eines MP3-Rekorders (es muss nicht immer Video sein) eine Art «Self-Monitoring» durch. Die Ergebnisse der zweiten Messung, ca. 6 Wochen später, zeigten ihr, dass sich ihr Verhalten aus Sicht der Schülerinnen und Schüler geändert hatte.

3.6 Von der Unterrichtsentwicklung zur Schulentwicklung

Die Verbesserung der Kooperation und des kollegialen Austauschs innerhalb des Kollegiums wie auch die Überwindung von Einzelkämpfertum und Isolation gehören nicht zu den Primärzielen, sind aber ausserordentlich erwünschte Nebenwirkungen von Massnahmen der Unterrichtsdiagnostik. Zudem gibt es Belege dafür, dass die Qualität der kollegialen Kooperation positive Effekte auf das Kompetenzniveau der Schülerinnen und Schüler hat (vgl. Huber, Ahlgrimm & Hader-Popp, 2012; Fussangel & Gräsel, 2012).

Das EMU-Team hat zusätzlich eine Software entwickelt, mit deren Hilfe viele individuelle EMU-Daten (aus Kollegialfeedback und/oder Schülerfeedback) auf Schulebene aggregiert und visualisiert werden können. Auf diese Weise sind die Ergebnisse nicht nur für individuelle Lehrpersonen und ihre Klassen von Nutzen, sondern die gesamte Schule kann in einen Lernprozess eingebunden werden, dem im Idealfall ein Dreischritt zugrunde liegt: (1) Diagnose der Ausgangslage, (2) Intervention, also vor allem Unterrichtsentwicklung, und (3) Evaluation, d.h. Wiederholung der Messung, wobei der Schwerpunkt auf einer Veränderung des Unterrichtsprofils der gesamten Schule liegt.

4 Elemente der Lernwirksamkeit einer kriteriengeleiteten Reflexion über Unterricht

Die Literatur zu «Teacher Change» (vgl. Richardson & Placier, 2002) zeigt, wie schwierig es ist, Veränderungen des Lehrens zu bewirken. Angesichts der oft beschriebenen (Wahl, 2001) Resistenz eingefahrener, unbewusst ablaufender Routinen, Gewohnheiten und Skripts, aber auch der Stabilität impliziter, subjektiver Unterrichtstheorien und Überzeugungen (Beliefs) kommt der Hospitation in Veränderungsprozessen ein grosses Potenzial zu, weil bei diesem Szenario mehrere Elemente zusammenkommen, welche jedes für sich genommen lernförderlich sind. Diese lernförderlichen Elemente werden im Folgenden kurz beschrieben.

4.1 Evidenzbasierung

Die Unterrichtsdiagnostik basiert auf empirischer Evidenz, auf Beschreibungen des Unterrichts, die sich auf Beobachtungen stützen. Dies eröffnet die Möglichkeit eines evidenzbasierten Austauschs über Unterricht sowie einer datenbasierten Reflexion und Massnahmenplanung im Sinne einer gezielten Verbesserung des Unterrichts.

4.2 Lernen durch Beobachtung

Lernen durch Beobachtung («Lernen am Modell») eröffnet der *hospitierenden Kollegin* oder dem *hospitierenden Kollegen* die Chance, das eigene Verhaltensrepertoire weiterzuentwickeln. Dabei geht es keineswegs nur um die Imitation von Verhalten, sondern auch um das Entdecken von Regeln, Ritualen und Methoden (welche man bisher selbst nicht angewendet hat) sowie um die Beobachtung von Ursache-Wirkungs-Zusammenhängen. Die Lernsituation ist aussichtsreich, weil man – anders als die unterrichtende Kollegin oder der unterrichtende Kollege – nicht unter Handlungsdruck steht und kein «cognitive overload» droht, sodass man den Blick auch auf normalerweise übersehene, wichtige Details richten kann.

Der *unterrichtenden Kollegin* oder dem *unterrichtenden Kollegen* bietet sich demgegenüber die Gelegenheit, neue Verhaltensweisen, Prozeduren, Szenarien oder Arrangements (z.B. eine neue Variante des kooperativen Lernens wie Gruppenpuzzle) zu erproben – verbunden mit der Chance, durch anschliessendes Feedback dazuzulernen (vgl. auch Abschnitt 6.3 zum Microteaching). Was für das Lernen von Schülerinnen und Schülern gilt, ist auch für das Lernen von Lehrpersonen relevant: Schaffung von Lernsituationen, in denen man auch einmal etwas ausprobieren kann und demnach auch Fehler machen darf, aus denen man mit Unterstützung der Kollegin oder des Kollegen lernen kann – dies im Gegensatz zu prüfungsähnlichen Leistungssituationen, wie sie z.B. bei Unterrichtsbesuchen mit dem Ziel der Personalbeurteilung durch die Schulleitung entstehen.

4.3 Potenzial des kriteriengeleiteten Perspektivenabgleichs

Die Konstellation einer kriteriengeleiteten, beobachtungsbasierten wechselseitigen Unterrichtsbeobachtung in Verbindung mit darauf beruhender gemeinsamer Reflexion bietet ein grosses kognitives Lernpotenzial. Sie stellt zum einen eine Gelegenheit dafür dar, auf eigene Routinen, Gewohnheiten und Marotten aufmerksam zu werden, vor allem aber auch dafür, eigener *handlungsleitender subjektiver Theorien* gewahr zu werden. Ziel dabei ist nicht die Zerschlagung dieser subjektiven Theorien (weitgehend synonym mit Alltagstheorien, naiven Theorien, impliziten Theorien), sondern vielmehr ihre «Öffnung» (Fauser, Rissmann & Weyrauch, 2009). Zum anderen sind kognitive Konflikte, Dissonanzen und Irritationen Katalysatoren für Lernprozesse, wie man seit Piaget weiss: Die Erfahrung solcher Widersprüche kann günstigenfalls dazu führen, bisherige Konzepte zu überdenken, d.h. zu schärfen oder zu kalibrieren, und den damit verbundenen Kurs zu ändern.

Auch der Forschungsansatz des *Conceptual Change* (Schnotz, 2006) ist in diesem Zusammenhang von Bedeutung: Letztendlich geht es ja um die Veränderung von Wissen, insbesondere darum, subjektiv vorhandene, aber dysfunktionale Vorstellungen zu korrigieren, sowie darum, «vorhandene Fehlkonzepte durch wissenschaftliche Konzepte zu ersetzen. ... Dementsprechend soll der Lernende mit neuen Erfahrungen konfrontiert werden, die bei ihm kognitive Konflikte auslösen und die bisherigen Konzepte in Frage stellen» (Schnotz, 2006, S. 78). Solche Prozesse der Wissensveränderung werden – so die Forschung zum *Conceptual Change* – vor allem dann gefördert, wenn es sich um authentische und persönliche Kontexte handelt und «wenn der Wissenserwerb in einer ermutigenden Lernkultur stattfindet, in der der kooperative Erwerb von Kompetenzen im Vordergrund steht» (Schnotz, 2006, S. 81). Genau dies ist die Essenz von EMU: Die Hospitationssituation ist *situiertes Lernen* in Reinform. Es handelt sich um Lernen «vor Ort», um ein persönlich hoch bedeutsames Thema, und der kollegiale Austausch im Anschluss an die Unterrichtsstunde findet auf Augenhöhe und im bewertungsfreien Raum statt.

4.4 Handlungsorientierung und Zielbezug

Diagnostik ist kein Selbstzweck, sondern hat ebenso wie eine aufwendige externe Evaluation nur eine dienende Funktion, nämlich die Qualität von Schule und Unterricht zu sichern und – falls nötig – zu verbessern. Konstitutiver Bestandteil des EMU-Szenarios nach einer gemeinsamen Besprechung einer Unterrichtsstunde anhand der dazu erhobenen Daten sind (a) die stichwortartige Zusammenfassung der Essentials des Gesprächs, (b) die Planung konkreter Massnahmen der Unterrichtsentwicklung, verbunden mit einer Schätzung des Unterstützungsbedarfs, und – bei mehreren Sitzungen im gleichen Schuljahr – (c) eine rückblickende Reflexion darüber, ob die Massnahmen überhaupt realisiert wurden und, wenn ja, wie effektiv (Hat es etwas gebracht? Wer hat davon profitiert? Wie habe ich das festgestellt?) und wie effizient (Stand der Aufwand in einem gesunden Verhältnis zum Ertrag? Wenn nein, was könnte man künftig optimieren?) diese Massnahmen waren.

4.5 Forschendes Lernen

Anders als beispielsweise bei der Unterrichtsbeobachtung durch Schulinspektorinnen und Schulinspektoren im Rahmen der externen Evaluation von Schulen geht es bei EMU ausdrücklich nicht darum, so genau wie möglich zu «messen». Im Gegenteil, man könnte sogar sagen, der Weg sei das Ziel: Im Prozess, Konzepte der Unterrichtsqualität präzise und wissenschaftlich zu bestimmen, unterschiedliche Definitionen voneinander abzugrenzen, unterrichtsbezogene Urteile auf beobachtbare Indikatoren zu beziehen und über deren Rolle und Gewichtung nachzudenken, liegt eine wesentliche Zielsetzung von EMU. Insofern umfasst die Situation bei EMU auch wesentliche Aspekte dessen, was man als «forschendes Lernen» bezeichnet (Dirks & Hansmann, 2002).

Dabei sind Widersprüche und Dissens kein unerwünschtes «Rauschen in den Daten» und keine «Fehlervarianz», sondern eine willkommene und anregende Gelegenheit für einen Diskurs über Unterricht, der nicht von allgemeinen didaktischen Konzepten und Theorien, sondern von beobachtbarem Verhalten ausgeht und von dort aus gegebenenfalls wieder vom Konkreten zum Abstrakten aufsteigt. Den Dreh- und Angelpunkt der Reflexion über Unterricht und des beobachtungsbasierten Feedbacks stellen Informationen auf der Ebene von *Items* dar. Aus diesem Grund bilden wir keine Summenscores (Mittelwert der Items eines Qualitätsbereichs). Insofern ist EMU zwar das *Ergebnis* langjähriger Unterrichts- und Lehrpersonenforschung, für quantitative Ansätze der Unterrichtsforschung, die zur Erfassung übergeordneter Merkmale (wie z.B. «Effizienz der Klassenführung») mit Summenscores arbeiten, jedoch nicht gedacht und auch nicht geeignet.

5 Ein Beispiel aus der Praxis von EMU

Zur Illustration des Einsatzes von EMU in der Schulpraxis soll in diesem Abschnitt ein (reales) Beispiel aufgeführt werden: In einer Schule hatte sich das Jahrgangsstufenteam der 7. Klasse darauf verständigt, gemeinsam die Qualität des Unterrichts zu sichten, um ihn – wo nötig – gezielt zu verbessern. Dazu wurde das EMU-Werkzeug eingesetzt. Bei einem Tandem kam es dabei zu einem Dissens zwischen der Einschätzung der unterrichtenden und derjenigen der hospitierenden Lehrperson. Erstere nannte als Grund für ihre Einschätzung von Item 24 («Die Unterrichtsstunde war für die Schüler/innen interessant.»), dem sie «eher nicht» zustimmte, dass sich nur wenige Schülerinnen und Schüler aktiv mit Meldungen am Unterricht beteiligt hätten – insbesondere bei schwierigeren Fragen. Aufgrund ihrer bisherigen persönlichen Erfahrung war sie der Ansicht, dass auch eine alltagsnahe Problemstellung in dieser Klasse zu keinem höheren Engagement führen würde. Sie glaubte, dass Schülerinnen und Schüler in diesem Alter generell weniger Interesse an Schule und Unterricht hätten. Die *hospitierende* Lehrperson hingegen teilte diese subjektive Theorie nicht. Ihrer Einschätzung zufolge hatten die Schülerinnen und Schüler einen durchaus anregenden und für sie interessanten Unter-

richt erlebt. Die geringe Anzahl sich beteiligender Schülerinnen und Schüler, gerade bei schwierigeren Fragestellungen, hatte die hospitierende Lehrperson zwar auch wahrgenommen, führte die geringe Anzahl an Meldungen jedoch nicht auf das Interesse, sondern auf die zur Verfügung stehende Antwortzeit zurück. Sie glaubte, die Schülerinnen und Schüler hätten schlichtweg zu wenig Zeit gehabt, um die Fragestellungen zu beantworten, da die unterrichtende Lehrperson stets schon nach kurzer Zeit eine Schülerin oder einen Schüler aufgerufen habe.

Die beiden Lehrpersonen zogen in einem entscheidenden zweiten Schritt die Schülerperspektive hinzu, um ihre Hypothesen zu verifizieren bzw. zu falsifizieren, und inspizierten die Verteilung der Schülerantworten (vgl. Abbildung 5).

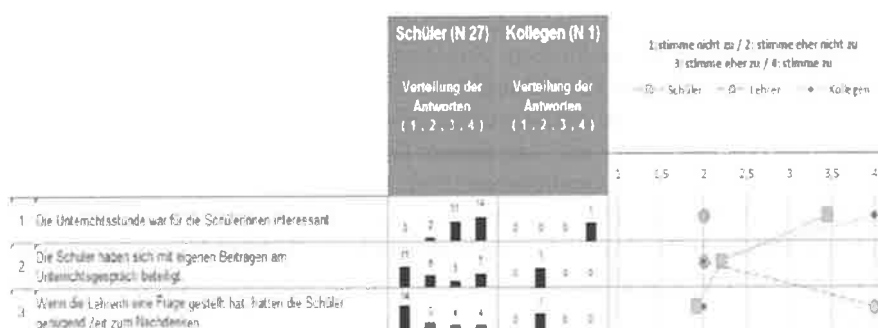


Abbildung 5: Visualisierung des Praxisbeispiels mithilfe der EMU-Software.

In diesem Beispiel fanden zwar auch die Schülerinnen und Schüler die Stunde interessant. Es stellte sich jedoch heraus, dass etwa zwei Drittel von ihnen nicht genügend Zeit zum Nachdenken hatten. Durch den Perspektivenabgleich wurde der unterrichtenden Lehrperson bewusst, dass ihre eigene Annahme auf einer subjektiven Theorie beruhte, diese jedoch nicht der Realität entsprach. Sie sieht ihre Lerngruppe nun mit anderen Augen und kann sich aufgrund des Feedbacks besser auf sie einstellen, indem sie den Schülerinnen und Schülern z.B. mehr Zeit zum Nachdenken gewährt. EMU unterstützt Lehrkräfte also darin, das Lernen der Schülerinnen und Schüler – ganz nach Hatties (2009) Motto – durch die Augen der Schülerinnen und Schüler zu sehen.

Des Weiteren hatte die oben genannte Situation zur Folge, dass sich die beteiligten Lehrpersonen zu einer ausgewählten Fragestellung – Rolle der Wartezeit – in der Fachliteratur informierten und auf diese Weise die empirische Befundlage zur Kenntnis nahmen – z.B. Rows (1986) Artikel mit dem vielsagenden Titel «Wait time: Slowing down may be a way of speeding up» –, die ihnen bis dato unbekannt gewesen war. Obwohl es seit Langem empirisch begründete untere «Schwellenwerte» für die *Wartezeit* («wait time») auf Schülerantworten gibt (sofern es sich nicht um automatisierte Prozesse han-

delt, wie z.B. beim Kopfrechnen), nämlich minimal drei Sekunden, begrenzt sich die Wartezeit von Lehrpersonen im Schnitt auf eine Sekunde oder weniger; anschliessend wiederholen sie die Frage, geben diese an andere Schülerinnen und Schüler weiter, brechen die Frage-Antwort-Sequenz ab oder geben die Antwort praktischerweise gleich selbst (Helmke et al., 2008; Hattie, 2012). Ein solches Verhalten wollten die beteiligten Lehrpersonen in ihrem Unterricht künftig vermeiden. Aber können Lehrpersonen im laufenden Unterricht überhaupt ein «Monitoring» ihres eigenen Verhaltens leisten? Hierzu meint der amerikanische Unterrichtsforscher Borich (2007, S. 13):

Within the busy schedule of a school day, teachers do not have many opportunities to reflect on the relative merits of the strategies and methods they use. To pause for contemplation during instruction could disrupt the rapidity of classroom events, and almost surely would result in a loss of momentum; to pause after class or at the end of the school day would require the ability to accurately recall events that may have occurred hours earlier. As a result, teachers frequently can be observed performing behaviors that they are unaware of, such as dominating discussions and allowing too little response time for students to think through an answer.

Ohne den Abgleich mit anderen Perspektiven und deren explizite Thematisierung bleiben subjektive Theorien meist implizit. Die Wahrnehmung des eigenen Lehrhandelns und der eigenen Person unterscheidet sich in der Regel von der Fremdwahrnehmung und der Realität (Clausen, 2002). Illustriert wird dies durch das sogenannte Johari-Fenster (vgl. Abbildung 6), eine Denkfigur aus der Gruppendynamik, deren Name sich aus den jeweils ersten Buchstaben der Vornamen der beiden «Erfinder» Joseph Luft und Harrington Ingham ableitet (Luft & Ingham, 1955).

<p>A öffentlich mir bekannt, anderen bekannt</p>	<p>C blinder Fleck mir nicht bekannt, anderen bekannt</p>
<p>B geheim mir bekannt, anderen nicht bekannt</p>	<p>D unbewusst mir nicht bekannt, anderen nicht bekannt</p>

Abbildung 6: Die vier Felder des Johari-Fensters.

Wie Abbildung 6 verdeutlicht, gibt es Bereiche, die der eigenen Wahrnehmung verborgen bleiben/sind (C, D). Während der Bereich des Unbewussten (D) weder der Selbst- noch der Fremdwahrnehmung zugänglich ist, beinhaltet der blinde Fleck (C) den Anteil des Verhaltens, den man selbst wenig wahrnimmt, obwohl er anderen offenkundig ist. Feedback kann helfen, die Fenstergrössen zu verändern bzw. den blinden Fleck zu reduzieren und dadurch die Handlungsfähigkeit zu steigern.

6 Nutzen von EMU für verschiedene Zielgruppen und Szenarien innerhalb der Lehrerinnen- und Lehrerbildung

Die bisherigen Ausführungen waren vor allem für die Schulpraxis und die Lehrpersonenfortbildung von Bedeutung. Das Programm der Unterrichtsdiagnostik eignet sich aber ebenso für die Lehrpersonenausbildung, nämlich sobald die Studierenden mit der Praxis eigenen Unterrichtens konfrontiert werden. Dadurch ergeben sich beispielsweise Lerngelegenheiten, um eigene Beobachtungskompetenzen zu entwickeln, wozu insbesondere eine klare Unterscheidung zwischen Beschreibung und Bewertung gehört. Erfahrungen aus der Lehrpersonenausbildung in Deutschland sowohl aus der ersten Phase (Universität, PH) wie auch aus der zweiten Phase (Studienseminar) zeigen die Tendenz, (zu) schnell zu einem global bewertenden Urteil zu gelangen. Die Vorschaltung einer differenzierten Analyse auf der Ebene von Qualitätsbereichen und einzelnen Items ergibt am Ende vielfach ein anderes Bild. Damit kann die Nutzung von EMU in der Lehrpersonenausbildung dazu beitragen, frühzeitig ein wissenschaftlich fundiertes Verständnis für eine methodisch fundierte Unterrichtsbeurteilung und für eine evidenzbasierte Unterrichtsentwicklung zu schaffen.

6.1 Konzepte für Lehrveranstaltungen

Für die Praxis der Lehramtsausbildung werden auf der Website von EMU (www.unterrichtsdiagnostik.info) konkrete und bewährte (von den Autorinnen und Autoren in ihren eigenen Lehrveranstaltungen mehrfach selbst praktizierte) Konzepte für eine Lehrveranstaltung zur Unterrichtsdiagnostik mit zwei Wochenstunden pro Semester (alternativ: mit mehreren Blockveranstaltungen) vorgeschlagen, die folgende Module umfassen:

- Information über zentrale fachübergreifende (und gegebenenfalls zusätzlich fachspezifische) Aspekte der Unterrichtsqualität (Input von Expertinnen und Experten: Dozierende oder Gast).
- Praktische Durchführung der Unterrichtsdiagnostik auf der Grundlage eines Unterrichtsvideos.
- Der datenbasierte Austausch über Unterricht im Kollegium bzw. in der Fachgruppe wird durch Kleingruppen von Studierenden simuliert, die Unterrichtsvideos verschiedener Fächer, Klassenstufen und Schularten individuell beobachten und bewerten, die Daten mithilfe der hierfür entwickelten speziellen Software eingeben und sie anschließend – auf der Basis der vom Programm grafisch illustrierten Profile – im Team diskutieren. Die Gruppen präsentieren ihre Ergebnisse im Plenum, wobei der Schwerpunkt auf folgenden Aspekten liegt: (a) Qualitätsbereiche und Sachverhalte (Items), bei denen die Übereinstimmung gering ist (Dissens), (b) Schwierigkeiten, Probleme oder offene Fragen bei der Interpretation der Ergebnisse, (c) ergänzende fachspezifische Aspekte der Unterrichtsqualität und (d) konstruktive Vorschläge zur Verbesserung der Instrumente mit dem Ziel, die Brauchbarkeit für die Lehrpersonenausbildung und die Schulpraxis zu steigern.

- Vergleich unterschiedlicher Instrumente zur Beurteilung einer konkreten Unterrichtsstunde, ihrer Inhalte, Antwortformate, Stärken und Schwächen.

Hierfür gibt es, ergänzend zu EMU, auch Übungsmaterial aus verschiedenen Bundesländern, beispielsweise die Broschüre BBBB (Beobachten – Beschreiben – Bewerten – Begleiten, vgl. <http://lehrerfortbildung-bw.de/allgschulen/bbbb/>). Das Potenzial des Tandems, bei dem in der Schulpraxis die Rollen von unterrichtender und hospitierender Lehrperson jeweils rotieren, lässt sich bei Lehramtsanwärterinnen und Lehramtsanwärtlern ebenfalls nutzen, jedoch in einer anderen Architektur: keine Tandems auf Augenhöhe und mit Rotation der Rollen (A unterrichtet und B hospitiert, anschließend umgekehrt), sondern als *Coaching-Situation* (Mentorin oder Mentor/Ausbilderin oder Ausbilder einerseits, Studentin oder Student/Mentee andererseits; vgl. dazu die Vorarbeiten von Staub 2001, 2005, sowie Staub & Niggli, 2009).

6.2 Videobasierte Unterrichtsanalyse im Team

Ein anderes, für Seminarveranstaltungen – und für andere Settings, in deren Rahmen Gruppen, Teams oder Fachschaften gemeinsam (videografierten) Unterricht reflektieren – besonders geeignetes Format ist die videobasierte Unterrichtsanalyse im Team (zum Potenzial der Nutzung von Videos für Coaching und Unterrichtsreflexion vgl. z.B. Sherin, 2003, sowie für den deutschen Sprachraum die Pionierarbeiten der Gruppe um Reusser, z.B. Petko & Reusser, 2005). Dieses Szenario sieht die folgenden Schritte vor: (1) Einführung in Fragen der Unterrichtsqualität und Vertrautmachen mit einem konkreten Diagnoseinstrument, z.B. EMU, (2) gemeinsame Bearbeitung des Unterrichtsbeobachtungsbogens, (3) Eingabe der Daten und (4) Diskussion der visualisierten Ergebnisse, die z.B. so aussehen könnten, wie in Abbildung 7 dargestellt.

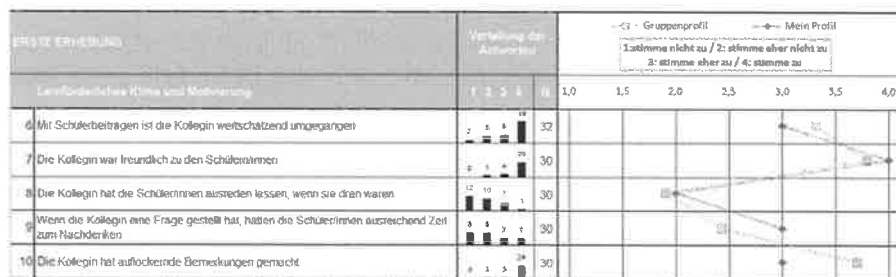


Abbildung 7: Visualisierte Ergebnisse einer Unterrichtsbeobachtung als Grundlage für die Unterrichtsanalyse im Team.

Dabei fallen vier Arten von Informationen an, die Grundlage für gemeinsame und/oder individuelle Reflexion sein können:

- *Analyse des Profils*: Wo zeigen sich auf der Basis des (über alle Urteilenden gemittelten) Gruppenprofils Stärken und Schwächen im Unterricht?

