

Pädagogische Diagnostik im Schulalltag

Stefanie van Ophuysen und Katrin Lintorf

„Guten“, das heißt akkuraten Diagnosen von Schülermerkmalen wird inzwischen ein hohes Gewicht für eine valide und faire Leistungsbeurteilung und darauf aufbauend für eine adäquate Förderung der Schülerinnen und Schüler beigemessen. Dennoch fehlt es bislang an Überlegungen zum diagnostischen Handeln als Bedingung der Diagnosequalität. In dem vorliegenden Beitrag wird zunächst der Begriff der pädagogischen Diagnostik eingegrenzt, um darauf aufbauend ein heuristisches Prozessmodell zu entwickeln, das das alltägliche Diagnostikhandeln in fünf Phasen untergliedert (Identifikation von Kriterien; Operationalisierung; Diagnoseplanung und -durchführung; Informationsspeicherung; Informationsintegration). Basierend auf empirischen Befunden wird herausgearbeitet, wie diagnostisches Handeln so gestaltet werden kann, dass es mit hoher Wahrscheinlichkeit in einer qualitativ hochwertigen Diagnose resultiert.

1. Einleitung

Verschiedene Befunde aus internationalen Schulleistungsstudien haben in Deutschland in den letzten Jahren zu einem verstärkten Interesse an den diagnostischen Kompetenzen von Lehrkräften geführt. Zu diesen Befunden zählen zum Beispiel die wiederholt belegten unbefriedigenden Leistungen eines wesentlichen Teils der deutschen Schülerschaft (z. B. Bos et al., 2003) wie auch die geringe Treffsicherheit von Lehrkräften bei der Identifikation von Risikoschülern (Artelt, Stanat, Schneider & Schiefele, 2001). Vor diesem Hintergrund wurden schwache Schülerleistungen unter anderem auf mangelnde diagnostische Kompetenzen der Lehrerinnen und Lehrer zurückgeführt (DPK, 2001).

Entsprechend hat dieses Thema sowohl in der Forschung als auch in der Lehrerbildung deutlich an Interesse gewonnen. So hat die Kultusministerkonferenz (KMK, 2004) das Beurteilen als einen der vier Kompetenzbereiche in die Standards für die Lehrerbildung aufgenommen und in Nordrhein-Westfalen beispielsweise erfährt das Thema Diagnostik im Rahmen der neuen Lehreraus- und -fortbildung eine stärkere Schwerpunktsetzung (s. LABG¹). Dahinter verbirgt sich die Annahme, dass ohne adäquate Diagnosen eine individualisierte Förderung und damit die Optimierung des Lernens kaum möglich ist (Helmke, Hosenfeld & Schrader, 2004; Rogalla & Vogt, 2008). Tatsächlich deuten erste Forschungsergebnisse an, dass diagnostische

¹ Gesetz über die Ausbildung für Lehrämter an öffentlichen Schulen (Lehrerausbildungsgesetz – LABG). Vom 12. Mai 2009 (GV. NRW. S. 308). Zugriff am 15.05.2013 unter: <http://www.schulministerium.nrw.de/BP/Schulrecht/Lehrerausbildung/LABGNeu.pdf>

Kompetenzen eine notwendige Voraussetzung für die Leistungsförderung sind (Anders, Kunter, Brunner, Krauss & Baumert, 2010; Schrader & Helmke, 1987).

Sowohl aus Sicht der Praxis als auch aus Sicht der Forschung kommt der adäquaten Diagnostik im Kontext Schule offensichtlich eine hohe Bedeutung zu. Dies wirft die Fragen auf, wodurch genau adäquate Diagnosen im schulischen Kontext definiert sind und wie man die Adäquatheit von Diagnosen günstig beeinflussen kann. Zur Klärung dieser Fragen beginnt der vorliegende Beitrag mit einigen einführenden Überlegungen zum Begriff der pädagogischen Diagnostik (Abschnitt 2). Anschließend wird in Abschnitt 3 ein heuristisches Prozessmodell vorgestellt, das abbildet, wie pädagogische Diagnostik als eine Komponente des professionellen Lehrerhandelns im Schulalltag umgesetzt werden kann. In Abschnitt 4 werden dann Forschungsbefunde zur Qualität des diagnostischen Lehrerhandelns zusammengefasst. Dabei wird deutlich, dass die Forschung dem diagnostischen Handeln im schulischen Alltag bislang wenig Rechnung getragen hat. Vielmehr stehen bisher die Akkuratheit von Lehrerurteilen (Abschnitt 4.1) und die kognitiven Prozesse der Informationsverarbeitung (Abschnitt 4.2) im Fokus der Forschung. Überlegungen zur Relevanz des diagnostischen Handelns machen jedoch das Potenzial bestimmter Handlungsaspekte zur Vermeidung von systematischen Urteilsfehlern deutlich (Abschnitt 4.3) und führen zur Forderung, diese professionellen Handlungsweisen empirisch genauer zu beleuchten.