- *Konsens und Dissens*: Wo stimmen die Urteilenden überein, wo nicht – und wenn Letzteres: warum?
- *Milde und Strenge*: Für alle Urteilenden lassen sich Kennwerte der individuellen Urteilstendenz berechnen, insbesondere in Bezug auf Milde versus Strenge.
- *Idiosynkratisches Urteilsprofil*: Wo weiche ich («mein Profil») vom Profil der gesamten Gruppe («mainstream») ab und woran könnte das liegen?

Der Gewinn von Unterrichtsanalysen im Team liegt vor allem darin, dass man

- durch die Notwendigkeit, die eigenen Urteile zu begründen und mit beobachtetem Verhalten zu belegen, zu genauen Beobachtungen und zur Explikation der eigenen Alltagskonzepte und des impliziten Verständnisses von gutem Unterricht ermuntert wird;
- das eigene Beobachtungsverhalten und die eigenen Beurteilungsmassstäbe durch den Vergleich mit anderen Lehrpersonen einzuschätzen lernt;
- durch einen intensiven Diskurs ein gemeinsames Begriffsverständnis entwickeln und in der Folge Differenzen minimieren kann. Lehrpersonen unterhalten sich über Unterricht häufig auf einer abstrakten Sprachebene mittels Begrifflichkeiten wie beispielsweise «kindgerecht», «alltagsnah» oder «schülerorientiert». Zum Teil definieren sie diese Begrifflichkeiten jedoch unterschiedlich und reden somit aneinander vorbei – ohne dies zu realisieren. Unterrichtsreflexion mit Fokus auf eine konkrete Stunde unterstützt Lehrpersonen dabei, abstrakte Begrifflichkeiten mit Inhalt zu füllen und Definitionen abzugleichen.

6.3 Microteaching und Unterrichtsdiagnostik

Ein überraschendes Ergebnis der Hattie-Studie war für viele der enorm positive Effekt des Microteachings (für Details vgl. Klinzing, 2002). Auf der Rangliste der 138 Effektstärken steht das Microteaching auf Platz 4 (Hattie, 2013, S. 134 f.). Die Kernelemente des Microteachings sind bemerkenswerterweise identisch mit jenen der evidenzbasierten Reflexion von Unterricht im Sinne von EMU: In beiden Fällen geht es um eine erfahrungsgestützte, beobachtungsbasierte kollegiale Reflexion über Unterricht. Der Unterschied besteht lediglich darin, dass das Microteaching ausserhalb des regulären Unterrichts angesiedelt ist und mit videogestütztem Feedback arbeitet, während EMU sich auf naturalistischen Unterricht bezieht und die Datenbasis aus der Bearbeitung des EMU-Werkzeugs bezieht.

Unter diesen Umständen erscheint es aussichtsreich, in der Lehrpersonenaus- und -fortbildung Situationen zu schaffen, in denen das nachweislich erfolgreiche Microteaching mit kriteriengeleiteter Unterrichtsdiagnostik und Reflexion über Unterricht gekoppelt wird. Kramis (1991) hat dies – lange vor Hattie und EMU – in seinem wegweisenden Aufsatz «Eine Kombination mit hoher Effektivität: Microteaching – Reflective Teaching – Unterrichtsbeobachtung» bereits vor über 20 Jahren empfohlen!

7 Ausblick

Zurück zu Hattie: Die aus unserer Sicht vielleicht überzeugendste Botschaft der Hattie-Studie (2009, 2013) ist, dass Methoden und Organisationsformen des Unterrichts per se eher unbedeutend sind. Man kann sehr viele unterschiedliche Szenarien des Unterrichts so anlegen, dass sie lernwirksam werden. Mit anderen Worten, der Einsatz bestimmter Methoden und die Nutzung bestimmter Werkzeuge per se ist nicht automatisch lernförderlich; es kommt auf das *Wie* an und weniger auf das *Ob*. Diese Kernaussage gilt selbstverständlich auch für die EMU-Unterrichtsdiagnostik. Das heisst: Kollegiale Hospitation unter Nutzung der EMU-Werkzeuge ist nicht automatisch ertragreich. Wird sie als «Wellness-Veranstaltung» betrachtet, verbunden mit gegenseitiger Schmeichelei (nach dem Motto «Nenn mich Goethe, dann nenn ich dich Schiller» – eine originelle Metapher, die von H.G. Rolff (2010, S. 69) stammt), dann verfehlt sie ihren Zweck. Wird sie hingegen ernsthaft und verantwortungsvoll angegangen, dann erfordert sie die strikte Beachtung von Feedbackregeln und die Fähigkeit, konstruktive Kritik zu geben und auch zu ertragen. Dabei soll nicht verschwiegen werden, dass das Verlassen «bewährter» Routinen, die Öffnung und vielleicht Revision subjektiver Theorien im Einzelfall durchaus auch anstrengend sein kann – «Lernen ist schmerzhaft», so betitelten Oser und Spychiger (2005) ihr folgenreiches Buch. Viel stärker ist allerdings das Entlastungspotenzial: Es kann erleichternd sein, zu sehen, dass andere auch nur «mit Wasser kochen» und vielleicht mit ähnlichen Problemen zu kämpfen haben wie man selbst und dass man diese Probleme jetzt gemeinsam angehen kann statt allein. Ausserdem ist regelgestütztes Feedback niemals destruktiv und nur negativ. Auch dies ist für Lehrpersonen eine Quelle für positive Motivation, denn in der Regel erhalten sie eher Feedback, wenn etwas nicht so gut läuft, als wenn es reibungslos und glatt läuft. Dazu kommt, dass es befriedigend sein kann, einmal schwarz auf weiss zu erfahren, dass man etwas gut kann, von dem man bislang lediglich vage vermutete, es zu können (vgl. das EMU-Feedbackgespräch unter www.unterrichtsdiagnostik.info/video). Vor allem aber verspricht die Investition von Zeit und Energie in EMU eine lohnende Rendite: die Förderung des Lehrens und Lernens und damit die Zufriedenheit mit dem eigenen Beruf. Aus diesem Grund hat das Kultusministerium Baden-Württembergs das Diagnosewerkzeug EMU als zentrales Modul in das landesweite Fortbildungsangebot «10plus – Motiviert und gesund bleiben im Lehrberuf» integriert (Landesakademie für Fortbildung und Personalentwicklung an Schulen, 2013).

Literatur

- Ade-Thurow, M.** (2011). *Empirische Untersuchungen zur Unterrichtsdiagnostik an Schulen*, Masterthesis im Masterstudiengang Schulentwicklung. Weingarten: Pädagogische Hochschule Weingarten.
- Altrichter, H. & Posch, P.** (2007). *Lehrerinnen und Lehrer erforschen ihren Unterricht* (4. Auflage). Bad Heilbrunn: Klinkhardt.
- Beywl, W.** (2013). Mit Taten zu Daten. Der Ansatz der unterrichtsintegrierten Selbstevaluation. *Journal für Schulentwicklung*, 17 (1), 7–14.

- Borich, G.D.** (2007). *Observation Skills for Effective Teaching* (5. Auflage). Upper Saddle River, NJ: Merrill Prentice Hall.
- Brosziewski, A. & Maeder, C.** (2007). *Fokus Unterricht. Unterrichtsentwicklung durch Beobachtung*. Zürich: Seismo.
- Clausen, M.** (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Dirks, U. & Hansmann, W.** (Hrsg.). (2002). *Forschendes Lernen in der Lehrerbildung. Auf dem Weg zu einer professionellen Unterrichts- und Schulentwicklung*. Bad Heilbrunn: Klinkhardt.
- Fauser, P., Rissmann, J. & Weyrauch, A.** (2009). Das Entwicklungsprogramm für Unterricht und Lernqualität. Lehrerfortbildung als theoriegeleitete Intervention und Ausbildung adaptiver Routinen. *Beiträge zur Lehrerbildung*, 27 (3), 361–371.
- Fussangel, K. & Gräsel, C.** (2012). Lehrerverbündung aus Sicht der Bildungsforschung. In E. Baum, T.-S. Idel & H. Ullrich (Hrsg.), *Kollegialität und Kooperation in der Schule: Theoretische Konzepte und empirische Befunde* (S. 29–40). Wiesbaden: Springer VS.
- Hattie, J.** (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J.** (2012). *Visible Learning for Teachers. Maximizing Impact on Learning*. London: Routledge.
- Hattie, J.** (2013). *Lernen sichtbar machen*. Überarbeitete deutschsprachige Ausgabe von «Visible Learning» besorgt von Wolfgang Beywl und Klaus Zierer. Baltmannsweiler: Schneider Verlag Hohengehren.
- Helmke, A.** (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (Schule weiterentwickeln – Unterricht verbessern: Orientierungsband) (4., überarbeitete Auflage). Seelze: Klett-Kallmeyer.
- Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. & Ade-Thurow, M.** (2011). Unterrichtsdiagnostik – Voraussetzung für die Verbesserung der Unterrichtsqualität. In A. Bartz, M. Dammann, S. Huber, C. Kloft & M. Schreiner (Hrsg.), *PraxisWissen SchulLeitung, AL 28* (Kapitel 30.71). Köln: Wolters Kluwer.
- Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. & Ade-Thurow, M.** (2013). *EMU – Unterrichtsdiagnostik*. Studienbrief Version 4.0. Kultusministerkonferenz: Projekt EMU (Evidenzbasierte Methoden der Unterrichtsdiagnostik). Landau: Universität Koblenz-Landau, Campus Landau.
- Helmke, A., Helmke, T. & Pham, G.** (2012). Von der externen zur internen Evaluation des Unterrichts. *Hamburg macht Schule*, Heft 3, 31–32.
- Helmke, A., Helmke, T. & Schrader, F.-W.** (2012). Unterrichtsdiagnostik mit EMU – Unterricht aus mehreren Perspektiven betrachten und diskutieren. PDF Download Unterrichtsqualität Sekundarstufe. In M. Bosen, W. Homeier & K. Tschekan (Hrsg.), *Unterrichtsqualität sichern – Sekundarstufe* (17. Ergänzung). Stuttgart: Raabe.
- Helmke, A. & Renkl, A.** (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schülerschulmerksamkeit im Unterricht. *Diagnostika*, 38 (2), 130–141.
- Helmke, T., Helmke, A., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K.** (2008). Die Videostudie des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 345–363). Weinheim: Beltz.
- Herlt, S. & Schaarschmidt, U.** (2007). Fit für den Lehrerberuf?! In U. Schaarschmidt & U. Kieschke (Hrsg.), *Gerüstet für den Schulalltag. Psychologische Unterstützungsangebote für Lehrerinnen und Lehrer* (S. 157–181). Weinheim: Beltz.
- Huber, S.G., Ahlgrimm, F. & Hader-Popp, S.** (2012). Kooperation in und zwischen Schulen sowie mit anderen Bildungseinrichtungen: Aktuelle Diskussionsstränge, Wirkungen und Gelingensbedingungen. In S.G. Huber & F. Ahlgrimm (Hrsg.), *Kooperation: Aktuelle Forschung zur Kooperation in und zwischen Schulen sowie mit anderen Partnern* (S. 323–372). Münster: Waxmann.
- Klinzing, H.G.** (2002). Wie effektiv ist Microteaching? Ein Überblick über fünfunddreißig Jahre Forschung. *Zeitschrift für Pädagogik*, 48 (2), 194–214.
- Kramis, J.** (1991). Eine Kombination mit hoher Effektivität: Microteaching – Reflective Teaching – Unterrichtsbeobachtung. *Unterrichtswissenschaft*, 19 (3), 260–277.

- Kultusministerkonferenz.** (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Bonn: KMK.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M.** (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Landesakademie für Fortbildung und Personalentwicklung an Schulen.** (2013). *Fortbildungsmaßnahme «10plus – Motiviert und gesund bleiben im Lehrberuf»*. Online verfügbar unter: <http://lehrerfortbildung-bw.de/qm/sonderaufgaben/gesundheit/10plus/tinfo.htm> (28.07.2013).
- Luft, J. & Ingham, H.** (1955). *The Johari Window, a graphic model for interpersonal relations*. Los Angeles: University of California, Western Training Laboratory in Group Development, Extension Office.
- Marzano, J.S., Pickering, D.J. & Pollock, R.J.** (2005). *Classroom Instruction that works*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Oser, F. & Spychiger, M.** (2005). *Lernen ist schmerzhaft. Zur Theorie des Negativen Wissens und zur Praxis der Fehlerkultur*. Weinheim: Beltz.
- Petko, D. & Reusser, K.** (2005). Praxisorientiertes E-Learning mit Video gestalten. In A. Hohenstein & K. Wilbers (Hrsg.), *Handbuch E-Learning. Expertenwissen aus Wissenschaft und Praxis* (1. Ergänzungslieferung) (S. 1–21). Köln: Deutscher Wirtschaftsdienst.
- Richardson, V. & Placier, P.** (2002). Teacher change. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4. Auflage) (S. 905–950). Washington: American Educational Research Association.
- Roff, H.G.** (2010). Trugschlüsse der Individualisierung. *Journal für Schulentwicklung*, Heft 3, 67–69.
- Rowe, M.B.** (1986). Wait time: Slowing down may be a way of speeding up. *Journal of Teacher Education*, 37 (1), 43–50.
- Schnotz, W.** (2006). Conceptual Change. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3., überarbeitete und erweiterte Auflage) (S. 77–82). Weinheim: Beltz Psychologie Verlags Union.
- Schön, D.A.** (1983). *The reflective practitioner. How professionals think in action*. New York: Basic Books.
- Schrader, F.-W.** (2011). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 683–698). Münster: Waxmann.
- Schrader, F.-W.** (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31 (2), 154–165.
- Sherin, M.G.** (2003). New perspectives on the role of video in teacher education. In J.E. Brophy (Hrsg.), *Using video in teacher education* (S. 1–27). Oxford: Elsevier.
- Staub, F.C.** (2001). Fachspezifisch-pädagogisches Coaching: Förderung von Unterrichtsexpertise durch Unterrichtsentwicklung. *Beiträge zur Lehrerbildung*, 19 (2), 175–198.
- Staub, F.C.** (2005). Videos im Fachspezifisch-Pädagogischen Coaching. *Journal für lehrerinnen- und lehrerbildung*, 5 (2), 26–30.
- Staub, F.C. & Niggli, A.** (2009). Zertifikatskurs der Universität Freiburg: Coaching und Mentoring in der Lehrerbildung. Certified Course at the University of Fribourg: Coaching and Mentoring in Teacher Education. *Beiträge zur Lehrerbildung*, 27 (1), 93–103.
- Wahl, D.** (2001). Nachhaltige Wege vom Wissen zum Handeln. *Beiträge zur Lehrerbildung*, 19 (2), 157–174.
- Weinert, F.E. & Helmke, A.** (1996). Der gute Lehrer: Person, Funktion oder Fiktion? In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule* (Zeitschrift für Pädagogik, 34. Beiheft) (S. 223–233). Weinheim: Beltz.

Autor und Autorin

Andreas Helmke, Prof. Dr., Universität Koblenz-Landau, Fachbereich Psychologie, helmke@uni-landau.de, andreas.helmke@uni-konstanz.de, www.andreas-helmke.de, www.unterrichtsdiagnostik.info

Gerlinde Lenske, Dr., Universität Duisburg-Essen, gerlinde.lenske@uni-due.de

Lehrpersonen mit Migrationshintergrund: Eine kritische Synthese der Literatur

Carola Mantel und Bruno Leutwyler

Zusammenfassung Lehrpersonen mit Migrationshintergrund wird in jüngerer Zeit von bildungspolitischer Seite ein besonderes Potenzial im Umgang mit migrationsbedingter Heterogenität in Schule und Unterricht zugeschrieben. Ein Überblick über den Forschungsstand zeigt ein vielfältiges Bild unterschiedlicher Hinweise auf ein besonderes Potenzial, allerdings kaum solide empirische Evidenz. Eine kritische Reflexion der Erkenntnisse zeigt auf, dass sowohl politische Zuschreibungen wie auch wissenschaftliche Überprüfungen eines zugeschriebenen Potenzials Gefahr laufen, ein essentialistisches Kulturverständnis zu begünstigen und damit die Herstellung von Differenzlinien zu unterstützen. Im Kontrast dazu wird angeregt, Forschungszugänge vermehrt auf der Basis eines konstruktivistischen Kulturverständnisses zu entwickeln und etwa danach zu fragen, ob, und falls ja, inwiefern eine familiäre Migrationsgeschichte zur Entwicklung interkultureller Kompetenz beiträgt, sowie in der Folge auch danach, inwiefern die Lehrerinnen- und Lehrerbildung diese Entwicklung fördern kann.

Schlagwörter Lehrperson mit Migrationshintergrund – Heterogenität – interkulturelle Pädagogik – interkulturelle Kompetenz

Teachers with an Immigrant Background: A Critical Synthesis of the Literature

Abstract It has lately been claimed by educational political bodies that teachers with an immigrant background have a specific potential for dealing with cultural heterogeneity in their teaching practice. An overview of the current state of research shows a variety of indications for a specific potential, but hardly any solid empirical evidence. Grounded in a critical appraisal of the findings, the authors point to the fact that political ascriptions of a specific potential as well as scientific examinations of such an ascribed potential run the risk of supporting an essentialist understanding of «culture» and thus of contributing to the construction of difference. In contrast, the authors suggest developing research approaches more decidedly based on a constructivist understanding of «culture», and asking, for instance, to what extent and in what ways an immigrant background contributes to the development of intercultural competence, and how teacher education can support such a development.

Keywords teachers with an immigrant background – heterogeneity – intercultural education – intercultural competence

1 Einleitung

Lehrpersonen mit Migrationshintergrund sind in den Schweizer Schulen nur selten anzutreffen. In den Lehrpersonenkollegien machen sie einen geringen Anteil aus, sodass diese Kollegien in ihrer kulturellen Zusammensetzung zumeist in einem starken Kontrast stehen zu einer ausgesprochen multikulturell zusammengesetzten Schülerschaft. So beträgt – um nur einen Indikator dafür zu nennen – der Anteil an Schülerinnen und Schülern mit ausländischer Staatsangehörigkeit auf der Primarstufe 23.5 Prozent (BFS, 2012), unter den Lehrpersonen hingegen nur gerade 2.3 Prozent.¹ Statistische Angaben zum Migrationshintergrund² liegen nicht vor, der Anteil dürfte aber in beiden Gruppen etwas höher liegen und das Ungleichverhältnis dabei ein ähnliches sein.

Mit Verweis auf ein solches Ungleichverhältnis wird verschiedentlich die Forderung nach «mehr Lehrpersonen mit Zuwanderungsgeschichte» laut, dies insbesondere in Deutschland, wo dieser Appell 2007 in den Nationalen Integrationsplan der Kultusminister aufgenommen wurde (Presse- und Informationsamt der Bundesregierung, 2007). Bestrebungen wurden in Gang gesetzt, um die Anzahl von Lehramtsstudierenden mit Migrationshintergrund zu erhöhen: Stiftungen setzen sich mit Stipendien für sie ein, Berlin und Hamburg haben bei der Vergabe von Referendariatsplätzen eine «Multi-kulti-Quote» eingeführt und in Nordrhein-Westfalen wurden Informationsworkshops durchgeführt und wurde ein Netzwerk für Lehrkräfte mit Zuwanderungsgeschichte gegründet (Rotter, 2009; Stiller & Zeoli, 2010).

In der Deutschschweiz kommt die politische Forderung, mehr Personen mit Migrationshintergrund zu Lehrpersonen auszubilden, insbesondere vonseiten der Expertinnen und Experten für interkulturelle Pädagogik. So wird im Schlussbericht der EDK (Schweizerische Konferenz der kantonalen Erziehungsdirektoren) zur interkulturellen Pädagogik in der Lehrerinnen- und Lehrerbildung festgehalten (EDK, 2000, S. 64):

Aufgrund von formalen und psychosozialen Hindernissen sind Studierende aus andern Kulturbereichen in der Lehrerinnen- und Lehrerbildung stark untervertreten. Durch eine vermehrte Aufnahme von Studierenden mit einem andern sprachlichen und kulturellen Hintergrund könnte der Austausch zwischen Kulturen während der Ausbildung direkt thematisiert und eins zu eins gelebt werden. Ausgebildete Lehrkräfte mit einem andern kulturellen Hintergrund könnten in multikulturellen Klassen wichtige Funktionen für die Verständigung, den Austausch und die Entwicklung von Wertschätzung und Akzeptanz «des Anderen» übernehmen und ebenso eine Bereicherung für das Lehrerkollegium bedeuten.

¹ Die Angabe zum Anteil ausländischer Lehrpersonen entstammt einer Auskunft des Bundesamts für Statistik (BFS). Beide Angaben zu den prozentualen Anteilen beziehen sich auf die ganze Schweiz und auf das Schuljahr 2010/2011.

² Ein «Migrationshintergrund» (zuweilen auch als «Zuwanderungsgeschichte» bezeichnet) wird unterschiedlich gefasst. Eine Möglichkeit ist die Bezeichnung als «zweite Generation», d.h. «in der Schweiz aufgewachsene, zumeist auch hier geborene Kinder ausländischer Eltern, die ihre Schulbildung ganz oder mehrheitlich in der Schweiz erworben haben» (Schweizerische Eidgenossenschaft, 2001, S. 1917).

Nebst der EDK empfiehlt auch die COHEP (Schweizerische Konferenz der Rektoren und Rektoren der Pädagogischen Hochschulen, 2007a, S. 6), Hindernisse für Studierende mit Migrationshintergrund und mehrsprachiger Biografie an den pädagogischen Hochschulen abzubauen und diese Studierenden während ihrer Ausbildung zu unterstützen.

Begründet werden solche Forderungen zumeist mit dem Verweis auf die oben erwähnte Unterrepräsentanz von Lehrpersonen mit Migrationshintergrund im Verhältnis zur Schweizer Wohnbevölkerung. Der Zusammenhang mit der interkulturellen Pädagogik und der Hinweis auf deren Funktion für eine höhere Wertschätzung und Akzeptanz des «Anderen» zeigen aber auch, dass mit der Forderung die Vermutung verbunden ist, Lehrpersonen mit Migrationshintergrund würden über ein besonderes Potenzial im Umgang mit migrationsbedingter Heterogenität verfügen, wovon sowohl die Schülerinnen und Schüler wie auch deren Eltern und die Lehrpersonenkollegien profitieren würden. Solchen Vermutungen eines besonderen Potenzials soll in diesem Beitrag in einer kritischen Weise nachgegangen werden. Dazu wird in einem ersten Teil (Abschnitt 2) ein Überblick über den gegenwärtigen Forschungsstand dargelegt. In einem zweiten Teil (Abschnitt 3) wird der Forschungsstand kritisch reflektiert. Dabei zeigt sich, dass sowohl politische Zuschreibungen wie auch wissenschaftliche Überprüfungen eines zugeschriebenen Potenzials Gefahr laufen, ein essentialistisches Kulturverständnis zu begünstigen und damit die Herstellung von Differenzlinien zu unterstützen. Im Kontrast dazu wird angeregt, Forschungszugänge vermehrt auf der Basis eines konstruktivistischen Kulturverständnisses zu entwickeln und etwa danach zu fragen, ob, und falls ja, inwiefern eine familiäre Migrationsgeschichte zur Entwicklung interkultureller Kompetenz beiträgt, sowie in der Folge auch danach, inwiefern die Lehrerinnen- und Lehrerbildung diese Entwicklung fördern kann.

2 Forschungsstand: Einige Hinweise, wenig empirische Evidenz

Im deutschsprachigen Raum gibt es bislang nur wenige wissenschaftliche Untersuchungen zum berufsspezifischen Potenzial von Lehrpersonen mit Migrationshintergrund. Im Schweizer Kontext hat sich Edelmann (2008) forschend mit der Thematik befasst und in Deutschland wurden unlängst die Ergebnisse zweier explorativer Studien publiziert (Georgi, Ackermann & Karakaş, 2011; Rotter, 2012).³ Auch in anderen Ländern, insbesondere in den USA sowie in Kanada, Grossbritannien und Australien, stösst das Thema auf Interesse und es erscheinen vereinzelte Publikationen zu Fra-

³ Weitere Erkenntnisse sind zudem aus einer Reihe laufender Forschungen zu erwarten, etwa an der Universität Bremen (Prof. Dr. Karakaşoğlu), der Johannes-Gutenberg-Universität Mainz (Prof. Dr. Bräu; vgl. Bräu, Karakaşoğlu & Rotter, 2013) und der Ruhr-Universität Bochum (Prof. Dr. Bellenberg).

gen der Spezifika von «immigrant teachers», «ethnic minority teachers», «teachers of color» oder «indigenous teachers», wobei sich in den genannten Begriffen auch die unterschiedlichen gesellschaftlichen Konstellationen und damit die beschränkte Übertragbarkeit der Forschungsergebnisse spiegeln.