2. Was ist Pädagogische Diagnostik?

Fragt man Lehrkräfte, ob sie im Rahmen ihrer Arbeit diagnostisch tätig sind, so erhält man häufig ein erstauntes, teilweise auch entsetztes „Nein“ zur Antwort. Auf Nachfragen wird deutlich, dass diese Verneinung in ihrem alltagssprachlichen Verständnis von Diagnose als dem Feststellen einer Krankheit – meist durch einen Arzt – begründet liegt. Diagnostisches Handeln meint demzufolge die Erfassung bestimmter Merkmale oder Symptome, die auf einen möglichst eindeutigen (Krankheits-)Zustand schließen lassen. Prototypische diagnostische Methoden sind dabei Laboruntersuchungen, Messung von physiologischen Werten wie Herzfrequenz, Blutdruck sowie technisch unterstützte Untersuchungen von Organen zum Beispiel mittels Sonographie oder Magnetresonanztomographie. Dass dieses Bild wenig mit dem zu tun hat, was Lehrkräfte in ihrem Berufsalltag leisten, ist offensichtlich. Bevor also die Frage geklärt werden kann, ob und in welcher Form Lehrerinnen und Lehrer in ihrem Berufsalltag diagnostisch handeln, ist der Begriff der Diagnostik genauer zu bestimmen.

Hält man sich zunächst an den griechischen Ursprung des Wortes Diagnose (*diagnósis*), so ist damit laut Duden schlicht eine *unterscheidende Beurteilung* gemeint. Im Sinne der beiden Wortbestandteile soll die Beurteilung durch (*dia*) (Er)Kenntnis (*gnósis*) gestützt sein. Demnach ist unter Diagnostik ein Erkenntnis bringender Pro-

zess zu verstehen, in dem Informationen systematisch gesammelt und anschließend zu einem Urteil verdichtet werden.

Was ist nun unter *pädagogischer* Diagnostik zu verstehen? Pädagogische Diagnostik lässt sich anhand ihrer Zielsetzung sowie anhand der verwendeten Kriterien beziehungsweise Methoden beschreiben und von anderer Diagnostik unterscheiden.

2.1 Zielsetzungen Pädagogischer Diagnostik

Wenngleich Diagnosen von den darauffolgenden Handlungen formal abzugrenzen sind, erfüllen sie keinen Selbstzweck. Sie haben immer zum Ziel, nachfolgende Handlungen zu begründen (Jäger & Petermann, 1999; Krohne & Hock, 2007). In seiner vielzitierten Definition versteht Klauer (1982) pädagogische Diagnostik daher auch als „das Insgesamt von Erkenntnisbemühungen im Dienste aktueller pädagogischer Entscheidungen“ (S. 5). Aus Diagnostik wird pädagogische Diagnostik, wenn die Diagnose einer *pädagogischen* Zielsetzung dient.

Im Schulkontext steht als ultimates Ziel die Optimierung des Lernens der Schülerinnen und Schüler im Mittelpunkt. Um dieses zu erreichen, werden pädagogische Entscheidungen getroffen, wobei je nach Handlungsfokus zwischen Modifikations- und Zuweisungs-/Selektionsentscheidungen unterschieden wird (Amelang & Schmidt-Atzert, 2006; Krohne & Hock, 2007). Zuordnungs-/Selektionsentscheidungen bestehen grundsätzlich in einer Koppelung von Personen und relativ feststehenden Bedingungen, während bei der Modifikation innerhalb einer bestehenden Person-Bedingungs-Koppelung über Veränderungen auf Personen- oder Bedingungebene entschieden wird.