Im Folgenden soll ein Überblick über den Forschungsstand⁴ gegeben werden, eingeteilt in zwei Bezugfelder im Handlungs- und Interaktionsradius von Lehrpersonen, wobei jeweils die Fragen zum Umgang mit migrationsbedingter Heterogenität im Zentrum stehen: Der erste Teil bezieht sich vor allem auf die Interaktionen zwischen Lehrpersonen mit Migrationshintergrund und ihren Schülerinnen und Schülern, der zweite Teil auf das weitere schulische Umfeld und damit vor allem auf Interaktionen zwischen den Lehrpersonen und den Eltern ihrer Schülerinnen und Schüler sowie auf Interaktionen innerhalb des Lehrpersonenkollegiums.

2.1 Lehrpersonen mit Migrationshintergrund im Umgang mit migrationsbedingter Heterogenität im Unterricht

Ein gelingender Umgang mit migrationsbedingter Heterogenität ist das Kernanliegen einer Pädagogik, die sich mit Interkulturalität befasst, und orientiert sich an den beiden Zielrichtungen der interkulturellen Erziehung aller Schülerinnen und Schüler einerseits und der Integration von Kindern und Jugendlichen mit Migrationshintergrund andererseits (COHEP, 2007b, S. 27). Im Bezug auf Erstere sollen hier die Fragen zu den heterogenitätsbezogenen Haltungen von Lehrpersonen mit Migrationshintergrund im Vordergrund stehen, in Bezug auf Letztere diejenigen zu den Leistungserwartungen von Lehrpersonen mit Migrationshintergrund sowie zu ihrer Rolle als Vorbilder.

Das Gelingen im Umgang mit migrationsbedingter Heterogenität ist nicht leicht zu definieren. Eine Möglichkeit, sich den Gelingensbedingungen anzunähern, ist es, nach den *Haltungen* der Lehrpersonen zu fragen, da solche grundlegenden Orientierungen wesentlich zu Erwerb und Anwendung der entsprechenden Handlungskompetenzen beitragen (Baumert & Kunter, 2006; vgl. auch Strasser & Steber, 2010). Haltungen von Lehrpersonen im Umgang mit migrationsbedingter Heterogenität wurden von Edelmann (2008) untersucht, die im Rahmen einer qualitativen Stichprobenuntersuchung in der Deutschschweiz ein breites Spektrum festgestellt und daraus eine Typologie entwickelt hat. Bei den 40 befragten Lehrpersonen – darunter 15 mit Migrationshintergrund – reicht das in sechs Typen aufgefächerte Spektrum von einer Haltung, die dazu führt, die Heterogenität weitgehend zu ignorieren («abgrenzend-distanziert»), bis hin zu einer Haltung, die gekennzeichnet ist durch eine anerkennende Berücksichtigung der Heterogenität auf der Basis ausgeprägter Eigeninitiative («individuell-synergieorientiert»)

⁴ Während der deutschsprachige Forschungsstand umfassend dargelegt wird, wurde der internationale Forschungsstand nur anhand englischsprachiger Publikationen erfasst. Zu erwarten sind hier noch Ergebnisse aus einer Forschungskoooperation unter der Leitung von Prof. Dr. Hanna Ragnarsdóttir an der University of Iceland, die unter dem Titel «Diverse Teachers for Diverse Learners» mit quantitativen und qualitativen Zugängen in Island, Norwegen, Finnland, Grossbritannien und Kanada forscht.

bzw. durch eine anerkennende Berücksichtigung der Heterogenität in Kooperation mit allen beteiligten Akteuren («kooperativ-synergieorientiert»). Edelmann stellt fest, dass Lehrpersonen mit Migrationshintergrund vom abgrenzend-distanzierten Typus in ihrem Sample nicht vertreten sind, während der individuell-synergieorientierte Typus nur bei Lehrpersonen vorzufinden ist, die entweder einen Migrationshintergrund haben oder in einer binationalen Familiensituation leben. Die Ergebnisse lassen es nachvollziehbar erscheinen, dass sich Lehrpersonen mit Migrationshintergrund angesichts ihrer Biografie von der Thematik kultureller Differenz kaum distanzieren können (Edelmann, 2008, S. 193 f.). Auch im österreichischen Kontext finden sich einige Hinweise darauf, dass sich «kulturell heterogene Verwandtschaftsbeziehungen» bei Lehrpersonen positiv auf einen situationsadäquaten Umgang mit kultureller Differenz im Unterricht auswirken können (Fillitz, 2003, S. 467). Fallstudien aus dem US-amerikanischen Raum zeigen, dass die Erfahrung der Minderheitenangehörigkeit einen starken Einfluss auf die Haltung einer Lehrperson haben kann und damit auch auf einen konstruktiven Umgang mit Multikulturalität (Smith, 2000). Ebenso wird erwähnt, dass ein solcher Erfahrungshintergrund zwar als Potenzial gesehen werden könne, dass die empirische Evidenz für einen Zusammenhang mit einem gelingenden Umgang mit kultureller Heterogenität aber auch für den US-amerikanischen Kontext weitgehend fehle (Tellez, 1999).

Das Anliegen der Integration von Kindern und Jugendlichen mit Migrationshintergrund ist eng verknüpft mit dem Anliegen der Chancengleichheit, dies vor dem Hintergrund wiederholt attestierter Benachteiligungen von Kindern und Jugendlichen mit marginalisierender sozioökonomischer Herkunft, wovon Kinder und Jugendliche mit Migrationshintergrund in der Schweiz überdurchschnittlich oft betroffen sind (Konsortium PISA.ch, 2011, S. 77). Die Benachteiligung zeigt sich darin, dass sich Schülerinnen und Schüler mit Migrationshintergrund häufiger in unteren Schullaufbahnen bewegen und dass sie öfter Sonderklassen zugewiesen werden, ohne dass diese Zuteilungen mit dem individuellen Leistungsstand der Schülerinnen und Schüler zu erklären wären. Fragen nach den Ursachen für diese Benachteiligungen werden kontrovers diskutiert (Beck, Jäpel & Becker, 2010; Kronig, 2007). Eine mögliche Ursache sind die sogenannten *Erwartungseffekte*. Es konnte gezeigt werden, dass sich die Leistungserwartungen von Lehrpersonen auf die Leistungen ihrer Schülerinnen und Schüler auswirken und dass solche Erwartungseffekte insbesondere auch für Selektionsentscheidungen mitbestimmend sind, dies oftmals zuungunsten von Kindern mit Migrationshintergrund (Kronig, 2003). Aus diesen Überlegungen heraus stellt sich die Frage, ob und inwiefern sich die Leistungserwartungen von Lehrpersonen mit Migrationshintergrund von denjenigen der Lehrpersonen ohne Migrationshintergrund unterscheiden. Oft wird auch gefragt, ob sich ein höherer Anteil an Lehrpersonen mit Migrationshintergrund positiv auf die schulischen Leistungen und damit auf eine verbesserte Chancengleichheit von Kindern und Jugendlichen mit Migrationshintergrund auswirken würde. Für den deutschsprachigen Raum gibt es dazu bislang lediglich Aussagen, die auf Selbsteinschätzungen beruhen. So gaben Lehrpersonen mit Migrationshintergrund an, sie würden hohe Leistungsanforderungen an Kinder mit Migrationshintergrund stellen (Edelmann, 2008; Georgi et

al., 2011, S. 268). Studien aus den USA zeigen ein sehr gemischtes Bild zu dieser Frage und reichen von dezidierten Aussagen zu positiven Wirkungen von «minority teachers» auf die Leistungen von «minority children» (Beady & Hansell, 1981; Meier, Stewart & England, 1989) bis zu kritischen Einschätzungen solcher als unterkomplex bezeichneten Zusammenhänge (Dee, 2004; Ehrenberg, Goldhaber & Brewer, 1995; Takei & Shouse, 2008). So schreibt Dee (2004, S. 59), die scheinbar zu beobachtenden Wirkungszusammenhänge in den Schulleistungen bei einem «racial match» zwischen Lehrenden und Lernenden sagten noch zu wenig darüber aus, ob, weshalb und inwiefern solche Wirkungen entstehen würden.

Eine weitere häufig genannte Erwartung – sowohl der Lehrpersonen mit Migrationshintergrund an sich selbst wie auch als von aussen an sie herangetragene – ist die Erwartung, als *Vorbild* zu wirken. So fand Edelmann (2008, S. 196) in den Äusserungen der von ihr befragten Lehrpersonen mit Migrationshintergrund eine starke Erwartung an sich selbst, für Kinder mit Migrationshintergrund ein Vorbild zu sein, dies im Sinne einer Wertschätzung für bilinguale Kompetenzen, einer gelebten transnationalen Identität und einer Absage an die Opferrolle. In deutschen Studien wird nebst den Selbstzuschreibungen auch die Auseinandersetzung mit den Fremdzuschreibungen thematisiert, da sich Lehrpersonen mit Zuwanderungsgeschichte herausgefordert sehen, auf die in jüngerer Zeit von bildungspolitischer Seite an sie herangetragenen Zuschreibungen zu reagieren. Rotter (2012) findet drei Typen möglicher Reaktionen: zum einen Lehrpersonen, die bereit sind, die an sie herangetragenen Erwartungen als Vorbild (und auch als «Kulturvermittler») zu erfüllen, bei denen also die Selbst- und Fremdzuschreibungen übereinstimmen; zum anderen Lehrpersonen, die zwar bereit sind, die an sie herangetragenen Funktionen zu erfüllen, gleichzeitig aber nicht darauf reduziert werden möchten, und schliesslich Lehrpersonen, die sich explizit gegen die an sie gestellten Erwartungen wehren, vor einer Kulturalisierung von Problemen warnen und sich selbst als Fachlehrpersonen positionieren (vgl. auch Georgi et al., 2011, S. 269; Karakaşoğlu-Aydın, 2000, S. 435; für den britischen Kontext Carrington, 2002 sowie Cunningham & Hargreaves, 2007; für den kanadischen Kontext Carr & Klassen, 1997). Ob und inwiefern von einer Vorbildwirkung ausgegangen werden kann, wäre Gegenstand vertiefter Auseinandersetzung. Bezieht man die Vorbildwirkung darauf, dass Stereotype herausgefordert werden, scheint der Zusammenhang zumindest plausibel zu sein: Da Lehrpersonen mit einer Zuwanderungsgeschichte das Bildungssystem selbst erfolgreich durchlaufen haben, können sie als Mitglied einer Gruppe, die stereotypisiert als schulisch leistungsschwach gilt (Kronig, 2003), als positives Beispiel wahrgenommen werden. Die negativen Stereotype «schulisch leistungsschwacher Immigrantenkinder» würden damit durch ein positives Beispiel herausgefordert, sowohl bei den Immigrantenfamilien selbst wie auch bei Angehörigen der Mehrheitskultur (vgl. Strasser & Steber, 2010, S. 114 f.). Dass es dazu auch gegenteilige Beispiele gibt, zeigt die US-amerikanische Fallstudie von Kern, Roehrig und Wattam (2012), in der das Unterrichtsverhalten eines Lehrers mit vietnamesischem Hintergrund untersucht und dabei festgestellt wird, dass dem Lehrer zwar eine erfolgreiche Bildungskarriere in den USA gelungen ist, dass

er aber *seinen* Weg des Erfolgs – Drill, extreme Leistungsorientierung und implizite Anpassung an die Regeln der machtausübenden Mehrheit – unreflektiert auch von all seinen Schülerinnen und Schülern erwartet, was in diesem Fall zu grossen Irritationen und letztlich zu einem unbefriedigenden Unterrichtserfolg führte.

2.2 Lehrpersonen mit Migrationshintergrund und ihre Rolle im schulischen Umfeld

Bezüglich der Rolle von Lehrpersonen mit Migrationshintergrund im weiteren schulischen Umfeld, etwa bezogen auf die Zusammenarbeit mit Eltern und Kollegium, zeigt die Literatur eine deutliche Rollenambivalenz. Einerseits besteht eine grosse Bereitschaft, sich mit dem biculturellen Hintergrund einzubringen, andererseits aber auch die Befürchtung, darauf reduziert zu werden.

Die Bereitschaft, sich mit dem biculturellen Hintergrund einzubringen, wird vielfältig beschrieben. Die von Edelmann (2008) in der Deutschschweiz befragten Lehrpersonen mit Migrationshintergrund geben an, der Zugang zu Eltern mit Migrationshintergrund würde ihnen leichtfallen, insbesondere, wenn diese über einen kulturell ähnlichen Hintergrund verfügten. Wenn Eltern die gleiche Herkunftssprache sprechen, werde diese in Kontakten häufig eingesetzt und es entstehe schnell eine Vertrauensbasis für offene Gespräche, wobei gesellige Anlässe als bedeutsam betrachtet würden, um immigrierte Eltern für die Kooperation mit der Schule zu gewinnen (Edelmann, 2008, S. 197 f.). Auch im deutschen Kontext wird von einem Vertrauensverhältnis zwischen Lehrpersonen und Familien mit Zuwanderungsgeschichte gesprochen (Georgi et al., 2011, S. 271). Berichtet wird zudem, dass der Lehrberuf von einigen Lehrpersonen als «sozio-kulturelle Mission» betrachtet werde mit dem Ziel, Familien mit Migrationshintergrund zu unterstützen (Karakaşoğlu, 2011, S. 126). Eine Schwierigkeit kann entstehen, wenn religiöse oder kulturelle Wertorientierungen bestimmter Einwanderergruppen den Prinzipien der Institution Schule widersprechen. Hier zeigt sich, dass die befragten Lehrenden mit Zuwanderungsgeschichte in Deutschland zwar bereit sind, kultursensibel auf Bedürfnisse von Zuwandererfamilien einzugehen, in der Tendenz aber die Werte der Schule vertreten und diese wenn nötig auch von den Eltern einfordern (Georgi et al., 2011, S. 271). Auch Studien aus den USA zeigen, dass Lehrende aus ethnischen Minderheiten eine hohe Bereitschaft haben, sich für Familien aus Minderheitengruppen einzusetzen und die Beziehung zu ihnen in einer kulturell angepassten Art zu gestalten (Gomez, Rodriguez & Agosto, 2008; King, 1993). Darüber hinaus wird eine gesellschaftskritische Haltung festgestellt. So schreiben Quiócho und Rios (2000), dass Lehrende aus Minderheiten eher über ein kritisches Bewusstsein in Bezug auf die ungleichen gesellschaftlichen Positionen verschiedener ethnischer Gruppen verfügten und dass dieses auf eigener Erfahrung mit Ungleichheit basiere. Weiter konstatieren sie, dass Lehrende aus Minderheiten öfter zur Aufdeckung institutioneller und persönlicher Diskriminierung bereit seien (vgl. auch Irvine, 1989; Su, 1997; für den britischen Kontext Cunningham & Hargreaves, 2007; für den kanadischen Kontext Carr & Klassen, 1997). Über die Bedeutung des Migrationshintergrunds in der Zusammenarbeit im

Lehrpersonenkollegium ist wenig bekannt. Die von Edelman (2008) in der Schweiz befragten Lehrpersonen mit Migrationshintergrund geben an, sie fühlten sich von ihren Kollegien umfassend akzeptiert. Ob diese Lehrpersonen ihren kulturellen Hintergrund als Ressource in die kollegiale Zusammenarbeit einbringen können, hängt hingegen davon ab, ob an ihrer Schule Teamstrukturen vorhanden sind und ob das Thema der kulturellen Heterogenität als wichtig erachtet wird (Edelman, 2008, S. 199; vgl. auch Georgi et al., 2011, S. 270).

Die Befürchtung und die Erfahrung, auf den Migrationshintergrund reduziert zu werden und eine Funktion als Fachperson für Migrationsfragen zugewiesen zu bekommen, ist allerdings ein sehr häufig thematisierter Aspekt der eigenen Rollenwahrnehmung, insbesondere im Zusammenhang mit Rollenerwartungen innerhalb eines Kollegiums (Edelman, 2008, S. 200; vgl. auch Georgi et al., 2011, S. 270; Rotter, 2012, S. 217). Es kann vermutet werden, dass die Bereitschaft von Lehrpersonen mit Migrationshintergrund, sich für interkulturelle Verständigung einzusetzen, damit einhergeht, dass gerade interkulturelle Konfliktsituationen als besonders hoher Stressor empfunden werden (Karakaşoğlu, 2011, S. 127). Erfahrungen von Rollenzuweisungen als «professional ethnic» werden auch im britischen (Blair & Maylor, 1993, S. 64), kanadischen (Carr & Klassen, 1997) und australischen (Santoro, 2007) Kontext gemacht. Zudem wird von noch stärkeren Zuschreibungen im Sinne mangelnder Anerkennung von Kompetenzen und diskriminierender Praktiken berichtet (für den britischen Kontext Blair & Maylor, 1993; Carrington, 2002; Cunningham & Hargreaves, 2007; für den US-amerikanischen Kontext Quirocho & Rios, 2000; für den kanadischen Kontext Carr & Klassen, 1997; für den australischen Kontext Santoro, 2007).

Der Überblick über den Forschungsstand zeigt eine breit gefächerte Palette verschiedener Aspekte eines möglichen Potenzials von Lehrpersonen mit Migrationshintergrund, die noch mit vielen Fragezeichen versehen ist und einen entsprechend grossen Forschungsbedarf offenlegt. Bei weiterer Forschung gilt es dabei zu beachten, dass Forschungszugänge problematische Zuschreibungen nicht unreflektiert reproduzieren. Im Folgenden soll daher zu einer Neuorientierung in der Thematik angeregt werden.

3 Offene Fragen: Anregung zu einer Neuorientierung in der Thematik

Erwartungen an Lehrpersonen mit Migrationshintergrund vonseiten der Bildungspolitik, der Lehrpersonenkollegien, der Eltern oder der Schülerinnen und Schüler, nicht zuletzt aber auch von Lehrpersonen an sich selbst, scheinen allgegenwärtig zu sein. Doch die gängigen Erwartungen, etwa Vertraute, Vorbilder, Brückenbauer, Kulturvermittler oder Übersetzer zu sein (Georgi et al., 2011, S. 19), können auf unterschiedliche Weise verstanden werden.

Einerseits können diese Erwartungen auf der Basis einer «Sie-für-sich-und-wir-für-uns»-Vorstellung gelesen werden: Die Vertrautheit mit einem «anderen» kulturellen Hintergrund wird dann verstanden als Befähigung, Menschen mit einem ebenso «anderen» Hintergrund besser zu verstehen, dies im Kontrast zu Lehrpersonen, die diesen «anderen» Hintergrund nicht genügend kennen. In einer solchen Vorstellung scheint es stabil existierende und voneinander abgrenzbare «kulturelle» Gruppen zu geben, die sich nach bestimmten «kulturell gegebenen» Regeln verhalten. So eine Vorstellung führt allerdings schnell in unlösbare Widersprüche, denn «kulturelle» Gruppen sind weder über Zeit stabil noch klar abgrenzbar oder homogen (Barth, 1969). Auch ist so ein Verständnis oft implizit mit der Vorstellung verbunden, «Kultur» habe eine Art übersubjektive Wesenheit, die sich objektiv erfassen und beschreiben lasse (Wicker, 2012, S. 228). Im Kontext einer solchen Vorstellung liegt die Vermutung nahe, es sei das Kennen einer «Kultur» oder das Wissen darüber, das für gelingende Kommunikation hauptsächlich bedeutsam sei. Der ihr zugrunde liegende essentialistische Kulturbegriff ist allerdings nur sehr beschränkt anwendbar.

Einem essentialistischen Kulturbegriff steht ein konstruktivistischer gegenüber. Die Konstruktion der Wirklichkeit (Berger & Luckmann, 1969) wird dabei als ein fortlaufender Prozess verstanden. «Kultur» bezieht sich auf dieser Verständnisbasis auf Verhaltensmuster, die durch Interaktionen entstehen und in weiteren Interaktionen auch laufend verändert werden. Dabei entstehen kollektive Muster, dies allerdings nicht im Sinne selbstständig existierender kultureller Gebilde, sondern in einem dynamischen Geschehen, das sich in Interaktionen immer wieder neu gestaltet (vgl. Bennett, 2005; Reckwitz, 2005; Wimmer, 2005). In einem solch dynamischen und vielfältigen Prozess stösst blosses – und möglicherweise vermeintliches – Wissen über die «Kultur» des Gegenübers schnell an Grenzen. Für gelingende, d.h. situationsadäquate Interaktion in «interkulturellen» Situationen sind vielmehr Kompetenzen erforderlich, die über blosses Wissen hinausgehen und die oft mit «interkultureller Kompetenz» benannt werden. Solche Kompetenzen sollen es gemäss Bender-Szymanski (2008, S. 204–205) ermöglichen, «auf Unvertrautes (Fremdes) nicht nur mit Inklusion oder Exklusion zu reagieren, sondern neue Erfahrungen auch ... kreativ ... zu verarbeiten». Dabei liesse sich anfügen, dass Individuen auf Fremdes nicht nur re-agieren, sondern im Sinne von «othering» (Reuter, 2002) eine Konstruktion als «anders» oder «fremd» in der Interaktion potenziell auch mitkreieren: Schulische Akteure mit ihren Handlungsrouninen sind selbst an der Herstellung von Heterogenität und von Differenzlinien beteiligt (Budde, 2012). Kompetentes Handeln in einem derart komplexen Vorgang erfordert daher einen hohen Grad an Selbstreflexion (Budde, 2012; Lanfranchi, 2008). Bennett (2005, S. 13) betont zudem die Bedeutung entwickelter interkultureller Sensibilität, die es ermöglicht, die Konstruktion kultureller Differenzen und Gemeinsamkeiten wahrzunehmen und ethnozentrische Perspektiven überwinden zu können.⁵

⁵ Für eine Einführung in die weitreichenden Debatten um adäquate Definitionen von «interkultureller Kompetenz» vgl. beispielsweise Deardorff (2009).

Eine politische Forderung nach «mehr Lehrpersonen mit Migrationshintergrund» läuft Gefahr, sich an einem essentialistischen Kulturverständnis zu orientieren. Es kann nicht davon ausgegangen werden, dass Lehrpersonen mit Migrationshintergrund im Umgang mit migrationsbedingter Heterogenität ein gewissermassen «kulturgegebenes» Potenzial haben. Zuschreibungen auf dieser Basis werden von Lehrpersonen mit Migrationshintergrund wohl zu Recht zurückgewiesen, da sie in zweifacher Hinsicht nicht zielführend sind: Zum einen wird statt zu einer kulturellen Öffnung der Schulen und einer kulturellen Sensibilisierung aller Beteiligten eher zu einer verstärkten Differenzlinienziehung und Segregation beigetragen (Georgi et al., 2011, S. 33; auch Dee, 2004). Zweitens bewirkt eine Rollenzuweisung an Lehrende mit Migrationshintergrund als «Sonderbotschafterinnen und Sonderbotschafter für Migrationsfragen» einen verengten Blick auf diesen einen Aspekt der «kulturellen» Herkunft, während Lehrende mit Migrationshintergrund wohl ebenso wie alle anderen Lehrpersonen in der ganzen Bandbreite ihrer Fähigkeiten und Kompetenzen wahrgenommen werden möchten und sollten.

Auch Forschungszugänge, die nach dem spezifischen Potenzial von Lehrpersonen mit Migrationshintergrund fragen, können Prämissen eines essentialistischen Kulturbegriffs enthalten. Eine Fragestellung etwa, die sich danach ausrichtet, ob und inwiefern Lehrpersonen mit Migrationshintergrund als «Brückenbauer» oder «Kulturvermittler» fungieren, scheint unreflektiert von essentialistisch verstandenen Grenzlinien zwischen «Heimischen» und «Migranten» auszugehen (vgl. Hummrich, 2009; Knappik & Dirim, 2012, S. 92). Auch die Unterscheidung in Lehrpersonen «mit» und Lehrpersonen «ohne» Migrationshintergrund – die wir zugegebenermassen hier auch selbst eingebracht haben – kann implizieren, dass der Migrationshintergrund und damit etwas «Kulturelles» als relevant vorausgesetzt wird. In der Folge wird mitunter versucht, etwas zu überprüfen, was bereits a priori als relevante Differenzkategorie in das Untersuchungsfeld gelegt wird und daher unter Umständen eher den Interessen der Forschenden entspricht als den Eigenschaften des Feldes (vgl. dazu Budde, 2012, S. 528) – woraus sich dann leicht zirkuläre Argumentationsstrukturen entwickeln.

Produktive Zugänge zur Thematik müssten, folgt man einem konstruktivistischen Kulturverständnis, soziale Grenzziehungen und Gruppenzugehörigkeiten als konstruiert und damit als situativ und performativ verstehen (Hormel, 2011, S. 100). Ein Forschungsdesign, das eine Unterscheidung in Lehrpersonen «mit» und Lehrpersonen «ohne» Migrationshintergrund enthält, sollte nicht von der Idee stabiler Gruppenzugehörigkeiten ausgehen, oder, um mit Brubaker (2007) zu sprechen, «Groupism» sollte vermieden werden. Stattdessen kann mit dem Begriff der «Kategorien» gearbeitet und danach gefragt werden, wann sie von wem und weshalb verwendet werden, wann also «ethnisierend» oder «kulturalisierend» gedeutet wird (vgl. Brubaker, 2007). Damit kann die Möglichkeit offengehalten werden dafür, dass sich die Unterscheidung in «mit» und «ohne» Migrationshintergrund je nach Kontext als irrelevant herausstellt und andere Faktoren als die mit einem Migrationshintergrund assoziierten eine bedeutsame Rolle

spielen. In der Konsequenz ist zu überlegen, wie die Analyse auch auf andere soziale Differenzkategorien ausgeweitet werden kann. Theoretische Überlegungen kommen dazu jüngst insbesondere aus der Intersektionalitätsdebatte. Walgenbach (2011, S. 113) regt diesbezüglich an, soziale Kategorien wie Gender, Ethnizität, Nation oder Klasse nicht isoliert oder additiv zu betrachten, sondern in ihrer Verwobenheit zu analysieren. Dabei ist zudem zu berücksichtigen, dass Differenzkonstruktionen von sozialen Macht- und Ungleichheitsverhältnissen durchzogen sind (Walgenbach, 2011, S. 120).

Für die inhaltliche Ausrichtung von Fragestellungen zur Thematik gibt es angesichts des Forschungsstands zahlreiche Möglichkeiten (vgl. Strasser & Steber, 2010, S. 118–121). Forschung, die der Lehrerinnen- und Lehrerbildung verpflichtet ist, richtet sich insbesondere danach, wie Potenzial gefördert und berufsspezifische Kompetenzentwicklung begünstigt werden kann. Bezogen auf die oben erwähnten «interkulturellen Kompetenzen» drängt sich die Frage auf, ob, und falls ja, inwiefern ein «Migrationshintergrund» in ihrer Entwicklung eine Rolle spielt. Es scheint durchaus plausibel zu sein, dass gerade Menschen mit einer familiären Immigrationsgeschichte vielfältigen Lebensentwürfen und kulturellen Mustern begegnen und dass sie aufgrund dieser Erfahrungen besonders gute Voraussetzungen dafür haben, interkulturelle Sensibilität und möglicherweise auch interkulturelle Kompetenz zu entwickeln. Hinweise darauf finden sich in Fallstudien aus dem australischen und amerikanischen Kontext (Quiocho & Rios, 2000; Santoro, 2007). Im Kontrast dazu zeigt das bereits erwähnte Fallbeispiel eines vietnamesischen Lehrers in den USA, dass die Erfahrung der Immigration und Minderheitenangehörigkeit kein Garant für eine produktive Entwicklung sind (Kern et al., 2012).

Für die Aus- und Weiterbildung von Lehrerinnen und Lehrern bedeuten diese Überlegungen, dass bei Lehrpersonen mit Migrationshintergrund möglicherweise ein Potenzial vorhanden ist, das bislang noch wenig beachtet worden ist. Bei der Beachtung dieses Potenzials kann es aber nicht darum gehen, durch eine Gruppeneinteilung in «mit» und «ohne» Migrationshintergrund Differenzlinien zu akzentuieren und diese dadurch mitzukonstruieren und zu Segregationen beizutragen. Vielmehr ist das Entwickeln interkultureller Kompetenz ein Anliegen in der Lehrerinnen- und Lehrerbildung, das alle Studierenden und Lehrpersonen gleichermaßen betrifft, bei dem aber Menschen mit einer familiären Migrationsgeschichte wegen ihres Erfahrungshintergrunds möglicherweise besonders günstige Voraussetzungen haben. Dass sich damit gleichsam von selbst interkulturelle Kompetenz und ein gelingender Umgang mit migrationsbedingter Heterogenität in Schule und Unterricht einstellen, ist keinesfalls selbstverständlich. Vielmehr setzt eine produktive Umsetzung von Erfahrungen einen langen und vielschichtigen Entwicklungsprozess voraus (Bender-Szymanski, 2008; Bennett, 2004), den es im Rahmen von Aus- und Weiterbildung zu unterstützen und zu fördern gilt.

Literatur

- Barth, F. (1969). Introduction. In F. Barth (Hrsg.), *Ethnic Groups and Boundaries: The Social Organization of Culture Difference* (S. 9–38). London: Allen & Unwin.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Beady, C.H. & Hansell, S. (1981). Teacher Race and Expectations for Student Achievement. *American Educational Research Journal*, 18 (2), 191–206.
- Beck, M., Jäpel, F. & Becker, R. (2010). Determinanten des Bildungserfolgs von Migranten. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer: Neue Ungleichheiten* (S. 313–337). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bender-Szymanski, D. (2008). Interkulturelle Kompetenz bei Lehrerinnen und Lehrern aus der Sicht der empirischen Bildungsforschung. In G. Auernheimer (Hrsg.), *Interkulturelle Kompetenz und pädagogische Professionalität* (S. 201–230). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bennett, M.J. (2004). Becoming Interculturally Competent. In J. Wurzel (Hrsg.), *Toward multiculturalism: A reader in multicultural education* (S. 62–77). Newton, MA: Intercultural Resource Corporation.
- Bennett, M.J. (2005). *Paradigmatic assumptions of intercultural communication*. Hillsboro: IDRInstitute.
- Berger, P.L. & Luckmann, T. (1969). *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*. Frankfurt am Main: Fischer.
- BFS [Bundesamt für Statistik]. (2012). *Bildungsstatistik 2011*. Neuenburg: BFS.
- Blair, M. & Maylor, U. (1993). Issues and Concerns for Black Women Teachers in Training. In I. Siraj-Blachford (Hrsg.), *«Race», Gender and the Education of Teachers* (S. 55–73). Buckingham: Open University Press.
- Bräu, K., Karakaşoğlu, Y. & Rotter, C. (2013). *Lehrerinnen und Lehrer mit Migrationshintergrund. Zur Relevanz einer Kategorie in Theorie, Empirie und Praxis*. Münster: Waxmann.
- Brubaker, R. (2007). *Ethnicität ohne Gruppen*. Hamburg: Hamburger Edition.
- Budde, J. (2012). Problematisierende Perspektive auf Heterogenität als ambivalentes Thema der Schul- und Unterrichtsforschung. *Zeitschrift für Pädagogik*, 58 (4), 522–540.
- Carr, P.D. & Klassen, T.R. (1997). Different Perceptions of Race in Education: Racial Minority and White Teachers. *Canadian Journal of Education*, 22 (1), 67–81.
- Carrington, B. (2002). Ethnicity, «role models» and teaching. *Journal of Research in Education*, 12 (1), 40–49.
- COHEP [Schweizerische Konferenz der Rektorinnen und Rektoren der Pädagogischen Hochschulen]. (2007a). *Empfehlungen zur Interkulturellen Pädagogik an den Institutionen der Lehrerinnen- und Lehrerbildung*. Bern: COHEP.
- COHEP [Schweizerische Konferenz der Rektorinnen und Rektoren der Pädagogischen Hochschulen]. (2007b). *Grundlagenbericht Interkulturelle Pädagogik in der Lehrerinnen- und Lehrerbildung in der Schweiz*. Bern: COHEP.
- Cunningham, M. & Hargreaves, L. (2007). *Minority Ethnic Teachers' Professional Experiences. Evidence from the Teacher Status Project*. Nottingham: University of Cambridge, Faculty of Education.
- Deardorff, D.K. (2009). *The Sage Handbook of Intercultural Competence*. Thousand Oaks: Sage.
- Dee, T.S. (2004). The race connection: Are teachers more effective with students who share their ethnicity? *Education Next*, 4 (2), 52–59.
- Edelmann, D. (2008). *Pädagogische Professionalität im transnationalen sozialen Raum. Eine qualitative Untersuchung über den Umgang von Lehrpersonen mit der migrationsbedingten Heterogenität ihrer Klassen*. Zürich: Lit.
- EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren]. (2000). *Interkulturelle Pädagogik in der Lehrerinnen- und Lehrerbildung. Zentrale Lernbereiche – Thesen – Literaturhinweise. Schlussbericht*. Bern: EDK.

- Ehrenberg, R.G., Goldhaber, D.D. & Brewer, D.J.** (1995). Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review*, 48 (3), 547–561.
- Fillitz, T.** (2003). *Interkulturelles Lernen. Zwischen institutionellem Rahmen, schulischer Praxis und gesellschaftlichem Kommunikationsprinzip*. Innsbruck: Studienverlag.
- Georgi, V.B., Ackermann, L. & Karakas, N.** (2011). *Vielfalt im Lehrerzimmer: Selbstverständnis und schulische Integration von Lehrenden mit Migrationshintergrund in Deutschland*. Münster: Waxmann.
- Gomez, M.L., Rodriguez, T.L. & Agosto, V.** (2008). Who are Latino prospective teachers and what do they bring to US schools? *Race Ethnicity and Education*, 11 (3), 267–283.
- Hormel, U.** (2011). Differenz und Diskriminierung: Mechanismen der Konstruktion von Ethnizität und sozialer Ungleichheit. In J. Bilstein, J. Earius & E. Keiner (Hrsg.), *Kulturelle Differenzen und Globalisierung* (S. 91–111). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hummrich, M.** (2009). Migration und Bildungsprozess. Zum ressourcenorientierten Umgang mit der Biographie. In V. King & H.-C. Koller (Hrsg.), *Adoleszenz – Migration – Bildung* (S. 103–120). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Irvine, J.J.** (1989). Beyond Role Models: An Examination of Cultural Influences on the Pedagogical Perspectives of Black Teachers. *Peabody Journal of Education*, 66 (4), 51–63.
- Karakaşoğlu-Aydın, Y.** (2000). *Muslimische Religiosität und Erziehungsvorstellungen: eine empirische Untersuchung zu Orientierungen bei türkischen Lehramts- und Pädagogik-Studentinnen in Deutschland*. Frankfurt am Main: IKO – Verlag für Interkulturelle Kommunikation.
- Karakaşoğlu, Y.** (2011). Lehrer, Lehrerinnen und Lehramtsstudierende mit Migrationshintergrund. Hoffnungsträger der interkulturellen Öffnung von Schule. In U. Neumann & J. Schneider (Hrsg.), *Schule mit Migrationshintergrund* (S. 121–135). Münster: Waxmann.
- Kern, A.L., Roehrig, G. & Wattam, D.K.** (2012). Inside a beginning immigrant science teacher's classroom: an ethnographic study. *Teachers and Teaching*, 18 (4), 469–481.
- King, S.H.** (1993). The Limited Presence of African-American Teachers. *Review of Educational Research*, 63 (2), 115–149.
- Knappik, M. & Dirim, İ.** (2012). Von Ressourcen zu Qualifikationen – Was es heisst, Lehrerin mit Migrationshintergrund zu sein. In K. Fereidooni (Hrsg.), *Das interkulturelle Lehrerzimmer* (S. 89–94). Wiesbaden: Springer VS.
- Konsortium PISA.ch.** (2011). *PISA 2009: Regionale und kantonale Ergebnisse*. Bern: BBT/EDK & Konsortium PISA.ch.
- Kronig, W.** (2003). Das Konstrukt des leistungsschwachen Immigrantenkindes. *Zeitschrift für Erziehungswissenschaft*, 6 (1), 126–141.
- Kronig, W.** (2007). *Die systematische Zufälligkeit des Bildungserfolgs. Theoretische Erklärungen und empirische Untersuchungen zur Lernentwicklung und zur Leistungsbewertung in unterschiedlichen Schulklassen*. Bern: Haupt.
- Lanfranchi, A.** (2008). Interkulturelle Kompetenz als Element pädagogischer Professionalität – Schlussfolgerungen für die Lehrerbildung. In G. Auernheimer (Hrsg.), *Interkulturelle Kompetenz und pädagogische Professionalität* (S. 231–260). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Meier, K.J., Stewart, J. & England, R.E.** (1989). *Race, class and education: The politics of second-generation discrimination*. Madison: University of Wisconsin Press.
- Presse- und Informationsamt der Bundesregierung.** (Hrsg.). (2007). *Der Nationale Integrationsplan. Neue Wege – Neue Chancen*. Berlin: Presse- und Informationsamt der Bundesregierung.
- Quijcho, A. & Rios, F.** (2000). The Power of Their Presence: Minority Group Teachers and Schooling. *Review of Educational Research*, 70 (4), 485–528.
- Reckwitz, A.** (2005). Kulturelle Differenzen aus praxeologischer Perspektive. Kulturelle Globalisierung jenseits von Modernisierungstheorie und Kulturesentialismus. In I. Srubar, J. Renn & U. Wenzel (Hrsg.), *Kulturen vergleichen* (S. 92–111). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Reuter, J.** (2002). *Ordnungen des Anderen: Zum Problem des Eigenen in der Soziologie des Fremden*. Bielefeld: Transcript.

- Rotter, C.** (2009). Lehrkräfte mit Zuwanderungsgeschichte: von der Politik umworben, von der Forschung vernachlässigt. *Tertium Comparationis*, 15 (1), 3–20.
- Rotter, C.** (2012). Lehrkräfte mit Migrationshintergrund: individuelle Umgangsweisen mit bildungspolitischen Erwartungen. *Zeitschrift für Pädagogik*, 58 (2), 204–222.
- Santoro, N.** (2007). «Outsiders» and «others»: «different» teachers teaching in culturally diverse classrooms. *Teachers and Teaching*, 13 (1), 81–97.
- Schweizerische Eidgenossenschaft.** (2001). *Botschaft zum Bürgerrecht für junge Ausländerinnen und Ausländer und zur Revision des Bürgerrechtsgesetzes*. Online verfügbar unter: <http://www.admin.ch/ch/d/ff/2002/1911.pdf> (13.04.2013).
- Smith, R.W.** (2000). The Influence of Teacher Background on the Inclusion of Multicultural Education: A Case Study of Two Contrasts. *Urban Review*, 32 (2), 155–176.
- Stiller, E. & Zeoli, A.** (2010). Lehrkräfte mit Zuwanderungsgeschichte – für einen ressourcenorientierten Perspektivwechsel in der Personalentwicklung. *Die Deutsche Schule*, 102 (4), 338–346.
- Strasser, J. & Steber, C.** (2010). Lehrerinnen und Lehrer mit Migrationshintergrund. Eine empirische Reflexion einer bildungspolitischen Forderung. In J. Hagedorn & L. Herwartz-Emden (Hrsg.), *Ethnizität, Geschlecht, Familie und Schule. Heterogenität als erziehungswissenschaftliche Herausforderung* (S. 97–126). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Su, Z.** (1997). Teaching as a profession and as a career: Minority candidates' perspectives. *Teaching and Teacher Education*, 13 (3), 325–340.
- Takei, Y. & Shouse, R.** (2008). Ratings in Black and White: Does Racial Symmetry or Asymmetry Influence Teacher Assessment of a Pupil's Work Habits? *Social Psychology of Education: An International Journal*, 11 (4), 367–387.
- Tellez, K.** (1999). Mexican-American Preservice Teachers and the Intransigency of the Elementary School Curriculum. *Teaching and Teacher Education*, 15 (5), 555–570.
- Walgenbach, K.** (2011). Intersektionalität als Analyseparadigma kultureller und sozialer Ungleichheiten. In J. Bilstein, J. Ecarus & E. Keiner (Hrsg.), *Kulturelle Differenzen und Globalisierung* (S. 113–130). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wicker, H.-R.** (2012). *Migration, Differenz, Recht und Schmerz*. Zürich: Seismo.
- Wimmer, A.** (2005). *Kultur als Prozess. Zur Dynamik des Aushandelns von Bedeutungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Autorin und Autor

Carola Mantel, M.A., wissenschaftliche Mitarbeiterin IZB, Pädagogische Hochschule Zug, carola.mantel@phzg.ch

Bruno Leutwyler, Prof. Dr., Co-Leiter IZB, Pädagogische Hochschule Zug, bruno.leutwyler@phzg.ch

Forschung an pädagogischen Hochschulen – Kurzberichte*

Bewegungsfördernde Spiel- und Lernumgebungen im Kindergarten

Evelyne Wannack, Regula Nyffeler und Thomas Balmer (Pädagogische Hochschule Bern)

Seit 2007 können Kindergartenlehrpersonen im Kanton Bern freiwillig am Projekt «Purzelbaum – mehr Bewegung im Kindergarten» teilnehmen. Es umfasst fünf Weiterbildungsmodule sowie über zwei Jahre verteilte Austauschtreffen. Die wissenschaftliche Begleitung fragte danach, wie Bewegungsangebote in die Spiel- und Lernumgebung im Kindergarten integriert und welche motorischen Fähigkeiten gefördert werden. Im Rahmen eines qualitativen Untersuchungsdesigns wurden die Erhebungsmethoden Raumanalyse, problemzentriertes Interview und videobasierte Unterrichtsbeobachtung eingesetzt. Die Stichprobe umfasste acht Kindergärten der zweiten Weiterbildungsstaffel.

Die Raumanalysen wurden vor Beginn und während der Umsetzungsphase durchgeführt, wobei die Einrichtungen und Materialien im Innenraum der Kindergärten mittels Skizzen und Fotos dokumentiert wurden. Hier zeigte sich, dass die Raumstruktur nur in zwei Kindergärten verändert wurde. Die Interviewdaten lassen hingegen verschiedene Veränderungen in der Unterrichtsgestaltung erkennen, z.B. dass die Lehrpersonen Bewegungsanteile gezielter, systematischer und konsequenter in den Kindergartenalltag integrieren. Die Analyse des Videomaterials zum Bewegungsangebot im freien Spiel machte verschiedene Bewegungsarrangements sichtbar. Bei den Arrangements mit konkreten Aufgabenstellungen üben und erweitern die Kinder ihre motorischen Fähigkeiten in höherem Masse als in sehr offenen Arrangements.

Die Ergebnisse der Studie werden nun im Rahmen von Weiterbildungsmodulen genutzt, um die Bewegungsangebote vertiefter zu analysieren, zu reflektieren und weiterzuentwickeln. (Weiterführende Informationen sind verfügbar unter: www.bewegungserziehung.phbern.ch.)

Publikation

Wannack, E. (2012). Bewegungsangebote und ihre Nutzung im freien Spiel des Kindergartens. In J. Košinár & U. Carle (Hrsg.), *Aufgabenqualität in der Grundschule* (S. 81–92). Baltmannsweiler: Schneider Verlag Hohengehren.

* Zusammengestellt von der Schweizerischen Koordinationsstelle für Bildungsforschung (SKBF), Entfelderstrasse 61, 5000 Aarau, www.skbf-csre.ch. Online-Datenbank und Suchportal: www.skbf-csre.ch/de/bildungsforschung/datenbank/.

Classroom-Management in der Schuleingangsstufe

Evelyne Wannack und Kirsten Herger (Pädagogische Hochschule Bern)

Welche Elemente des Classroom-Managements sind in der Schuleingangsstufe zu finden, wie werden sie im Rahmen der Unterrichtsgestaltung eingesetzt, und existieren auf dieser Stufe schon verschiedene Stile des Classroom-Managements? Diesen Fragen wurde im vorzustellenden Projekt mittels quantitativer und qualitativer Methoden nachgegangen. Als Erstes wurde eine Fragebogenerhebung bei Kindergarten- und Unterstufenlehrpersonen ($N=392$) durchgeführt. Anschliessend wurde kriteriengeleitet die Stichprobe für den qualitativen Teil der Studie gezogen. Bei sechs Kindergarten- und sechs Unterstufenlehrpersonen wurden je eine videobasierte Unterrichtsbeobachtung und ein fokussiertes Interview durchgeführt.

Anhand der Daten aus der Fragebogenerhebung wurden mittels Clusteranalyse drei Classroom-Management-Stile bestimmt: Lehrpersonen des regelorientierten Stils zeichnen sich durch eine hohe Strukturierung des Unterrichts aus, während prozedurenorientierte Lehrpersonen mehr Wert auf die Gestaltung der Spiel- und Lernumgebung legen. Lehrpersonen, die dem multidimensionalen Stil zugeordnet wurden, gaben an, offene und geführte Unterrichtssequenzen ausgewogen einzusetzen. Die qualitative Auswertung der Interviews erlaubte die Entwicklung eines Modells, welches aufzeigt, wie bestimmte Elemente und Funktionen des Classroom-Managements eingesetzt werden, um eine vielfältige Spiel- und Lernumgebung zu schaffen.

Für die Zusammenführung der quantitativen und qualitativen Daten wurde auf die drei zuvor definierten Stile des Classroom-Managements zurückgegriffen. Für jeden Stil wurde eine verdichtete Beschreibung erstellt. Welcher Stil zum Einsatz kommt, scheint von situativen Bedingungen abhängig zu sein. Als Ergebnis ihrer Studie können die Forscherinnen ein eigenständiges und ausdifferenziertes Konzept zum Classroom-Management für die Schuleingangsstufe vorlegen. Dieses wird nun einerseits durch die Verwendung in Aus- und Weiterbildung einer Prüfung unterzogen und andererseits soll es als Grundlage für weitere empirische Studien dienen. (Weiterführende Informationen sind verfügbar unter: www.schuleingangsstufe.phbern.ch.)

Publikationen

Wannack, E. (2012). Classroom Management und seine Bedeutung für die Gestaltung von Spiel- und Lernaktivitäten. In F. Hellmich, S. Förster & F. Hoya (Hrsg.), *Bedingungen des Lehrens und Lernens in der Grundschule* (S. 69–72). Wiesbaden: Springer VS.

Wannack, E. & Herger, K. (2011). Classroom Management als verbindendes Element zwischen Kindergarten und Grundschule. In D. Kucharz, T. Irion & B. Reinthoffer (Hrsg.), *Grundlegende Bildung ohne Brüche* (S. 127–132). Wiesbaden: VS Verlag für Sozialwissenschaften.

Intervention zur Förderung von Sozial- und Selbstkompetenzen in der Schule (InSSel)

Markus P. Neuenschwander und Nicole Frank (Pädagogische Hochschule der Fachhochschule Nordwestschweiz, Zentrum Lernen und Sozialisation)

Kinder und Jugendliche halten sich in der Schule nicht immer an die Regeln. Daraus können Situationen entstehen, welche für die Lehrpersonen, die Eltern und die Mitschülerinnen und Mitschüler, aber auch für die nicht regelkonformen Schülerinnen und Schüler selbst belastend sind. In diesem Kontext wurde im Herbst 2011 an sieben Schulen des Kantons Aargau das vom Zentrum Lernen und Sozialisation (ZLS) der Pädagogischen Hochschule FHNW entwickelte Programm «Intervention zur Förderung von Sozial- und Selbstkompetenzen in der Schule» (InSSel) eingeführt. Dieses Programm zielt darauf ab, die Sozial- und Selbstkompetenzen der Jugendlichen zu fördern und auf diese Weise zum Erreichen zweier wichtiger Lehrplanziele beizutragen. Dadurch soll den Jugendlichen in der Schule wie auch bei der Suche nach einer Lehrstelle geholfen werden.

Das Interventionsprogramm wurde in der Startphase (Herbst 2011 bis Mai 2012) evaluiert (ebenfalls vom ZLS), um zum einen Informationen über seine Wirkung zu erhalten und zum anderen Grundlagen für eine Weiterentwicklung zu schaffen. Die Ergebnisse zeigen, dass InSSel zu einer Verringerung der Unterrichtsstörungen beiträgt und die Sozial- und Selbstkompetenzen von Jugendlichen fördert. Weiter verbessern sich durch das Programm auch die Beziehungen der Schülerinnen und Schüler zu den Lehrpersonen und ihre Identifizierung mit der Schule. Aufgrund der ersten positiven Erfahrungen sind nun alle aargauischen Schulen eingeladen, das InSSel-Programm einzuführen und damit einen wirksamen Beitrag zur Förderung der Selbst- und Sozialkompetenzen ihrer Schülerinnen und Schüler zu leisten. (Weiterführende Informationen sind verfügbar unter: <http://www.fhnw.ch/ph/zls/inssel/>.)

Publikationen

Frank, N. & Neuenschwander, M.P. (2013). Förderung von Sozial- und Selbstkompetenzen – das Programm InSSel und seine Wirkungen. In M.P. Neuenschwander (Hrsg.), *Selektion in Schule und Arbeitsmarkt* (im Druck). Zürich: Rüegger.

Neuenschwander, M.P. & Frank, N. (2011). Förderung der Sozial- und Selbstkompetenzen in der Schule (InSSel): Beschreibung eines neuen Interventionsprogramms. *Sozialmagazin*, 36 (11), 43–49.

Der Zuteilungsentscheid beim Übertritt in die Sekundarstufe im Kanton Aargau: Einschätzungen von Schülerinnen, Schülern, Lehrpersonen und Eltern

Markus P. Neuenschwander und Rebekka Hartmann (Pädagogische Hochschule der Fachhochschule Nordwestschweiz, Zentrum Lernen und Sozialisation)

Im Zentrum dieser Untersuchung standen die Fragen, wie die Kinder, ihre Eltern und Lehrpersonen im Kanton Aargau das Übertrittsverfahren in die Sekundarstufe I erleben, mit welchen Belastungen und Ressourcen sie konfrontiert sind und wo subjektiv empfundene Stärken und Schwächen des Übertrittsverfahrens liegen. Um diese Fragen zu beantworten, wurden je sechs Jugendliche der 5. Klasse, ihre Eltern und Lehrpersonen zwischen April und Juni 2012 unter Beizug eines halbstrukturierten Interviewleitfadens zum Aargauer Übertrittsverfahren befragt. Die Interviews wurden danach mittels einer qualitativen Inhaltsanalyse ausgewertet. Einerseits wurde jedes Tripletts einzeln analysiert, um die Eigenheiten jedes einzelnen Falles prägnant und mehrdimensional darzustellen, andererseits wurden Quervergleiche zwischen den Interviews vorgenommen.

Die Ergebnisse zeigen, dass es bei den Lehrpersonen wie auch bei den Eltern Unsicherheiten im Hinblick auf die Umsetzung der Promotionsverordnung gibt. So zeigen sich Varianten bezüglich der Mitsprache des Kindes beim Selektionsentscheid, aber auch bezüglich der Gewichtung verschiedener Selektionskriterien (Noten in Haupt- und Nebenfächern, überfachliche Kompetenzen). Lehrpersonen unterscheiden sich in der Anzahl der Prüfungen und Lernkontrollen, die für einen validen Übertrittsentscheid als erforderlich erachtet werden. Trotzdem stimmt das von der Lehrperson empfohlene Niveau meistens mit den Erwartungen der Eltern überein. Dies belegen auch Daten von über 350 Kindern sowie ihren Eltern und Lehrpersonen, die im Rahmen des gleichzeitig laufenden, vom Nationalfonds finanzierten Forschungsprojekts «Wirkungen der Selektion» (WiSel) im Kanton Aargau gesammelt wurden. Alle Akteurinnen und Akteure (Kinder, Eltern, Lehrpersonen) berichteten über Belastungen während des Verfahrens, weil den Selektionsentscheiden für die Schulkarriere der betreffenden Kinder ein hoher Stellenwert eingeräumt wird.

Publikationen

Neuenschwander, M.P. (2012). Selektion in die Sekundarstufe I – Von Noten, Leistungen und der Schülerpersönlichkeit. *Schulblatt Aargau/Solothurn*, Nr. 23, 38.

Neuenschwander, M.P. (2013). Selektionsentscheidungen beim Übergang in die Sekundarstufe I und in den Arbeitsmarkt im Vergleich. In M.P. Neuenschwander (Hrsg.), *Selektion in Schule und Arbeitsmarkt* (im Druck). Zürich: Rüegger.

Geschlechtsuntypische Studienwahl: Weshalb Frauen Ingenieurwissenschaften studieren und Männer Primarlehrer werden

Christine Bieri Buschor, Simone Berweger, Andrea Keck Frei und Christa Kappler (Pädagogische Hochschule Zürich)

Diese Studie fragte in verschiedenen Phasen nach dem Studienwahlprozess von jungen Frauen, die sich für ein Studium in den technisch-ingenieurwissenschaftlichen Fächern interessieren, und von jungen Männern, die den Beruf des Volksschullehrers in Erwägung ziehen. Die Daten wurden zu drei Zeitpunkten erhoben: kurz nach der Profilwahl im Gymnasium, anlässlich von Aktivitäten, welche die Entscheidungsfindung erleichtern sollen (Studienwochen, Informationstage usw.) und anlässlich des Übergangs zwischen Gymnasium und Hochschule. Die Ergebnisse zeigen unter anderem, dass Frauen mit einem Interesse an MINT-Fächern eine Passion für Mathematik aufweisen müssen. Sie brauchen ihre besten Noten nicht unbedingt in diesen Fächern zu erzielen, sollten aber ein intrinsisches Interesse dafür aufweisen; d.h. es sollte sich dabei um ihre Lieblingsfächer handeln. Bei männlichen Jugendlichen hingegen, die eine Ausbildung zum Volksschullehrer in Betracht ziehen, erweisen sich neben der Freude an der Arbeit mit Kindern und Jugendlichen und dem Wunsch, Wissen zu vermitteln, auch extrinsische Motive wie Sicherheit als ausschlaggebend für die Studienwahl. Die Würfel scheinen recht früh zu fallen. Bei Fünfzehnjährigen haben sich geschlechtsstereotype Studien- und Berufsabsichten in der Regel schon stark verfestigt. Dass frühe Erfahrungen im ausserschulischen, aber unterrichtsnahen Bereich von Bedeutung sind, zeigt sich auch daran, dass oft jene jungen Männer Lehrer werden wollen, die einst in der freiwilligen Jugendarbeit, in Sportvereinen usw. Verantwortung für Jüngere übernommen haben. Die in jener Zeit gesammelten Lernerfahrungen und die dabei aufgebauten Selbstwirksamkeitsüberzeugungen sind von grosser Bedeutung für die Laufbahnentscheide.

Publikationen

Kappler, C., Keck Frei, A. & Bieri Buschor, C. (2012). «Männer im Lehrberuf sind äusserst wichtig, aber diese Überlegung war kein Grund für meine Berufswahl» – Eine qualitative Studie zur Bedeutung von Geschlecht bei der Berufswahl von angehenden Lehrern. *Schulpädagogik heute*, 3 (5), 1–20.

Kappler, C., Keck Frei, A. & Bieri Buschor, C. (2012). Männer sind gesucht! – Eine qualitative Studie zur Bedeutung von Geschlecht bei der Berufswahl angehender Lehrer. In D. Bosse, L. Criblez & T. Hascher (Hrsg.), *Reform der Lehrerbildung in Deutschland, Österreich und der Schweiz. Teil I: Analysen, Perspektiven und Forschung* (S. 211–231). Immenhausen: Prolog-Verlag.

Evaluation der Aufnahmeprüfungen für die Berufsmittelschulen des Kantons Luzern

Werner Wicki und Roland Künzle (Pädagogische Hochschule Luzern, Institut für Lehren und Lernen)

Der Eintritt in eine Berufsmaturitätsschule setzt im Kanton Luzern in der Regel das Bestehen einer Eintrittsprüfung voraus; sind gewisse Bedingungen erfüllt, ist aber auch ein prüfungsfreier Eintritt möglich. Die hier vorzustellende Arbeit fragte nach der Gültigkeit der Aufnahmeprüfungen und basierte auf einer fachdidaktischen Analyse der Prüfungsaufgaben sowie auf einer gründlichen statistischen Analyse aller Noten, welche 462 Schülerinnen und Schüler von vier Schulen, die im Jahr 2011 ein Berufsmaturitätszeugnis ausgestellt erhielten, bei den Eintritts- und bei den Schlussprüfungen erhalten hatten. Die Analyse der Noten erlaubt Antworten auf Fragen wie diejenige nach der relativen Leistungsfähigkeit der prüfungsfrei Eingetretenen und jener, die eine Aufnahmeprüfung bestehen mussten, oder auf die Frage der prognostischen Gültigkeit der Aufnahmeprüfungsergebnisse für das Erlangen der Berufsmaturität zwei Jahre später. Je 200 Deutsch- und Mathematikaufnahmeprüfungen wurden ausserdem auf der Ebene der Aufgaben mittels Rasch-Analysen untersucht. Nichts aussagen kann die Evaluation hingegen über die Gültigkeit der Aufnahmeprüfungen im Hinblick auf das Potenzial jener, die sie nicht bestanden haben.

Die prognostische Validität der Aufnahmeprüfungen wurde insgesamt bestätigt; auch zeigte sich erwartungsgemäss, dass die (aufgrund des Notendurchschnitts der abgehenden Schulen) prüfungsfrei Zugelassenen an der Abschlussprüfung in verschiedenen Fächern bessere Noten erzielten als die via Aufnahmeprüfung zugelassenen Schülerinnen und Schüler. Die Übereinstimmung mit dem Lehrplan war insgesamt gegeben; Spielraum für Verbesserungen besteht aber durchaus. Auf dem Niveau der Aufgabenstellungen schlugen die Studienautoren deshalb verschiedene Anpassungen vor.

Publikation

Wicki, W. & Künzle, R. (2012). *Evaluation der Aufnahmeprüfungen für die Berufsmittelschulen des Kantons Luzern* (Forschungsbericht 33). Luzern: Pädagogische Hochschule Zentralschweiz, Hochschule Luzern.

Selbstkonzepte bei Kindern in der Schuleingangsstufe: Zusammenhänge mit Leistungen und dem Wohlbefinden in der Schule

Nadja Abt Gürber (Pädagogische Hochschule St. Gallen und Universität Bern, Institut für Psychologie)

Ziel dieser im Rahmen einer Dissertation am Institut für Psychologie der Universität Bern durchgeführten Untersuchung war die Beantwortung dreier Fragen: 1) Lassen schulische Leistungen Vorhersagen bezüglich der ein Jahr später erhobenen Selbstkon-

zepte zu? 2) Lassen sich der «Big-Fish-Little-Pond»-Effekt (BFLPE) und das Internal/External-Frame-of-Reference-Modell (I/E-Modell) auf Kinder in der Schuleingangsstufe übertragen? (Gemäss dem BFLPE-Modell wird ein Kind ein umso niedrigeres akademisches Selbstkonzept haben, je leistungsstärker die Schulklasse ist, der es angehört. Das I/E-Modell nimmt darüber hinaus an, dass für das Selbstkonzept neben den sozialen Vergleichen, die für den BFLPE verantwortlich sind, auch intraindividuelle Vergleiche eine Rolle spielen, etwa wenn ein Kind sich sagt: «In Mathematik bin ich aber um einiges besser als in Deutsch!».) 3) Lassen sich Selbstkonzept und Wohlbefinden in der Schule empirisch voneinander abgrenzen?

Die für die Untersuchung verwendeten Daten stammen von 821 Kindern im Alter zwischen 6 und 7 Jahren aus den ostschweizerischen Schulversuchen zur Grund- bzw. Basisstufe (vgl. Projektmeldung 11:005), in denen es um eine Zusammenführung von Kindergarten und Primarschulunterstufe zu einem einzigen Zyklus ging. Mittels standardisierter Tests wurden die Sprach- und die Mathematikleistungen erhoben; die bereichsspezifischen Selbstkonzepte sowie das Wohlbefinden der Kinder wurden mittels Skalen erfasst. Zusätzlich dazu wurden auch die Eltern zum Wohlbefinden ihrer Kinder befragt. Korrelations-, regressions- und faktorenanalytische Auswertungen dieser Daten führten zu folgenden Erkenntnissen: 1) Das Selbstkonzept hängt signifikant positiv mit den Leistungen und dem Wohlbefinden in der Schule zusammen bzw. wird von früheren Leistungs- und Wohlbefindensausprägungen beeinflusst. 2) Der BFLPE (negativer Effekt der durchschnittlichen Klassenleistung auf das individuelle Selbstkonzept) ist klar replizierbar, das I/E-Modell (z.B. negativer Effekt der Sprachleistung auf das Selbstkonzept in Mathematik) dagegen nur eingeschränkt. 3) Das Selbstkonzept lässt sich empirisch durchaus vom Wohlbefinden in der Schule abgrenzen.

Publikation

Abt Gürber, N. (2011). *Bereichsspezifische Selbstkonzepte bei Kindern in der Schuleingangsstufe: Zusammenhänge mit Leistungen und Wohlbefinden in der Schule*. Unveröffentlichte Dissertation. Bern: Universität Bern, Philosophisch-humanwissenschaftliche Fakultät.

Datenbank der SKBF

Die Datenbank der SKBF umfasst Projekte der schweizerischen Bildungsforschung seit 1987 («Information Bildungsforschung») sowie Einträge zu Bildungsforschungsinstitutionen und Bildungsforschenden in der Schweiz. Diese Informationen sind auf der Website der SKBF (www.skbf-csre.ch) frei einsehbar. Durch die Aufnahme in die Datenbank (www.skbf-csre.ch/de/bildungsforschung/datenbank/) erfahren die Projekte eine nationale und internationale Verbreitung.

Im Mail-Versand «Information Bildungsforschung» werden regelmässig die neusten Projektmeldungen bekannt gemacht. Eine Übersicht über Forschungsprojekte (inklusive Abstract), die an pädagogischen Hochschulen angesiedelt waren oder die von Mitarbeitenden der pädagogischen Hochschulen durchgeführt worden sind, erscheint jeweils in den BzL.

Abgeschlossene Projekte mit Publikationen (gedruckt oder elektronisch) können eingereicht werden unter: info@skbf-csre.ch oder per Post an: SKBF, Entfelderstrasse 61, 5000 Aarau.

Kommentierte Literaturliste zur pädagogischen Diagnostik

Alois Buholzer und Martina Brülisauer

Die pädagogische Diagnostik erlebt zurzeit eine «Renaissance» (Hesse & Latzko, 2011, S. 15 [vgl. Buchbesprechung unten]), die sich unter anderem an der Vielzahl von Publikationen zu dieser Thematik ablesen lässt. Die folgende kommentierte Literaturliste soll Einblick geben in die vielfältigen Aktivitäten zur Konzeptualisierung der pädagogischen Diagnostik und zu ihrer empirischen Überprüfung im schulischen Kontext. Es handelt sich dabei um eine subjektiv getroffene Auswahl an Beiträgen, die sich mit der pädagogischen Diagnostik im Allgemeinen sowie mit der Erforschung der diagnostischen Kompetenz von Lehrpersonen im Speziellen befassen. Um einen illustrierenden Charakter bemüht, befinden sich darunter neben Übersichtswerken und langjährigen «Klassikern» auch Forschungsstudien zur Diagnosekompetenz von Lehrpersonen und Texte zu ausgewählten Aspekten (sonder)pädagogischer Diagnostik. Bei den Forschungsbeiträgen wurden aktuelle Studien aus dem deutschsprachigen Gebiet berücksichtigt, welche aus der Sicht des Autors und der Autorin einen wichtigen Beitrag zur empirischen Fundierung der Diagnosekompetenz bei Lehrpersonen leisten.

Zunächst werden Überblickswerke zur pädagogischen Diagnostik vorgestellt (in alphabetischer Reihenfolge), anschliessend folgen Forschungsstudien und Beiträge zu ausgewählten Themen der pädagogischen Diagnostik. Am Schluss folgt ein zusammenfassendes Fazit.

A) Übersichtswerke

Hesse, I. & Latzko, B. (2011). *Diagnostik für Lehrkräfte*. Opladen: Barbara Budrich.

«Diagnostik für Lehrkräfte» will eine theoretisch fundierte und gleichzeitig praxisnahe Einführung in die pädagogisch-psychologische Diagnostik liefern. Die Publikation richtet sich an Lehrpersonen wie auch an Studierende der Lehrerinnen- und Lehrerbildung. Ausgangspunkt der Ausführungen bilden die oftmals beklagten Mängel bei den diagnostischen Kompetenzen von Lehrpersonen. Mit der Publikation sollen Lehrpersonen Hilfen und Lernmöglichkeiten geboten werden, damit sie ihre Lücken schliessen und ihre diagnostischen Handlungsrountinen reflektieren und optimieren können. Studierende der Lehrerinnen- und Lehrerbildung erhalten eine solide Einführung in die pädagogisch-psychologische Diagnostik sowie Lernmöglichkeiten, um ihre Diagnosekompetenz weiterzuentwickeln.

Das in vier Teile gegliederte Buch geht eingangs auf das Konstrukt der diagnostischen Kompetenz und Beeinflussungstendenzen im Lehrpersonenurteil ein. Teil 2 orientiert

über die messtheoretischen Grundlagen und greift Themen wie die Unterscheidung von Alltags- und professioneller Diagnostik, die diagnostische Urteilsbildung sowie Gütekriterien und Methoden zur Datenerfassung auf. Im umfangreichen Teil 3 werden verschiedene Anlässe für das Erstellen von Diagnosen beschrieben. Exemplarisch werden dazu Instrumente und Hilfsmittel vorgestellt, die zur Klärung von diagnostischen Fragen beigezogen werden können. Abgerundet wird die Publikation in Teil 4 mit Aufgaben zum Reflektieren, Anwenden und Üben des erworbenen Wissens.

Es ist der Anspruch der Autorinnen, die pädagogisch-psychologische Diagnostik in Bezug auf konkrete Diagnoseanlässe im schulischen Kontext zu besprechen. Diese Forderung wird eingelöst, indem von Praxisbeispielen ausgegangen und anschließend aufgezeigt wird, wie auf Instrumenten basierend und entlang von transparenten Kriterien diagnostische Hypothesen erstellt werden. Es ist jedoch zu fragen, wie praxisnah die beschriebene pädagogische Diagnostik tatsächlich ist. Eine praxisnahe Diagnostik für Lehrpersonen müsste unseres Erachtens verstärkt Konzepte und Modelle aufgreifen, welche die Diagnoseprozesse in laufenden Lehr-Lern-Prozessen situieren. Lehrpersonen benötigen nicht nur eine Diagnostik *über* bestimmte Situationen, sondern eine Diagnostik, die *im* aktuellen Lehr-Lern-Prozess diagnostisch relevante Informationen gewinnt. Darüber hinaus wäre insbesondere eine stärkere Verbindung von diagnostischen Analysen und pädagogisch-didaktischem Handeln zu wünschen.

Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.

Karlheinz Ingenkamp hat 1968 das erste deutschsprachige Buch zur pädagogischen Diagnostik vorgelegt. Mittlerweile ist in Zusammenarbeit mit Urban Lissmann die 6. Auflage erschienen. Das sieben Kapitel umfassende Werk ist als Lehrbuch konzipiert und beinhaltet eine fundierte Einführung in die zentralen Themen der pädagogischen Diagnostik. Die Kapitel werden jeweils mit einer Zusammenfassung und Angaben für weiterführende Lektüre abgeschlossen.

Grundlegend für das Verständnis von pädagogischer Diagnostik ist die Umschreibung des diagnostischen Handelns, welches all jene Tätigkeiten umfasst, «durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmässiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren. Zur pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen sowie die mehr die gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder Erteilung von Qualifikationen zum Ziel haben» (S. 13). Im weiteren Verlauf des ersten Kapitels wird auf die in der Definition der pädagogischen Diagnostik angelegte Spannung der beiden Funktionen (Diagnostik zur Verbesserung der Lernens

und Diagnostik zur Erteilung von Qualifikationen) eingegangen. Weiter stellen Ingenkamp und Lissmann in diesem Grundlagenkapitel dar, wie sich das Verständnis und die Vorgehensweise der pädagogischen Diagnostik immer wieder gewandelt haben. Gemäss den Autoren zeigen die verschiedenen Positionen innerhalb der pädagogischen Diagnostik, dass gegensätzliche Vorgehensweisen keine einander ausschliessenden Alternativen sind, sondern komplementäre Möglichkeiten darstellen. Erst im Zusammenspiel der verschiedenen Positionen, Aufgaben und Zielsetzungen ergibt sich ein vollständigeres Bild der pädagogischen Arbeit im Schulalltag.

Im zweiten Kapitel geht es um die Methodenfrage, wobei zuerst Grundfragen des Messens sowie die methodischen Gütekriterien ausgeführt werden. Anschliessend werden Verhaltensbeobachtung, Befragungsmethoden und Testmethoden erklärt. Es folgen Ausführungen zu den Anwendungsgebieten im Bereich der Schulleistungsdiagnostik (Kapitel 3), der Entwicklungs-, Intelligenz- und Eignungsdiagnostik im Bildungswesen (Kapitel 4) und der Diagnostik sozialer und emotionaler Merkmale (Kapitel 5). Abschliessend folgt ein Kapitel zur Beratung im Bildungswesen sowie zu den Repräsentativerhebungen mit Schulleistungstests.

Der an Studierende der Lehrerinnen- und Lehrerbildung sowie Lehrpersonen gerichtete «Klassiker» führt fundiert und systematisch in das Thema der pädagogischen Diagnostik ein und hilft, die grundlegenden Herausforderungen im Bereich des Diagnostizierens und Beurteilens einzuordnen und zu reflektieren. Aufgrund der reichhaltigen Literatur, die im Lehrbuch verarbeitet worden ist, dient es ausserdem als Nachschlagewerk. Neuere Themen (wie beispielsweise Repräsentativerhebungen mit Schulleistungstests) werden aufgenommen, aktuelle Forschungsergebnisse (z.B. zur diagnostischen Kompetenz von Lehrpersonen) fehlen hingegen. Auch Bezüge, die aufzeigen, wie diagnostische Analysen zur adaptiven Gestaltung von Unterricht genutzt werden können, werden weitgehend ausgeblendet.

Jäger, R.S. (2007). *Beobachten, beurteilen und fördern! Lehrbuch für die Aus-, Fort- und Weiterbildung*. Landau: Verlag Empirische Pädagogik.

Das Lehrbuch «Beobachten, beurteilen und fördern» von Reinhold S. Jäger leitet die Leserin und den Leser dazu an, die eigene Beobachtungs- und Beurteilungskompetenz zu reflektieren und weiterzuentwickeln. Dazu werden in jedem Kapitel Beispiele zur Verdeutlichung, Übungen zur Vertiefung und Verfestigung sowie Aufgaben zur Reflexion und Wiederholung angeboten. Einleitend werden Probleme beim Beobachten und Beurteilen im schulischen Kontext ausgeführt sowie Ansätze diskutiert, die zur Vermeidung dieser Fehler beitragen. Daran anschliessend wird der diagnostische Prozess beschrieben. Danach folgen Überlegungen zum Förderbegriff und zur Umsetzung von Förderung mittels verschiedener Fördermodelle. Grundlegend für alle Vorgehensweisen ist die der Förderung zugrunde gelegte Diagnostik. Diese stellt fest, wo

Förderbedürftigkeit identifiziert wurde, und bildet die Basis für die Förderziele und -massnahmen. Anhand des Lesens, der Lesekompetenz und an mathematischen Kompetenzen wird aufgezeigt, wie diagnostische Aussagen und die Planung der Förderung miteinander verzahnt werden.

In den weiteren Kapiteln folgen Ausführungen zur sozialen, individuellen, sachlichen sowie zur fähigkeitsorientierten Bezugsnorm. Mithilfe der verwendeten Vergleichsgrössen nehmen Lehrpersonen nicht nur Beurteilungen, sondern auch Bewertungen vor. Als Bewertung bezeichnet Jäger «das Beziffern eines Verhaltens relativ zu einer Norm. Dieses Beziffern kann in Form von Zahlen oder mit Hilfe von verbalen Beschreibungen vorgenommen werden» (S. 133). In einem ausführlichen Kapitel werden verschiedene Methoden zur Leistungserfassung vorgestellt. Die Vor- und Nachteile der jeweiligen Methodik werden herausgearbeitet und diskutiert. Das letzte Kapitel greift die Notengebung auf und thematisiert das Validitätsproblem im Zusammenhang mit Schulnoten sowie die verschiedenen Vorgehensweisen zur Bewertung mit Schulnoten (intuitive Methoden, rationale Methoden, rational-empirische Methoden).

Das Buch ist – wie im Untertitel vermerkt – primär auf die Aus- und Weiterbildung von Lehrpersonen ausgerichtet. Die zahlreichen Übungen, Aufgaben und Illustrationen regen zum Nachdenken, Reflektieren und zur Vertiefung der behandelten Themen an. Diese reichhaltige Sammlung von Aufgaben – verbunden mit den theoretischen Bezügen – bietet für die Aus- und Weiterbildung zweifellos einen guten Fundus. Die Vielzahl an Querverweisen, Übungen und Anregungen bewirkt auf der anderen Seite, dass die Systematik im Buch teilweise verloren geht.

B) Ausgewählte Forschungsstudien

Dünnebier, K., Gräsel, C. & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung: Eine experimentelle Studie zu Ankereffekten. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 187–197.

Da die schulische Leistungsbeurteilung immer wieder in der Kritik steht, die Gütekriterien von Messungen in ungenügender Weise zu erfüllen, setzen die Autorinnen die Erkennung der Bedingungen von Urteilsfehlern ins Zentrum ihrer Studie. Die Forschungsstudie widmet sich der Frage, inwiefern Ankereffekte zu Urteilsverzerrungen in der schulischen Leistungsbeurteilung beitragen. Speziell wird untersucht, inwieweit und unter welchen Bedingungen Ankereffekte bei Urteilsleistungen wirksam werden.

Inwiefern Ankereffekte die Notengebung bei Expertinnen und Experten bzw. Novizinnen und Novizen beeinflussen, haben die Autorinnen in einem Quasiexperiment mit 128 Beteiligten untersucht. Die Expertengruppe bestand aus 64 erfahrenen Lehrpersonen und die Novizengruppe aus 64 Lehramtsstudierenden. Beiden Gruppen wurden

eine neutrale Schülerbeschreibung sowie je eine Deutsch- und eine Mathematiklausur vorgelegt, welche sie in einem ersten Schritt bewerten sollten. In einem zweiten Schritt wurde ihnen Ankerinformation mitgeteilt, welche sie ebenfalls zu beurteilen hatten. Ein dritter Schritt bestand darin, eine Prognose für die weitere Schullaufbahn des beschriebenen Schülers abzugeben. Somit wurde der Ankereffekt auch im Hinblick auf das Verarbeitungsziel (Klausurbewertung oder Prognose für eine Schullaufbahn) untersucht. Die Ergebnisse belegen, dass bei der Beurteilung der Deutschleistung durch Expertinnen und Experten nur dann ein Ankereffekt eintritt, wenn ein erster Eindruck gebildet werden sollte. Bei der Erstellung bedeutsamer Diagnosen gehen Expertinnen und Experten in der Regel merkmalsbasiert vor, indem sie alle relevanten Merkmale einer Person berücksichtigen, was folglich zu einer geringen Anfälligkeit für Fehltritte und Verzerrungen führt. Bei den Novizinnen und Novizen machte sich unabhängig vom Verarbeitungsziel ein Ankereffekt bemerkbar. Bei der Beurteilung der Mathematikleistung trat erstaunlicherweise sowohl bei den Expertinnen und Experten als auch bei den Novizinnen und Novizen und unabhängig vom Verarbeitungsziel ein Ankereffekt auf, der sich verzerrend auf die Urteilsbedingungen auswirkte.

In ihrem Beitrag zeigen die Autorinnen den Einfluss einer Ankerheuristik auf die Leistungsbeurteilung in Abhängigkeit von Expertise und Verarbeitungsziel auf. Des Weiteren scheint der Einfluss bei den Expertinnen und Experten aber auch abhängig vom jeweiligen Fach zu sein, wie das Eintreten des Ankereffekts bei der Mathematikleistung unabhängig vom Verarbeitungsziel bzw. in Abhängigkeit vom Verarbeitungsziel bei den Deutschleistungen verdeutlicht. Hinter diesem generell eintretenden Ankereffekt bei der Beurteilung der Mathematikleistungen vermuten die Autorinnen einen Zusammenhang mit der Motivation der Versuchspersonen, sich den anspruchsvoll erscheinenden Mathematikaufgaben zu widmen, oder auch, «dass die Lehrpersonen den beiden Unterrichtsfächern einen unterschiedlichen Einfluss auf die Prognose der Schullaufbahn zuschreiben» (S. 194). Allerdings müsste aus unserer Sicht den unterschiedlichen Ergebnissen zu beiden (promotionsrelevanten) Fächern nachgegangen werden und es müsste die Frage gestellt werden, wie die offensichtlich verschiedenen Heuristiken von Lehrpersonen einer Bearbeitung zugänglich gemacht werden könnten.

Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Münster: Waxmann.

In ihrer Dissertation leitet Karina Karst ein theoretisches Strukturmodell zur diagnostischen Kompetenz aus bedeutsamen Vorarbeiten ab. Wegleitend ist hierfür die folgende Umschreibung von diagnostischer Kompetenz: Es handelt sich um «eine auf das fachspezifische unterrichtliche Handeln bezogene Leistungsdisposition des Lehrers, die sich zunächst funktional auf die informellen bis semiformalen Diagnoseanforderungen in der Lehrertätigkeit bezieht» (S. 85). Die Lehrperson ist dabei gefordert, die Lernenden in Bezug auf deren kognitive Lernvoraussetzungen angemessen einzuschät-

zen, um neben der Herstellung einer förderlichen Unterrichtsatmosphäre auch eine kognitive Aktivierung der Schülerinnen und Schüler zu bewirken und eine effektive Unterrichtsgestaltung zu gewährleisten.

Hinter dieser Kompetenz stehen aufseiten der Lehrperson verschiedene Wissensstrukturen, die Karst als «Niveauwissen», «Heterogenitätswissen» und «Profilwissen» bezeichnet. Das Niveauwissen ist sowohl aufgaben- wie auch schülerbezogen. Das aufgabenbezogene Niveauwissen umfasst das Wissen der Lehrperson über den Schwierigkeitsgrad von Aufgaben oder Aufgabentypen. Das schülerbezogene Wissen bezieht sich auf die Kenntnisse der Lehrperson über das durchschnittliche Leistungsniveau der Klasse, die sie unterrichtet. Das Heterogenitätswissen über die Schülerinnen und Schüler bezeichnet die Fähigkeit der Lehrperson, das leistungsbezogene Heterogenitätsprofil zu erkennen und die relative Position einer Schülerin oder eines Schülers zu den Mitschülerinnen und Mitschülern zu bestimmen. Das Heterogenitätswissen bezüglich der Aufgaben besteht darin, das kognitive Potenzial einer einzelnen Aufgabe in Relation zu den anderen Aufgaben des gleichen Bereichs bestimmen zu können. Das Profilwissen schliesslich umfasst das sehr spezifische Wissen der Lehrperson über das Leistungsprofil einer einzelnen Schülerin oder eines einzelnen Schülers (S. 91 f.).

Die im Rahmen des PERLE-Projekts («Persönlichkeits- und Lernentwicklung von Grundschulkindern») erhobenen Schülerleistungen von 735 Schülerinnen und Schülern sowie die Urteile von 52 Mathematiklehrpersonen dienten als Datengrundlage für die Prüfung des entwickelten Kompetenzmodells. Aufgrund ihrer Analysen kommt Karst zum Schluss, dass der Begriff «diagnostische Kompetenz» irreführend sei, weil «mit ihm eine übergreifende Fähigkeit, sichere Urteile zu fällen, suggeriert wird. Denn eine solche liess sich aufgrund der Ergebnisse der Studie nicht empirisch nachweisen» (S. 215). Sie schlägt stattdessen vor, von «situationsspezifischen diagnostischen Kompetenzen» zu sprechen. Dieses Verständnis bildet die offensichtlich unabhängig voneinander wirkenden Dimensionen der Diagnosekompetenz besser ab und könnte als Ausgangspunkt für weitere Forschungen dienen. Zudem stellt sich die Frage, welche Konsequenzen zur Unterstützung von Lehrpersonen aus den Ergebnissen abgeleitet werden können.

Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess: Der Lehrer als flexibler Denker. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 175–187.

Der Artikel von Sabine Krolak-Schwerdt, Matthias Böhmer und Cornelia Gräsel widmet sich der Frage, inwiefern schülerbezogene Informationen einen Einfluss auf die von Lehrpersonen getätigte Leistungsbeurteilung haben. Ausgehend von der Theorie, dass Lehrpersonenurteile oft eine niedrige Validität, Objektivität und Reliabilität aufweisen, gelangen Aspekte wie soziale Herkunft, Geschlecht und Beliebtheit als Störfaktoren der Notengebung ins Blickfeld. Die Erklärung von Urteilsfehlern und die da-

raus resultierenden kategorialen Erwartungen leiten die Autorinnen und der Autor aus Forschungsarbeiten zur sozialen Kognition ab, wobei zwei Verarbeitungsarten unterschieden werden: Bei der kategoriengeleiteten Verarbeitung werden die individuellen Merkmale einer Person kaum berücksichtigt, da eine soziale Kategorie aktiviert wird, die folglich auch andere Personenmerkmale miteinbezieht. Bei der merkmaleitenden Verarbeitung werden die spezifischen Merkmale einer Person berücksichtigt, ohne dass dabei eine soziale Kategorie aktiviert wird (S. 176).

Um die zentrale Frage, «inwieweit unterschiedliche Ziele bei der Verarbeitung beobachtbarer Merkmale von Schülern ... die Aufmerksamkeitsallokationen und die Gedächtnisleistung für diese Merkmale beeinflussen» (S. 175), zu beantworten, wurden zwei Experimente durchgeführt. Beim ersten Experiment wirkten erfahrene Lehrpersonen als Expertinnen und Experten mit und beim zweiten Experiment Studierende in der Laienrolle. Beiden Gruppen wurden Fallbeschreibungen von Schülerinnen und Schülern vorgelegt. Sie wurden angewiesen, sich als Ziel entweder einen Eindruck über die Schülerin oder den Schüler zu bilden oder eine Vorhersage über die künftige Entwicklung der Schülerin oder des Schülers zu formulieren. Erhoben wurden sowohl die Lesezeit wie auch die anschliessende Reproduktion der Fallbeschreibungen. Die Ergebnisse machen deutlich, dass Expertinnen und Experten im Hinblick auf das Ziel der Eindrucksbildung die Informationen (schüler)kategoriengeleitet und die Verhaltensvoraussage merkmalsorientiert angingen. Laiinnen und Laien verwendeten hingegen unabhängig vom Ziel dieselbe kognitive Verarbeitungsstrategie. Es kann also nachgewiesen werden, dass sich Strategien der Informationsverarbeitung zwischen Expertinnen und Experten und Laiinnen und Laien unterscheiden. Erfahrene Lehrpersonen erweisen sich «als flexible Denker, deren Verarbeitung schülerbezogener Informationen als zielgeleiteter diagnostischer Prozess verstanden werden kann» (S. 184).

Die sorgfältig ausgearbeitete Forschungsstudie belegt, wie sich unterschiedliche Zielvorgaben auf die kognitiven Verarbeitungsprozesse (im Rahmen von Urteilsleistungen) auswirken. Inwieweit sich die Theorie sozialer Informationsverarbeitungsprozesse tatsächlich auf die Leistungsdiagnostik in der Schule übertragen lässt, bleibt mit der vorliegenden Studie unbeantwortet. Gemäss eigenen Angaben wollen die Autorinnen und der Autor dieser Frage in weiteren Studien anhand realer Fälle, bei denen zusätzliche Kontextinformationen zur Verfügung stehen, nachgehen.

C) Ausgewählte Aspekte der pädagogischen Diagnostik

Graf, U. & Moser Opitz E. (Hrsg.). (2007). *Diagnostik und Förderung im Elementarbereich und Grundschulunterricht*. Baltmannsweiler: Schneider Verlag Hohengehren.

Im ersten Teil zum Thema «pädagogische Diagnostik» erörtert Silvia-Iris Beutel die schwierige Balance zwischen Lerndiagnostik und Leistungsbeurteilung. Zwei Beispiele

verbaler Leistungsbeurteilung illustrieren die Anforderungen an Lernstandsdiagnosen und Leistungsbeurteilung. Argyro Panagiotopoulou greift das Thema der Lernbeobachtung im Unterrichtsalltag auf und spricht dessen Relevanz für die Aus- und Weiterbildung von Lehrpersonen an. Die Schlüsselsituationen des pädagogisch-diagnostischen Lernens in der Lehrerinnen- und Lehrerbildung sind das Thema des Artikels von Ulrike Graf, wobei sie unter Schlüsselsituationen Einblicke in Lernprozesse von Kindern oder Studierenden versteht. Die Bedeutung der Diagnostik und die pädagogischen Möglichkeiten bei Kindern mit ADHS werden im Beitrag von Kurt Czerwenka thematisiert.

Der zweite Teil zum Thema «bereichsspezifische Diagnostik» enthält Beiträge zum Sprach- und Schriftspracherwerb sowie zum mathematischen Lernen und zur interdisziplinären Sachbildung. Zum Sprach- und Schriftspracherwerb finden sich Beiträge zu sprachlichen Fähigkeiten von Schulanfängerinnen und Schulanfängern (Petra Schulz), zur Beobachtung der Literacy-Erfahrungen im Übergang vom Kindergarten zur Schule (Sven Nickel), zur Einschätzung des Sprachstands bei mehrsprachigen Kindern (Stefan Jeuk) und zur Beobachtung der Lernprozesse beim Orthografieerwerb (Cordula Löffler). Zum Thema «mathematisches Lernen» liefern Elisabeth Moser Opitz, Ursina Christen und Renate Vonlanthen Perler einen Artikel zum räumlichen bzw. geometrischen Denken von Kindern im Übergang vom Elementar- zum Primarbereich. Die freie mathematische Eigenproduktion und dazugehörige Lernprozesse sind das Thema von Kerensa Lee Hülswitt. Ein Projekt für Lehramtsstudierende an der Universität Erfurt (Heike Hahn & Dorothea Möller) widmet sich der Förderung von materialgeleitetem Lernen im Mathematikunterricht. Bildung und nachhaltige Entwicklung von Kindern (Meike Wulfmeyer) und das diagnostische Handeln im Sachunterricht (Simone Seitz) schliesslich sind die Themen der Beiträge zur interdisziplinären Sachbildung.

Der Band zeichnet sich durch eine Balance von theoretisch fundierten und praxisnahen Beiträgen im Kontext einer «fachdidaktisch orientierte[n] Diagnostik und Förderung» (S. 7) aus. Dabei werden nicht nur die individuellen Voraussetzungen der betreffenden Schülerin oder des betreffenden Schülers, sondern auch die kontextuellen Bedingungen miteinbezogen. Viele Beiträge im Buch verdeutlichen, dass diagnostische Prozesse nicht losgelöst von fachwissenschaftlichen und fachdidaktischen Konzepten erfolgen können, sondern auf das Engste damit verbunden sind. Diese Verbindung kann als besondere Stärke dieses Buches bezeichnet werden.

Horstkemper, M. (2006). Fördern heisst diagnostizieren. In G. Becker, M. Horstkemper, E. Risse, L. Stäudel, R. Werning & F. Winter (Hrsg.), *Friedrich Jahresheft XXIV: Diagnostizieren und Fördern* (S. 4–7). Seelze: Friedrich Verlag.

Marianne Horstkemper hält in ihrem Beitrag fest, dass pädagogische Diagnostik mehr ist als schulische Leistungsbeurteilung. Eingeschlossen in diagnostische Analysen sind

auch Aussagen über menschliches Verhalten und Erleben in Lehr-Lern-Prozessen, motivationale und emotionale Aspekte wie auch kognitive, motorische und sensorische Fähigkeiten. Mittels diagnostischen Handelns gilt es Klarheit über den erreichten Entwicklungsstand und seine entwicklungsfähigen Potenziale zu gewinnen. Dabei werden unterschiedliche Ziele verfolgt und diverse Strategien angewandt. Idealtypisch kann zwischen der Selektions- bzw. Auslesediagnostik und der Modifikations- bzw. Förderdiagnostik unterschieden werden. Diese Unterscheidung darf nicht als trennscharfe Aufteilung in zwei distinkte Klassen verstanden werden, sondern als polare Anordnung auf einem Kontinuum. Diese beiden Ausrichtungen pädagogischer Diagnostik werden einander auf der Grundlage verschiedener Kriterien gegenübergestellt und charakterisiert.

Horstkemper begründet, inwiefern Diagnosen für pädagogisches Handeln von hoher Relevanz sind: Sie dienen zum einen der Passung, weil der Unterricht nur dann den Voraussetzungen der Lerngruppe angepasst werden kann, wenn die Lehrperson Kenntnis über die Lernausgangslage der Kinder hat. Zum anderen wirken pädagogische Diagnosen als Frühwarnsystem, damit gefährdete Kinder rechtzeitig Unterstützung erhalten (Prävention). Weiter bilden sie die Grundlage der Intervention bzw. Förderung einzelner Kinder und ermöglichen die Förderung von Stärken (integrative Begabtenförderung). All diese Gründe belegen zudem, welche grosse Bedeutung Diagnosen insgesamt für das pädagogische Handeln von Lehrpersonen besitzen.

In ihrem Beitrag kritisiert Horstkemper zu Recht, dass die notenbasierte Leistungsbeurteilung (im Hinblick auf die Selektionierung) häufig einen zu grossen Stellenwert einnehme, und verweist gleichzeitig auf die Wichtigkeit, Diagnostik im Dienste der Förderung zu betreiben. Die Autorin fordert deshalb abschliessend, Lehrpersonen auf diese Aufgaben hin besser vorzubereiten. Zudem müssten in den Schulen ausreichend zeitliche Ressourcen bereitgestellt werden, damit diagnostische Aufgaben auch tatsächlich wahrgenommen werden können: «Diagnostische Kompetenz braucht Ressourcen. Zeit – sowohl für individuelle Analysearbeit als auch für kollegialen Austausch und Kooperation mit Spezialisten – ist dabei von fundamentaler Bedeutung» (S. 7).

Kany, W. & Schöler, H. (2009). *Diagnostik schulischer Lern- und Leistungsschwierigkeiten*. Stuttgart: Kohlhammer.

Das Buch von Werner Kany und Hermann Schöler soll Lehrpersonen in Primarschulen und an Sonderschulen sowie Studierenden der Lehrerinnen- und Lehrerbildung Unterstützung bei ausgewählten diagnostischen Fragen bieten. Im Vordergrund steht dabei die Diagnostik schulischer Lern- und Leistungsschwierigkeiten. Im ersten Kapitel werden die relevanten Fragestellungen einer pädagogisch-psychologischen Diagnostik und die damit einhergehenden Aufgaben vorgestellt. Im zweiten Kapitel werden die Grundlagen und Prinzipien der pädagogisch-psychologischen Diagnostik ausgeführt.

So bezeichnen die Autoren das Messen und das Vergleichen mit Normen als zentrale Aufgabe diagnostischen Handelns. Dazu wird auf mögliche Fehlerquellen, die sowohl bei der Erfassung wie auch bei der Beurteilung vorkommen, eingegangen. Im dritten und vierten Kapitel werden gängige Methoden sowie einzelne Verfahren zur Gewinnung diagnostischer Informationen mit Kurzdarstellungen beschrieben. Die Autoren haben aus der Fülle von Instrumenten solche ausgewählt, welche sie zur Beantwortung der aufgeworfenen Fragestellung als empfehlenswert erachten. Im Vordergrund stehen dabei Instrumente, die der Erfassung der Entwicklungsbereiche Sprache und Kognition dienen.

Im fünften Kapitel wird das Vorgehen bei der Erstellung eines pädagogisch-psychologischen Gutachtens zur Feststellung des sonderpädagogischen Förderbedarfs vorgestellt. Hierzu halten die Autoren verschiedene Empfehlungen fest: (1) Mit jedem Kind, das hinsichtlich eines sonderpädagogischen Förderbedarfs abgeklärt wird, soll ein Intelligenztest durchgeführt werden, um abschätzen zu können, über welche kognitiven Leistungsmöglichkeiten das Kind verfügt. (2) Das Kind ist prozessorientiert in seinem Lern- und Leistungsverhalten zu beobachten. (3) Schulleistungen sind vergleichend mit schulischen Leistungstests zu erheben, da Studien immer wieder verdeutlichen, wie kontextabhängig die Leistungsbeurteilung ausfällt. Diese Fehleinschätzungen können vermieden werden, wenn standardisierte Tests zum Einsatz gelangen. (4) Mit den Eltern und weiteren Bezugspersonen sind explorative und anamnestiche Gespräche durchzuführen, um Informationen zu Entwicklungsbedingungen und -voraussetzungen zu erheben. Diese Daten unterstützen den Diagnosefindungsprozess.

Das Buch bietet einen interessanten Orientierungsrahmen zur Diagnostik des sonderpädagogischen Förderbedarfs. Die Ausführungen zeigen, auf welche Aspekte bei der Bestimmung des sonderpädagogischen Förderbedarfs zu achten ist und wie ein Gutachten aufgebaut wird. Dazu stellen die Autoren eine Checkliste vor. Inwiefern die dargelegten Empfehlungen und Vorgehensweisen in der Praxis tatsächlich umgesetzt bzw. eingehalten werden können, ist aufgrund der Bedingungen im schulischen Kontext hingegen fraglich.

D) Fazit

Aus ihren unterschiedlichen Positionen tragen die vorgestellten Texte dazu bei, das vielfältige diagnostische Handeln von Lehrpersonen besser zu verstehen. Die Beiträge machen jedoch auch deutlich, dass die Vorstellungen von pädagogischer Diagnostik (hinsichtlich ihrer Funktionen, Aufgaben und Methoden) unterschiedlich sind und es zuweilen schwerfällt, die verschiedenen Denkrichtungen miteinander zu verbinden. Die in den Beiträgen vorhandenen wissenschaftlichen Grundpositionen lassen sich der pädagogischen Psychologie, der empirischen Unterrichtsforschung, der Sonderpädagogik, den Fachwissenschaften und den Fachdidaktiken zuordnen. Dabei lässt sich fest-

stellen, dass bei diagnostischen Fragen den Fachwissenschaften und Fachdidaktiken zunehmend eine grössere Beachtung geschenkt wird (vgl. Graf & Moser Opitz, 2007). Während langer Zeit befasste man sich im Zusammenhang mit der pädagogischen Diagnostik primär mit dem pädagogischen Urteil, Urteilsfehlern, der Urteilsgröße sowie den verschiedenen Methoden der Datenerfassung und -interpretation (vgl. Ingenkamp & Lissmann, 2008; Hesse & Latzko, 2011). Dass Fachinhalte und fachdidaktische Überlegungen stärker zu berücksichtigen sind, zeigt eine Reihe von interessanten Forschungsergebnissen. Diese belegen, dass es sich bei der Diagnosekompetenz nicht um eine einheitliche Grundkompetenz handelt, welche sich über die verschiedenen Fächer und diagnostischen Aufgaben hinweg generalisieren lässt, sondern dass sie spezifisch und stark mit den Fachinhalten verbunden ist (vgl. Karst, 2012; Dünnebier et al., 2009).

Für Lehrpersonen werden diagnostische Analysen dann interessant, wenn sie aus deren Ergebnissen Konsequenzen für ihr unterrichtliches Handeln ableiten können. Zuweilen fokussiert die pädagogische Diagnostik primär das einzelne Kind und insbesondere Kinder mit Auffälligkeiten oder Lern- und Leistungsproblemen (vgl. Kany & Schöler, 2009). Diagnostik dient jedoch nicht bloss dazu, einzelne Kinder zu fördern, sondern sie steht ebenso im Dienste der Unterrichtsgestaltung (vgl. Horstkemper, 2006). Diagnosen werden so als Teil von didaktischen Überlegungen verstanden, um die Lernangebote adaptiv an die Schülerinnen und Schüler anzupassen. Dazu allerdings, wie Unterrichtsplanung, diagnostische Analysen sowie Lernförderung und -beratung sinnvoll miteinander verbunden werden, geben die hier vorgestellten Beiträge insgesamt nur wenige Hinweise. Schliesslich ist auch zu fragen, wie angehende Lehrpersonen in der Ausbildung und Lehrpersonen im Schuldienst in ihren diagnostischen Fertigkeiten systematisch unterstützt werden können. Jäger (2007) oder Hesse und Latzko (2011) machen in ihren Beiträgen Vorschläge für die Aus- und Weiterbildung von Lehrpersonen. Insgesamt besteht jedoch ein erheblicher Bedarf, vertiefte Kenntnisse darüber zu gewinnen, wie Lehrpersonen in ihren diagnostischen Fähigkeiten durch adaptive Angebote der Aus- und Weiterbildung gezielt unterstützt werden können.

Autor und Autorin

Alois Buholzer, Prof. Dr., PH Luzern, Institut für Schule und Heterogenität (ISH), alois.buholzer@phlu.ch
Martina Brülisauer, Master of Science in Education, PH Luzern, Institut für Schule und Heterogenität (ISH), martina.bruelisauer@phlu.ch

Buchbesprechungen

Drollinger-Vetter, B. (2011). Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht. Münster: Waxmann, 358 Seiten.

Haben Sie mathematische Inhalte bloss *gelernt* oder haben Sie diese auch *verstanden*? Dass Lernen und Verstehen nicht gleichgesetzt werden können, liegt zwar auf der Hand, wird aber in der Unterrichtspraxis wie auch in der Didaktik nicht immer genügend sorgfältig berücksichtigt. Aber was bedeutet «Verstehen von mathematischen Inhalten» ganz genau? Und welche Folgen hat das für die Gestaltung von Lehr- und Lernprozessen? – Barbara Drollinger-Vetter zeigt in ihrer äusserst fundierten Dissertation am Beispiel des Satzes des Pythagoras auf, was Verstehen bedeutet, welche Anforderungen dabei an Schülerinnen und Schüler gestellt werden und welche Konsequenzen dies für die Gestaltung von Unterricht durch die Lehrpersonen hat.

Die Autorin geht der Frage nach, durch welche fachdidaktischen Qualitätsmerkmale das Verstehen eines mathematischen Konzepts unterstützt werden kann. Dazu legt sie im ersten Teil ihrer Arbeit in einem grossen Kapitel die theoretischen Grundlagen zu mathematischen Verstehensprozessen aus Sicht der Kognitionspsychologie und aus Sicht der Disziplin dar. Diese fruchtbare Darstellung mündet in ein weiteres theoretisches Kapitel, in welchem die Autorin am Beispiel der Satzgruppe des Pythagoras aufzeigt, wie im konkreten Inhaltsbereich Verknüpfungen und Sinn hergestellt werden können. Anschliessend stellt Drollinger-Vetter eine von ihr entwickelte Theorie des fachdidaktisch unterstützten Verstehensaufbaus vor, deren Kern sogenannte «Verstehenselemente» darstellen, die immer inhaltsbezogen sind und sich unmittelbar auf das zu verstehende mathematische Konzept beziehen. Diese Verstehenselemente lassen sich mit weiteren fachdidaktischen Qualitätsmerkmalen verbinden und erlauben es, zu untersuchen, inwiefern ein fachliches Konzept inhaltlich kohärent und bezüglich seiner Struktur klar entwickelt wird.

Im empirischen Teil der Arbeit untersucht Drollinger-Vetter in 38 Klassen aus der schweizerisch-deutschen Videostudie «Unterrichtsqualität, Lernverhalten und mathematisches Verständnis» (Oktober 2002 bis Dezember 2006) eine dreistündige Einführung in den Satz des Pythagoras und analysiert, wie Verstehen unterstützt und aufgebaut wird. Dabei zeigt sich, dass eine höhere fachdidaktische Qualität der Theoriephasen des Unterrichts einen Zusammenhang mit der Fachleistung der Schülerinnen und Schüler aufweist: Wo die fachdidaktische Qualität des Verstehensaufbaus hoch ist, ist auch die Leistung der Schülerinnen und Schüler hoch. Qualitativ hochstehende fachliche Unterstützung beim Entwickeln von mathematischen Konzepten trägt also Früchte! Verstehen ist somit nicht identisch mit Lernen und verlangt primär eine inhaltsbezogene, fachlich durchdachte und kohärente Bearbeitung.

Ein weiterer Befund, der sehr interessant ist, bezieht sich auf offensichtlich unterschiedliche Sichtweisen von Fachdidaktik und Allgemeiner Didaktik. So weist die Einschätzung der kognitiven Aktivierung aus der Perspektive der Allgemeinen Didaktik keinen Zusammenhang mit der eingeschätzten fachdidaktischen Qualität auf. Dieser Befund irritiert und ist gleichermaßen fruchtbar, bedeutet er doch, dass Verstehensqualität im Fachunterricht eine fachdidaktisch orientierte Analyse und damit einen engen Bezug zum Inhalt braucht. Gerade deshalb ist die Arbeit von Drollinger-Vetter in hohem Masse geeignet, um Fachdidaktik und Allgemeine Didaktik vermehrt miteinander ins Gespräch zu bringen.

Zu glauben, dass sich das Buch nur an Mathematikdidaktikerinnen und Mathematikdidaktiker richten würde, wäre ein Irrtum. Die Arbeit von Barbara Drollinger-Vetter wirft auch Fragen zum Verhältnis von Allgemeiner Didaktik und Fachdidaktik auf. Indem die Autorin auf die kleinsten inhaltlichen Bausteine des fachlichen Verstehens fokussiert und danach fragt, wie ein Inhalt – hier die Satzgruppe des Pythagoras – bearbeitet werden sollte, damit die Zusammenhänge verstanden und nicht nur gelernt werden können, stellt sich auch die Frage, ob und inwiefern Unterrichtsqualität aus einer allgemeinen Perspektive überhaupt angemessen analysiert werden kann. Diese Frage ist für professionell Unterrichtsbeobachtende und -beurteilende, für Dozierende an pädagogischen Hochschulen wie auch für Mentorinnen und Mentoren gleichermaßen von Bedeutung.

Aber auch das von Drollinger-Vetter vorgelegte Verstehensmodell (S. 190), von ihr wohl etwas zu bescheiden und einschränkend «Pythagoras-Verstehensmodell» genannt, ist in seiner hervorragenden Darstellung von drei unterschiedlichen Verknüpfungsebenen nicht nur für das Verstehen von mathematischen Konzepten bedeutsam. Drollinger-Vetter entwirft in diesem Modell drei Verknüpfungsebenen (die Verstehens-elemente als kleinste Bausteine eines Konzepts, die Formen seiner Repräsentation und die Vernetzung mit weiteren Konzepten innerhalb der Disziplin), die in einer grundsätzlichen Weise darlegen, was für einen verstehensbasierten Konzeptaufbau notwendig ist. Damit ist die Arbeit für alle Personen, die sich für das Verstehen von fachlichen Konzepten – nicht nur in der Mathematik – interessieren, von grossem Belang und hochinteressant.

Für Mathematikdidaktikerinnen und Mathematikdidaktiker, aber auch für Lehrpersonen der Sekundarstufe I ist die Arbeit darüber hinaus eine Fundgrube, wenn es darum geht, das Verstehen der Satzgruppe des Pythagoras zu fokussieren und sich darüber kundig zu machen, welche spezifischen Verknüpfungen für das Verstehen dieses fundamentalen Satzes der euklidischen Geometrie notwendig sind, wie unterschiedlich sie repräsentiert werden können und welche Vorstellungen von Schülerinnen und Schülern in diesem Zusammenhang vorherrschen und zu Schwierigkeiten führen können.

Esther Brunner, Prof. Dr., Dozentin für Mathematikdidaktik, Pädagogik und Sonderpädagogik an der Pädagogischen Hochschule Thurgau, esther.brunner@phtg.ch

Engi, P. (2012). Lehrplan 21 – Instrument in einem interkantonalen Bildungsorchester. Chur: Rüegger, 148 Seiten.

Das Volksschulwesen wird in der Schweiz weitgehend von den Kantonen gesteuert. Jeder Kanton verfügt in Schulfragen über eine in der Vergangenheit viel gepriesene Entscheidungshoheit sowie über eine eigene Gesetzgebung und erlässt unter anderem eigene Lehrpläne, Lektionentafeln und Promotions(ver)ordnungen. Bei grossen Schulreformen der vergangenen Jahrzehnte reifte zwar das Verständnis für kantonsübergreifende Lösungen. Sie sind aber gemäss gegenwärtiger Schweizer Rechtsetzung im Volksschulbereich gleichwohl nur über Konkordate bzw. interkantonale Vereinbarungen zu erzielen. Lehrziele und Lehrinhalte bzw. Lernziele und Lerninhalte können sich in enger räumlicher und staatspolitischer Verflechtung, beispielsweise auf die Schweiz als Staatsgebilde bezogen, nicht mehr grundlegend von Kanton zu Kanton unterscheiden. Die Entwicklung eines gemeinsamen Lehrplans ist somit als Idee nahelegend und als Projekt pädagogisch, didaktisch, entwicklungspsychologisch wie auch ökonomisch nutzbringend.

«Ein Lehrplan für alle – oder: für fast alle?» – Diese Frage stand als Titel über der Masterarbeit, die als Grundlage des vorliegenden Buches diente. Sie steht auch bildungspolitisch nach wie vor im Raum und kann dort nicht leichtfertig mit Ja oder mit Nein beantwortet werden. Aus der Perspektive des Verständnisses von Bildung und aus der Perspektive des Verständnisses von Unterricht darf sie aber aus Überzeugung mit Ja beantwortet werden. Wenn der Lehrplan 21 der Lehrperson genügend gestalterische und berufliche Kernaufträge betreffende Freiheiten einräumt, sie zugleich zur Verantwortung gegenüber pädagogischen, psychologischen und soziologischen Fragestellungen mahnt und sie zu einem stets entwickelbaren, auf die Bedürfnisse der Heranwachsenden ausgerichteten Unterricht verpflichtet, wird er weiterhin ein wichtiges normatives Leitinstrument der Schule bleiben. Darüber hinaus kann er das Potenzial vieler an Unterricht und Schule interessierter Expertinnen und Experten sowie Praktikerinnen und Praktiker in einen innovativen Entwicklungsprozess einbinden. Dass der Lehrplan 21 Qualitäten des Unterrichts sichern und entwickeln hilft, wird als grosse Hoffnung des Autors spürbar.

Das Buch beantwortet auf eine eigenständige und differenzierte Weise die Frage, welche Aspekte des wandelbaren gegenwärtigen Lehr- und Lernverständnisses sowie eines zu hinterfragenden Unterrichtsverständnisses in die Entwicklung einer neuen Lehrplangeneration einfließen. Dem Autor ist es dank grossen Aufwands und aufgrund seines Vorwissens gelungen, dieses komplexe Verhältnis zu durchdringen und aufschlussreiche Klärungen zu liefern. Im ersten Teil werden aus einer Bildungsperspektive drei Aspekte einer Analyse unterzogen: gemässigter Konstruktivismus, Selbstwirksamkeit und Schlüsselqualifikationen. Im zweiten Teil werden aus einer Unterrichtsperspektive die drei Aspekte Theorien des Lernens, Theorien des Lehrens und Steuerung des Bildungssystems fokussiert. Im abschliessenden Teil wird neben einer zusammenfas-

senden Beantwortung der Haupt- und Präzisionsfragestellungen eine persönliche Wertung des Autors mit einem Ausblick vorgenommen.

Die vorliegende Arbeit ist aus verschiedenen Gründen bemerkenswert. Sie ist erstens ein gelungenes Beispiel einer Theorie-Praxis-Verbindung. In einer Zeit einer mehr oder weniger offensichtlichen Theoriefeindlichkeit – in den Worten des deutschen Sängers Herbert Grönemeyer im Lied «Ich versteh»: «Es verläuft sich die Theorie / Sie kennt genau das Was und Wie / Sie spricht nur, damit's sie gibt / Sie verfliegt sonst ins Nichts» – und einer ebenso offensichtlichen Praxisorientierung oder -glorifizierung – «Alles wurde schon mal erlebt / Man denkt genau, wie alles geht / Und kennt auch jeden klugen Spruch / Jeden Kniff und jeden Dreh» (Herbert Grönemeyer, «Ich versteh») – sind Verbindungsleistungen zwischen Theorie und Praxis von entscheidender Bedeutung. Denn es ist und bleibt unbestritten: Neben «Wisdom of Practice» braucht es auch «Foundations of Education». Diese Verbindung ist dem Autor durch eine vertiefte Auseinandersetzung mit verschiedenen theoretischen Ansätzen, die eine tragfähige theoretische Basis für Antworten auf aktuelle bildungspolitische Fragen bilden, gelungen.

Die Arbeit ist zweitens ein gelungenes Zeugnis eines reflektierten und kritisch-kreativen Praktikers. In einer Zeit eines mediengesteuerten gesellschafts- und bildungspolitischen Diskurses ist Reflexion gefordert – vor allem dann, wenn unter dem Deckmantel eines angenommenen Handlungsdrucks Reflexion durch Reflex ersetzt und Reflexion zum Unwort des Jahres deklariert wird. In seiner Arbeit hat der Autor seine Erkenntnisse in Form überlegter, strukturierter und prägnant formulierter Aussagen auf den Punkt gebracht und die Ergebnisse in Bezug auf die Fragestellungen und die theoretischen Aspekte kritisch analysiert.

Die in diesem Buch dargelegten reflektierten theorie- und praxisorientierten Antworten sorgen für eine interessante aktualitätsbezogene Lektüre. Persönliches Interesse und eigene Leistung werden gut sichtbar; sie zeigen sich in eigenständigen, kritischen und selbstkritischen kreativen Ansätzen zur Lösung der gestellten Probleme und in der Beantwortung der beiden Hauptfragen «In welchem Masse werden in der Konstruktion des zu entwickelnden interkantonalen Lehrplans ausgewählte Elemente des Bildungs- und Unterrichtsverständnisses aufgegriffen und integriert?» und «Findet der Konstruktivismus als bedeutender pädagogisch-didaktischer Ansatz des 20. Jahrhunderts seinen Niederschlag im Lehrplan 21? Wenn ja, welchen?». Und nicht zuletzt gilt es festzuhalten, dass diese Arbeit durch ihre Leserfreundlichkeit überzeugt, die sich in einer verständlichen Sprache und einem hohen Niveau in den Formulierungen zeigt.

Hans Berner, Prof. Dr., Pädagogische Hochschule Zürich, hans.berner@phzh.ch

Gluck, M.A., Mercado, E. & Myers, C.E. (2010). Lernen und Gedächtnis. Vom Gehirn zum Verhalten. Heidelberg: Spektrum Akademischer Verlag, 628 Seiten.

Im alten Ägypten galt das Herz als wichtigstes Organ und machte das Wesen einer Person aus; das Gehirn hingegen war wenig bedeutsam. Dies hat sich markant geändert: In den Neurowissenschaften ist das Gehirn die zentrale Instanz für Lernen und Gedächtnis. Dies zu verstehen, stellt Lehrpersonen vor grosse Herausforderungen. Hierfür leisten Mark A. Gluck, Eduardo Mercado und Catherine E. Myers mit ihrem Lehrbuch Herausragendes. Als Professoren bzw. Professorin in diesem Fachgebiet kennen sie sich bestens in der Thematik aus und sind selbst experimentell-psychologisch bzw. neurowissenschaftlich forschend tätig. Ihr gut sechshundertseitiges, angenehm lesbares Lehrbuch umfasst dreizehn Kapitel, ein Glossar sowie ein Literatur- und Stichwortverzeichnis.

Zuerst wird Historisches präsentiert (Kapitel 1) und eine leicht zugängliche Einführung in die Neurowissenschaft von Lernen und Gedächtnis geboten (Kapitel 2). Deshalb ist eine Vertrautheit mit psychologischen Grundkenntnissen oder neurowissenschaftlichen Konzepten nicht notwendig. Detaillierter geht es dann in den folgenden Kapiteln zur Sache, die stets ausführlich Verhaltensprozesse, Gehirnsubstrate und kurz gefasst auch klinische Perspektiven thematisieren. In Kapitel 3 werden wichtige Grundlagen und die Unterschiede zwischen dem episodischen und dem semantischen Gedächtnis dargestellt. Wenn Sie daran interessiert sind, wie Sie lernen sollten, damit Sie das Gelernte besser behalten können, dann finden Sie hier Antworten. Tiefe Verarbeitung, Interferenz, falsche Erinnerungen, Gedächtnisverluste, die Interaktion von Hippocampus und Cortex bei der Gedächtniskonsolidierung und viele andere Themen mehr werden zur Sprache gebracht. Ebenso viel Wissenswertes und für das Unterrichten Relevantes wird in Kapitel 4 zum Fertigkeitengedächtnis referiert, beispielsweise Fertigkeitenerwerb und Fertigkeitenzerfall, Expertise und Talent, Übung und Transfer oder implizites Lernen.

Dem Arbeitsgedächtnis und der exekutiven Kontrolle ist Kapitel 5 gewidmet. «Ist das Arbeitsgedächtnis der Schlüssel zur Intelligenz?» – Dies ist nur eine der vielen Fragen, die sich bei dieser kognitions- und neuropsychologisch in extenso untersuchten Gedächtnisform stellen. Prominent wird Baddeleys Arbeitsgedächtnismodell erörtert. Zielsetzung und Planung, Aufgabenwechsel, Reizselektion und Reaktionsunterdrückung sind ebenso Thema wie flüchtige Erinnerungen, verschiedene Formen des Wiederholens und Verarbeitens, Verbesserung des Arbeitsgedächtnisses, Aktivität der Frontallappen bei Arbeitsgedächtnisaufgaben oder klinische Beispiele zu Schizophrenie und Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung. Die nachfolgenden drei Kapitel 6 bis 8 ermöglichen profunde Einblicke in die Geheimnisse des nicht assoziativen Lernens als Habituation, Sensibilisierung und perzeptives Lernen, der klassischen und der instrumentellen Konditionierung. Die theoretische Fundierung und die Bedeutung

dieser basalen Lernformen bei Mensch und Tier werden deutlich gemacht. Generalisierung, Diskrimination und die Repräsentation von Ähnlichkeit (Kapitel 9) sind weitere grundlegende Prozesse und werden mit dem Kategorienerwerb, Stereotypen, der Alzheimerkrankheit oder der Rehabilitation von Kindern mit Spracherwerbsstörungen in Zusammenhang gebracht.

«Emotionale Aktivierung erhöht die Wahrscheinlichkeit, dass Erinnerungen gespeichert werden» (S. 426). – Belege für diese Aussage finden sich in Kapitel 10 über emotionales Lernen und Gedächtnis. Hier erfährt man, aus welchen Komponenten Emotionen bestehen, wie sie das Speichern und den Abruf von Gedächtnisinhalten beeinflussen, wie Vermeidungslernen entsteht, welche Gehirnssubstrate beteiligt sind etc. In Kapitel 11 wird beschrieben, wie Menschen und Tiere durch Beobachtung lernen. Nach der Lektüre ist man über diese faszinierende und effiziente Form des Lernens gut informiert. Anschliessend ist von Lernen und Gedächtnis im Laufe des Lebens die Rede (Kapitel 12): «Macht klassische Musik Babys klüger?» (S. 478). Dieser und viele andere Aspekte kommen dabei ins Blickfeld, bevor mit einem instruktiven Kapitel zum Spracherwerb (Kapitel 13), das Fremdsprachenerwerb und Sprachleistungen bei Tieren einschliesst, der weite Forschungsbereich «Lernen und Gedächtnis» in diesem Band abgeschlossen wird.

Ziemlich umfassend, fundiert, mit konkreten Alltagsbeispielen und vielen farbigen Grafiken und Bildern illustriert, bietet das Werk eine präzise geistige Landkarte. Die übersichtliche Gliederung, eingeflochtene Zwischenfazits und Selbsttests sowie Konzept-Checks, wichtige Punkte und Schlüsselbegriffe am Ende jedes Kapitels fördern das Verstehen, Behalten und Anwenden. Eine nicht aufdringliche Didaktik regt dazu an, das neue Wissen auf Alltagssituationen zu beziehen oder kritisch zu denken. Dieses von Katja van den Brink geschickt ins Deutsche übersetzte Buch ist eine willkommene Lehrhilfe einer Autorenschaft mit Kompetenz in der Sache, ohne Trivialisierungen und ohne die vielen falschen Töne verbreiteter Lehrbücher. Das Lesevergnügen und die Lernwirksamkeit werden durch narrative Elemente gesteigert. Wünschenswert wäre eine zweite Auflage, um auch neuste Forschungsergebnisse zu repräsentieren.

Was kann nun einem Einsatz in der Lehrerinnen- und Lehrerbildung entgegenstehen? Leider die üblicherweise zu gering kalkulierte Lernzeit für diese zentrale Thematik und der Wunsch nach simplen Rezepten. Beides bedient Oberflächliches und verhindert Verstehen. Für die Verwendung dieses Lehrbuches spricht demgegenüber vieles – ja noch mehr: Für das notwendige professionelle Weiterlernen im Beruf kann damit ein tragfähiges Fundament erarbeitet werden, das den Zugang zu und das Interesse an anspruchsvoller Fachliteratur nachhaltig zu stimulieren vermag.

Christoph Schmid, Prof. Dr., Pädagogische Hochschule Zürich, Leiter Fachbereich Entwicklung und Berufsidealität, christoph.schmid@phzh.ch

Kreis, A. (2012). Produktive Unterrichtsbesprechungen. Lernen im Dialog zwischen Mentoren und angehenden Lehrpersonen. Bern: Haupt, 348 Seiten.

Ein Problem von Praktikumsbesprechungen besteht darin, dass diese weitgehend im Personenkreis von Studentin oder Student und Praxislehrperson eingekapselt bleiben. Die Studie von Annelies Kreis stellt etwas Öffentlichkeit her, indem sie einen Einblick in Unterrichtsgespräche und deren Lernertrag aufseiten der Studierenden gibt. Grundlage der Studie bilden 31 auf Video aufgezeichnete Vor- und Nachbesprechungen von Unterricht. Etwas später wurden die betreffenden Studierenden dazu befragt, ob sie in diesen Besprechungen etwas gelernt hätten. 16 Praxislehrpersonen hatten zuvor an einer umfangreichen Weiterbildung in fachspezifisch-pädagogischem Coaching teilgenommen und bildeten die Interventionsgruppe. Die übrigen 15 Praxislehrpersonen hatten (als Kontrollgruppe) vorgängig keine Weiterbildung besucht.

Bei der Ergebnisdarstellung fokussiert die Autorin auf den Unterschied zwischen jenen Gesprächen, die in der Wahrnehmung der Studierenden keinen Lernertrag ergaben, und jenen Gesprächen, bei denen die Studierenden das Gefühl hatten, etwas gelernt zu haben. Sowohl hinsichtlich der Art der Gesprächsführung als auch hinsichtlich der Inhalte liessen sich Differenzen feststellen. Hier in Kurzform einige interessante Ergebnisse:

- Für produktive Unterrichtsvor- und -nachbesprechungen muss genügend Zeit zur Verfügung stehen. Gespräche, in welchen gemäss der Selbstdeklaration der Studierenden wenig oder kein Lernen stattfand, dauerten weniger lang als jene mit einem Lernertrag.
- In der Grundausbildung erweitern Studierende durch Unterrichtsbesprechungen vor allem ihr allgemeindidaktisches Wissen sowie ihre Planungs- und Reflexionsstrategien. «Nur wenige Lernereignisse betreffen den Bereich des *Wissens über die Lernenden*. Die Ergebnisse zeigen auch, dass *fachspezifisch-pädagogisches* und *fachspezifisches Wissen*, das ein zentraler Bestandteil professioneller Kompetenz ist, und damit auch Lerninhalt zukünftiger Lehrpersonen sein sollte ..., in den selbst berichteten Lernerträgen relativ selten genannt wurde» (S. 302, Hervorhebung im Original).
- Produktive Unterrichtsgespräche zeichnen sich durch ein spezifisches, nämlich genrespezifisches Gesprächshandeln aus. So ist es wichtig, zwischen der Gesprächseröffnung und dem Gesprächsabschluss strukturierende Elemente wie Rekapitulationen, positive Bewertungen, Paraphrasierungen, Zusammenfassungen von Erkenntnissen, Problematisierungen oder einen Ausblick auf zukünftigen Unterricht einzubauen. Derartige Strukturierungselemente helfen, eine ko-konstruktive Atmosphäre entstehen zu lassen.
- Als besonders produktiv werden Unterrichtsbesprechungen dann eingeschätzt, wenn sie eine ko-konstruktive *Folgestruktur* aufweisen: «In Unterrichtsbesprechungen mit hohem Lernertrag werden Gesprächsgegenstände während bedeutend längerer

Dauer von beiden Interaktanten in einem ko-konstruktiven Dialog gemeinsam bearbeitet, als dies in Besprechungen mit geringem Lernertrag der Fall ist» (S. 322).

Diese Resultate können aus dem umfangreichen Untersuchungsteil abgeleitet werden (S. 119–285). Vorbereitet wird dieser durch zwei Kapitel, in denen theoretische Grundlagen ausgeführt werden. In einem ersten Teil geht es um die bisherigen Erkenntnisse der Forschungsgemeinschaft zum Lernen im Praktikum. Ausgehend von den verschiedenen empirischen Erkenntnissen der letzten zwanzig Jahre und den daraus folgenden Theorien zum Bereich des Praxislernens erarbeitet Annelies Kreis ein ko-konstruktivistisches «Kontextmodell des Lernens in der berufspraktischen Ausbildung» (S. 84). Wenn in der Dyade zwischen Praxislehrperson und Praktikantin oder Praktikant die Expertise der Praxislehrperson bewusst und kommunizierbar gemacht wird, dient dies der Praktikantin bzw. dem Praktikanten zur Weiterentwicklung des eigenen Unterrichts. Die gängigen Theorien gehen in dieses Kontextmodell ein, insbesondere Erkenntnisse der Expertise- und Kompetenzforschung, des situierten Lernens, der Verbindung praktischer und theoretischer Studieninhalte, des reflexiven Praktikums, des fachspezifisch-pädagogischen Coachings, des Drei-Ebenen-Modells, des Empowerment-Ansatzes und der videounterstützten Unterrichtsbesprechung.

Die zweite Grundlage, auf der Kreis aufbaut, besteht aus gesprächsanalytisch-pragmalinguistischen Theorien (ab S. 91). In diesem interessanten Kapitel werden Möglichkeiten und Grenzen von lernförderlichen Gesprächen aufgearbeitet, beginnend bei der Sprechakttheorie von Austin über Konversationsanalysen bis hin zu Theorien über die Folgestrukturen von Gesprächen. Psycholinguistische Studien über das Lernen in tutoriellen Dialogen und Möglichkeiten der Beschreibung von Gesprächen auf der Makroebene runden die Ausführungen ab. Dieses Kapitel enthält sehr viele interessante Erkenntnisse, die jedoch in einem expliziten Sinne im Untersuchungsteil leider nicht ausgewertet und fruchtbar gemacht werden.

Insgesamt ist die Lektüre des Buches von Annelies Kreis ein Muss für all jene, die auf der Hochschulstufe Unterrichtsbesprechungen mit Studierenden oder Schulungen von Praxislehrpersonen durchführen. Es ist theoretisch aktuell, gibt Einblicke in eine intelligente, videogestützte Forschungsarbeit und lässt zahlreiche Ableitungen für die Gestaltung der Berufsstudien zu. Kritisch festzuhalten ist hingegen, dass die Autorin eine multiperspektivische Zugangsweise zur Thematik gewählt hat. Es gelingt ihr nicht immer, all die verschiedenen Fäden zu ordnen und beisammenzuhalten.

Nach der Untersuchung von Jürg Schüpbach («Über das Unterrichten reden», 2007) liegt mit der Studie von Annelies Kreis eine weitere originelle Arbeit vor, welche das Besprechungs- und Reflexionsverhalten in Praktika analysiert und für die Lehrerinnen- und Lehrerbildung produktiv genutzt werden kann.

Michael Fuchs, Prof. Dr., Pädagogische Hochschule Luzern, michael.fuchs@phlu.ch

Kuckartz, U. & Rädiker, S. (Hrsg.). (2012). Erziehungswissenschaftliche Evaluationspraxis: Beispiele – Konzepte – Methoden. Weinheim: Beltz, 232 Seiten.

Der zu rezensierende Sammelband ist im Forschungsschwerpunkt «Beratung – Methoden – Evaluation. Bildung im Wandel» an der Philipps-Universität Marburg entstanden. Die beiden Herausgeber gaben den Autorinnen und Autoren der Beiträge zwei Orientierungspunkte vor:

- (1) die Fokussierung auf die für Evaluationen im Bildungsbereich konstitutive «Komplexität» und
- (2) die von der *Gesellschaft für Evaluation* publizierten «Standards für Evaluation» (im Anhang abgedruckt).

Teil I enthält sieben Beispiele erziehungswissenschaftlicher Evaluationspraxis aus Schule, Hochschule, Weiterbildung und Rehabilitationspädagogik, darunter die Folgenden: Stefan Rädiker sucht in den Selbstberichten von 50 testierten Weiterbildungsorganisationen nach Herausforderungen für die interne Evaluation. Bei Datenerhebungsinstrumenten zeigten sich starke Spannungen zwischen Standardisierung und Kontextualisierung. Die Verarbeitung der Daten und deren Nutzung für die Weiterentwicklung von Programm und Organisation falle schwer. Gross sei der Bedarf an angepassten, methodisch präzisen und sparsamen Erhebungsinstrumenten. Sabine Lauber berichtet über eine Evaluation von Schulungen zu lebenspraktischen Fähigkeiten von Blinden/Sehbehinderten. Hierzu absolvierte sie eine Schulung unter der Augenbinde. Während Interviews zu drei Erhebungszeitpunkten gut durchführbar waren, erwies sich der vorzulesende standardisierte Fragebogen als ermüdend – was nicht zuletzt einen Anlass darstellt, über den Mehrfacheinsatz solcher Instrumente auch bei Nichtblinden nachzudenken. Julia Schehl und Philip Boddin berichten über die Evaluation der dezentralisierten fachdidaktischen Ausbildung von Marburger Lehramtsstudierenden in 15 verschiedenen Fächern. Wie schätzen Studierende diese mit Blick auf die Schulpraxis ein? Welchen Nutzen sehen Dozierende für den Professionalisierungsprozess? Das Evaluationsbeispiel verdeutlicht, wie ein multimethodisches Vorgehen bei knappen Ressourcen zu relevanten Ergebnissen führen kann – im konkreten Fall zur Spannweite der Qualität der verschiedenen Fachdidaktiken. Heike Ackermann und Diana Rebecca Kosanovic evaluieren ein neues Format für die Lernzeitplanung einer gymnasialen Eingangsstufe (offener Anfang, Wochenplanarbeit, 80-Minuten-Einheiten) in Anbetracht individuell stark differierender Förderbedarfe. Sie arbeiten Forschungsergebnisse zur Lernzeitorganisation auf und integrieren diese in die fallorientierte Evaluation. Reflektiert wird dabei auch die Evaluationsrolle, unter anderem im Spannungsfeld der Kriterien «Leistung» versus «schulisches Wohlbefinden» oder zwischen den Aufgaben der Beurteilung und jenen der Organisationsentwicklung. Das multimethodische, überwiegend qualitative Design ermöglicht differenzierte Einschätzungen zu differenziellen Effekten für unterschiedliche Schülergruppen. Problematisiert wird allerdings der dafür erforderliche Zeitaufwand.

Teil II umfasst sechs Beiträge zu evaluationstheoretischen, -methodischen und -ethischen Aspekten sowie zu den dafür erforderlichen Kompetenzen: Susanne Maria Weber beschäftigt sich mit dem wertepositionierten Evaluationsmodell der «Transformativen Evaluation». Udo Kuckartz und Julia Busch untersuchen den Einsatz von Mixed-Methods (einschliesslich einer darauf bezogenen Inhaltsanalyse von Zeitschriftenartikeln). Claus Stefer lotet die Potenziale und Grenzen von Online-Fragebögen aus, während sich Uta-Kristina Meyer und Stefan Wißmach mit dem Einbezug von Kindern und Jugendlichen in die Evaluation befassen. Magdalena Chlond, Felicitas Bergmann und Heike Schnoor wiederum behandeln theoretisch begründete Messkonzepte für die Konstrukte «Krankheitsbewältigung» und «Lebensqualität» bei Parkinsonbetroffenen. Wie können diese Konstrukte nutzbringend in die Evaluation rehabilitationspädagogischer Massnahmen einbezogen werden? Verschiedene theoriebasierte Konstruktionen eines Krankheitsbildes sowie daraus als Erfolg versprechend abgeleitete Interventionen haben Auswirkungen auf die Methodenwahl (standardisierter Fragebogen versus erzählgenerierendes Interview).

Der Schlussbeitrag von Stefan Rädiker und Aljosha Niklesz fragt schliesslich danach, welche Kompetenzen Evaluierende benötigen und wie sie erworben werden können. Die Kompetenzbeschreibungen international verbreiteter Taxonomien werden auf verschiedene Evaluationsberufsprofile bezogen, z.B. «in der Hauptprofession Evaluation» oder «in einer anderen tätig». Verdeutlicht wird dabei, dass professionelle Evaluationskompetenz mehr erfordert, als ein Studium zum Lehramt oder der Erziehungswissenschaften bieten kann. Leistbar sei hier gleichwohl, Grundlagen und eine Motivation für eine eventuelle spätere Weiterbildung in Evaluation zu vermitteln.

Der durch die beiden Orientierungspunkte integrierte Sammelband weist viele Stärken auf. Die *Wertebasis* (z.B. Präferenzen für Wertschätzung, Lösungsorientierung oder Partizipation) der meist formativen, nutzungsfokussierten Evaluationen wird offengelegt. Dieser und weitere Punkte belegen die *Orientierungskraft der «Evaluationsstandards»*, die sich auch in den Reflexionen über Rolle, Methodenwahl oder Nutzungsvorbereitung zeigt. Das Buch enthält Gegenbeispiele zur ansonsten oft defizitären Nutzung von Forschungsergebnissen durch Evaluationen. Es eröffnet auch die Perspektive, praxisnahe Evaluation im Sinne von «evaluation-grounded research» für Forschung zu nutzen. Unbefriedigend ist hingegen die geringe Konzeptualisierung von «Komplexität», wozu einschlägige Evaluationsliteratur hätte herangezogen werden können.

Das Buch schliesst im deutschsprachigen Raum insofern die Lücke zwischen den wenigen Evaluationslehrbüchern und den zahlreichen Evaluationsberichten, als theoretisch und methodologisch begründet über Fallbeispiele der Evaluation reflektiert wird. Aus diesem Grund ist es für die Evaluationsaus- und -weiterbildung von Lehrpersonen äusserst wertvoll.

Wolfgang Beywl, Prof. Dr., Pädagogische Hochschule FHNW, wolfgang.beywl@fhnw.ch

Neuerscheinungen

Allgemeine Pädagogik und Schulpädagogik

- König, J. & Seichter, S.** (Hrsg.). (2013). *Menschenrechte. Demokratie. Geschichte. Transdisziplinäre Herausforderungen an die Pädagogik*. Weinheim: Beltz Juventa.
- Wagener, A.L.** (2013). *Partizipation von Kindern an (Ganztags-)Grundschulen. Ziele, Möglichkeiten und Bedingungen aus Sicht verschiedener Akteure*. Weinheim: Beltz Juventa.

Bildungs- und Unterrichtsforschung

- Andresen, S., Koch, N., Richter, M., Wolf, A.-D. & Wrobel, K.** (2013). *Familien als Akteure der Ganztagschule*. Weinheim: Beltz Juventa.
- Berninger, V.W.** (Hrsg.). (2012). *Past, present, and future contributions of cognitive writing research to cognitive psychology*. New York: Taylor & Francis.
- Dedering, K., Tillmann, K.-J., Goecke, M. & Rauth, M.** (2013). *Wenn Experten in die Schule kommen. Schulentwicklungsberatung – empirisch betrachtet*. Berlin: Springer.
- Jude, N. & Klieme, E.** (Hrsg.). (2013). *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung*. 59. Beiheft der Zeitschrift für Pädagogik. Weinheim: Beltz Juventa.
- Lange, D.** (2013). *Inhaltsanalytische Untersuchung zur Kooperation beim Bearbeiten mathematischer Problemaufgaben*. Münster: Waxmann.
- Lenhard, W.** (2013). *Leseverständnis und Lesekompetenz. Grundlagen – Diagnostik – Förderung*. Stuttgart: Kohlhammer.

Didaktik / Fachdidaktik / Mediendidaktik

- Ahrenholz, B. & Knapp, W.** (Hrsg.). (2012). *Sprachstand erheben – Spracherwerb erforschen*. Stuttgart: Klett.
- Bayrhuber, H., Harms, U., Muszynski, B., Ralle, B., Rothgangel, M., Schön, L.-H., Vollmer, H.J. & Weigand, H.-G.** (Hrsg.). (2012). *Formate fachdidaktischer Forschung: empirische Projekte – historische Analysen – theoretische Grundlegungen*. Münster: Waxmann.
- Bender, U.** (2013). *Ernährungs- und Konsumbildung. Perspektiven und Praxisbeispiele für den Hauswirtschaftsunterricht. Fachdidaktische Entwicklungen in Deutschland, Österreich und der Schweiz*. Bern: Schulverlag blmv.
- Brunner, E.** (2013). *Innermathematisches Beweisen und Argumentieren in der Sekundarstufe I. Mögliche Erklärungen für systematische Bearbeitungsunterschiede und leistungsförderliche Aspekte*. Münster: Waxmann.
- DePietro, P.** (2013). *Transforming Education with New Media. Participatory Pedagogy, Interactive Learning, and Web 2.0*. New York: Peter Lang.
- Hüttis-Graff, P. & Jantzen, C.** (Hrsg.). (2012). *Überarbeiten lernen – Überarbeiten als Lernen*. Stuttgart: Klett.
- Krauthausen, G.** (2012). *Digitale Medien im Mathematikunterricht der Grundschule*. Berlin: Springer Spektrum.
- Merklinger, D.** (2012). *Schreiben lernen durch Diktieren. Theoretische Grundlagen und Praxisbeispiele für Diktiersituationen*. Berlin: Cornelsen.
- Philipp, M.** (2012). *Motiviert lesen und schreiben. Dimensionen, Bedeutung, Förderung*. Seelze: Klett-Kallmeyer.
- Prasse, D.** (2012). *Bedingungen innovativen Handelns in Schulen: Funktion und Interaktion von Innovationsbereitschaft, Innovationsklima und Akteursnetzwerken am Beispiel der IKT-Integration an Schulen*. Münster: Waxmann.

- Rauch, M.** (2012). *Jugendliteratur der Gegenwart. Grundlagen, Methoden, Unterrichtsvorschläge*. Seelze: Friedrich Verlag.
- Reiss, K. & Hammer, C.** (2013). *Grundlagen der Mathematikdidaktik. Eine Einführung für den Unterricht in der Sekundarstufe*. Basel: Birkhäuser.
- Riegel, U. & Macha, K.** (Hrsg.). (2013). *Videobasierte Kompetenzforschung in den Fachdidaktiken*. Münster: Waxmann.
- Rosebrock, C. & Bertschi-Kaufmann, A.** (Hrsg.). (2013). *Literalität erfassen: bildungspolitisch, kulturell, individuell*. Weinheim: Beltz Juventa.
- Ulrich, W.** (Hrsg.). (2012). *Deutsche Sprache in Kindergarten und Vorschule*. Baltmannsweiler: Schneider Verlag Hohengehren.

Lehrerinnen- und Lehrerbildung / Weiterbildung von Lehrpersonen

- Hußmann, S. & Selter, C.** (Hrsg.). (2013). *Diagnose und individuelle Förderung in der MINT-Lehrerbildung. Das Projekt dortMINT*. Münster: Waxmann.
- Kahl, R.** (2013). *Treibhäuser der Zukunft – Wie in Deutschland Schulen gelingen. Dreifach DVD mit ausführlichem Booklet* (4. Auflage). Weinheim: Beltz.
- Rost, D.H.** (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung* (3., vollständig überarbeitete und erweiterte Auflage). Bad Heilbrunn: Klinkhardt.
- Terhart, E.** (2013). *Erziehungswissenschaft und Lehrerbildung*. Münster: Waxmann.
- Wahl, D.** (2013). *Lernumgebungen erfolgreich gestalten. Vom trägen Wissen zum kompetenten Handeln* (3. Auflage). Bad Heilbrunn: Klinkhardt.

Pädagogische Psychologie / Entwicklungspsychologie

- Hasselhorn, M. & Gold, A.** (2013). *Pädagogische Psychologie. Erfolgreiches Lehren und Lernen* (3., vollständig überarbeitete Auflage). Stuttgart: Kohlhammer.
- Langford, P.E.** (2013). *Vygotsky's Developmental and Educational Psychology*. New York: Taylor & Francis.
- Möller, J. & Köller, O.** (2013). *Pädagogische Psychologie 1. Schüler und Lehrer*. Stuttgart: Kohlhammer.
- Schermer, F.J.** (2013). *Lernen und Gedächtnis* (5., überarbeitete Auflage). Stuttgart: Kohlhammer.

Sonder- und Integrationspädagogik / Hochbegabung

- Döbert, H. & Weishaupt, H.** (Hrsg.). (2013). *Inklusive Bildung professionell gestalten. Situationsanalyse und Handlungsempfehlungen*. Münster: Waxmann.
- Feuser, G. & Kutscher, J.** (Hrsg.). (2013). *Entwicklung und Lernen*. Stuttgart: Kohlhammer.

Zeitschriftenspiegel

Allgemeine Pädagogik und Schulpädagogik

- Emmerich, M. & Hormel, U.** (2013). Equity oder die Individualisierung der Bildungsbenachteiligung. *Journal für LehrerInnenbildung*, 13 (1), 8–15.
- Fransson, G. & Grannäs, J.** (2013). Dilemmatic spaces in educational contexts – towards a conceptual framework for dilemmas in teachers work. *Teachers and Teaching: Theory and Practice*, 19 (1), 4–17.

Bildungs- und Unterrichtsforschung

- Lotz, M., Gabriel, K. & Lipowsky, F.** (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, 59 (3), 357–380.
- Reusser, K. & Pauli, C.** (2013). Verständnisorientierung in Mathematikstunden erfassen. Ergebnisse eines methodenintegrativen Ansatzes. *Zeitschrift für Pädagogik*, 59 (3), 308–335.
- Saalbach, H., Eckstein, D., Andri, N., Hobid, R. & Grabner, R.H.** (2013). When language of instruction and language of application differ: Cognitive costs of bilingual mathematics learning. *Learning and Instruction*, 26, 36–44. Online verfügbar unter: <http://www.sciencedirect.com/science/article/pii/S0959475213000108> (14.06.2013).
- van de Pol, J. & Elbers, E.** (2013). Scaffolding student learning: A micro-analysis of teacher-student interaction. *Learning, Culture and Social Interaction*, 2 (1), 32–41.

Didaktik / Fachdidaktik / Mediendidaktik

- Blatt, I. & Frahm, S.** (2013). Explorative Analysen zur Entwicklung der Rechtschreibkompetenz im Rahmen der NEPS-Studie (Klassenstufe 5–7). *Didaktik Deutsch*, Heft 34, 12–36.
- Francek, M.** (2013). A compilation and review of over 500 geoscience misconceptions. *International Journal of Science Education*, 35 (1), 31–64.
- Golonka, E.M., Bowles, A.R., Frank, V.M., Richardson, D.L. & Freynik, S.** (2012). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*. Online-Publikation, doi: 10.1080/09588221.2012.700315.
- Lane, R. & Coutts, P.** (2012). Students' alternative conceptions of tropical cyclone causes and processes. *International Research in Geographical and Environmental Education*, 21 (3), 205–222.
- Means, B., Toyama, Y., Murphy, R. & Baki, M.** (2013). The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. *Teachers College Record*, 115 (3), 1–47.
- Newcombe, N.S. & Stieff, M.** (2012). Six myths about spatial thinking. *International Journal of Science Education*, 34 (6), 955–971.
- Reinfried, S. & Tempelmann, S.** (2013). The impact of secondary school students' preconceptions on the evolution of their mental models of the greenhouse effect and global warming. *International Journal of Science Education*. Online verfügbar unter: <http://dx.doi.org/10.1080/09500693.2013.773598> (07.06.2013).
- Wastiau, P., Blamire, R., Kearney, C., Quittre, V., van de Gaer, E. & Monseur, C.** (2013). The Use of ICT in Education: a survey of schools in Europe. *European Journal of Education*, 48 (1), 11–27.
- Wouters, P., van Nimwegen C., van Oostendorp H. & van der Spek, E.D.** (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*, 105 (2), Online-Publikation, doi: 10.1037/a0031311.

Lehrerinnen- und Lehrerbildung / Weiterbildung von Lehrpersonen

- Goeze, A., Hettfleisch, P. & Schrader, J.** (2013). Wirkungen des Lernens mit Videofällen bei Lehrkräften. Welche Rolle spielen instruktionale Unterstützung, Personen- und Prozessmerkmale? *Zeitschrift für Erziehungswissenschaft*, 16 (1), 79–113.
- Hecht, P.** (2013). Selbstwirksamkeitsüberzeugungen im Berufseinstieg von Lehrpersonen. *Unterrichtswissenschaft*, 41 (2), 108–124.
- Kleinknecht, M. & Schneider, J.** (2013). What do teachers think and feel when analyzing videos of themselves and other teachers teaching? *Teaching and Teacher Education*, 33, 13–23.
- Kocher, M., Wyss, C. & Baer, M.** (2013). Unterrichten im Berufseinstieg – Wirkung der Praxiserfahrung und Vergleich mit erfahrenen Lehrpersonen. *Unterrichtswissenschaft*, 41 (2), 125–151.
- Philipp, A. & Kunter, M.** (2013). How do teachers spend their time? A study on teachers' strategies of selection, optimisation, and compensation over their career cycle. *Teaching and Teacher Education*, 35, 1–12.
- Saalfrank, W.-T., Weiss, S., Schramm, S. & Kiel, E.** (2013). Berufswunsch Geografielehrer/in – Motive und Selbstbild im Vergleich zu Studierenden anderer Unterrichtsfächer. *Geografie und ihre Didaktik*, 41 (2), 81–95.
- Seidel, T., Blomberg, G. & Renkl, A.** (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, 34, 56–65.

Pädagogische Psychologie / Entwicklungspsychologie

- Bäuerlein, K., Linkert, C., Stumpf, E. & Schneider, W.** (2013). Kurz- und langfristige Effekte außerfamiliärer Kleinkindbetreuung auf die kognitive und sprachliche Entwicklung unter besonderer Berücksichtigung der Betreuungsqualität. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45 (2), 57–65.
- Zaubauer, A.C.M., Gebauer, S.K., Retelsdorf, J. & Möller, J.** (2013). Motivationale Veränderung von Grundschulkindern in Englisch, Deutsch und Mathematik im Immersions- und Regelunterricht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45 (2), 91–102.

Sonder- und Integrationspädagogik / Hochbegabung

- Abmus, D. & Förster, F.** (2013). ViStAD – Erste Ergebnisse einer Video-Studie zum analogen Denken bei mathematisch begabten Grundschulkindern. *Mathematica Didactica*, 36, 45–65.
- Brunner, E. & Gyseler, D.** (2013). Heilpädagogische Förderung auch für hochbegabte Schülerinnen und Schüler? *Schweizerische Zeitschrift für Heilpädagogik*, Nr. 6, 36–42.
- Melzer, C. & Hillenbrand, C.** (2013). Aufgaben sonderpädagogischer Lehrkräfte für die inklusive Bildung: empirische Befunde internationaler Studien. *Zeitschrift für Heilpädagogik*, 64 (5), 194–202.

Impressum

Beiträge zur Lehrerbildung

<http://www.bzl-online.ch>

Redaktion

Vgl. Umschlagseite vorn.

Inserate und Büro

Kontakt: Heidi Lehmann, Büro CLIP, Schreinerweg 7, Postfach 563, 3000 Bern 9, Tel. 031 305 71 05, heidilehmann@bluewin.ch

Layout

Büro CLIP, Bern

Druck

Suter Print AG, Postfach, 3072 Ostermundigen

Abdruckerlaubnis

Der Abdruck redaktioneller Beiträge ist mit Genehmigung der Redaktion gestattet.

Abonnementspreise

Mitglieder SGL: im Mitgliederbeitrag eingeschlossen.

Nichtmitglieder SGL: CHF 60.-; Institutionen: CHF 70.-

Das Jahresabonnement dauert ein Kalenderjahr und umfasst jeweils drei Nummern.

Bereits erschienene Hefte eines laufenden Jahrgangs werden nachgeliefert.

Abonnementsmitteilungen/Adressänderungen

Schriftlich an: bzl@goffice.ch

Hier können auch Einzelnummern der BzL zu CHF 20.-/EUR 20.- (exkl. Versandkosten) bestellt werden (solange Vorrat).

Schweizerische Gesellschaft für Lehrerinnen- und Lehrerbildung (SGL)

<http://www.sgl-ssfe.ch>

Die Schweizerische Gesellschaft für Lehrerinnen- und Lehrerbildung SGL wurde 1992 als Dachorganisation der Dozierenden, wissenschaftlichen Mitarbeitenden und Assistierenden der schweizerischen Lehrerinnen- und Lehrerbildungsinstitute gegründet. Die SGL initiiert, fördert und unterstützt den fachlichen Austausch und die Kooperation zwischen den pädagogischen Hochschulen bzw. universitären Instituten und trägt damit zur qualitativen Weiterentwicklung der Lehrerinnen- und Lehrerbildung bei. Sie beteiligt sich an den bildungspolitischen Diskursen und bringt die Anliegen der Lehrerinnen- und Lehrerbildung in den entsprechenden Gremien ein.

Friedrich-Wilhelm Schrader Diagnostische Kompetenz von Lehrpersonen

Constance Karing und Cordula Artelt Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften

Anna-Katharina Praetorius Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter – Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis

Alois Buholzer und Sandra Zulliger Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung

Guido McCombie und Damian Länge Der Educational Profiler: Ein Instrument für die Diagnose der pädagogischen Bezugssysteme angehender Lehrpersonen

Andreas Helmke und Gerlinde Lenske Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung

Carola Mantel und Bruno Leutwyler Lehrpersonen mit Migrationshintergrund: Eine kritische Synthese der Literatur

Alois Buholzer und Martina Brülisauer Kommentierte Literaturliste zur pädagogischen Diagnostik