Von *Selektion* wird gesprochen, wenn aus einer Gruppe von Personen diejenigen ausgewählt werden, die für bestimmte Anforderungen geeignet erscheinen (z. B. Auswahl unter den an einer weiterführenden Schule angemeldeten Schülerinnen und Schülern oder Auswahl von Schülerinnen und Schülern, die an einem Förderprogramm teilnehmen). Diese Art der Entscheidung impliziert in der Regel, dass weniger Plätze als Bewerber verfügbar sind.

Gibt es keine Beschränkung an verfügbaren Plätzen, so sprechen wir von einer *Zuweisungsentscheidung* (manchmal auch als Bedingungsselektion bezeichnet). Bei der Zuweisungsentscheidung stehen für eine Person unterschiedliche Bedingungen zur Auswahl, aus denen die am besten passende ausgewählt werden kann. Klassische Beispiele sind die Wahl einer Schulform und die Entscheidung über die Versetzung/Nichtversetzung.

Eine der grundlegenden Annahmen von Selektions- und Zuweisungsentscheidungen ist, dass aus früherem Verhalten auf zukünftiges Verhalten geschlossen werden kann beziehungsweise die zukünftige Entwicklung auf Basis der bisherigen Entwicklung extrapoliert werden kann. Damit wird vorausgesetzt, dass Verhalten beziehungsweise Entwicklungsverläufe stabil sind. Im Gegensatz dazu basieren Modifikationsentscheidungen auf der Prämisse der Veränderbarkeit. Ihr Ziel ist die Anpassung der Bedingungen an die Person oder der Person an die Bedingungen.

Im Rahmen einer *Bedingungsmodifikation* gilt es, die Lernumgebung an die jeweiligen Potenziale der Schülerinnen und Schüler anzupassen. Solche Anpassungen können sich auf die gesamte Klasse beziehen, zum Beispiel wenn statt der geplanten Fortsetzung im Stoff eine Wiederholung notwendig ist. Sie können sich aber ebenso auf einzelne Schülerinnen und Schüler beziehen, zum Beispiel wenn individuelles Lernmaterial zur Binnendifferenzierung eingesetzt wird.

Die personenbezogene Modifikationsentscheidung (*Verhaltensmodifikation*) wird dadurch begründet, dass es notwendig und möglich ist, den Lernenden an die Lernumgebung anzupassen. Verhaltensmodifikation findet zum Beispiel statt, wenn Eltern ihr Kind zu einer sorgfältigen Hausaufgabenbearbeitung anleiten oder wenn ein sozial gehemmter Schüler angeleitet wird, sich erfolgreich in eine Gruppenarbeit einzubringen.

Alle vorgenannten Entscheidungen zielen letztlich darauf ab, das Lernen des Schülers zu fördern und sind damit als Zielsetzung pädagogischer Diagnostik gleichermaßen sinnvoll. Sieht man aber die Optimierung des Lernens als einzige Zielsetzung pädagogischer Diagnostik an, wird dies der Schulpraxis nicht gerecht. Neben der Optimierung des Lernens haben Lehrerinnen und Lehrer eine Reihe anderer Aufgaben, die die Erfassung und Analyse von Schülerinformationen erfordern.

Schülerinformationen werden beispielsweise auch benötigt, um Eltern im Rahmen von Elterngesprächen über Lernstand und Verhalten ihres Kindes zu informieren. Ebenso dienen Zeugnisse einer Berichtsfunktion und geben Auskunft über die Leistungen des Schülers und sind darüber hinaus die Basis für die Erteilung von Qualifikationen (z. B. Abitur als Hochschulzugangsberechtigung). In keinem dieser Fälle ist mit der Diagnose unmittelbar ein pädagogisches Anschlusshandeln beziehungsweise eine pädagogische Entscheidung verbunden. Die Schülerinformationen werden nicht in einem pädagogischen Sinne genutzt. Nichtsdestotrotz erscheint es im Schulkontext sinnvoll, auch die auf diese Nutzungsarten hinführende Sammlung und Analyse von Informationen als pädagogische Diagnostik zu fassen. Wir verstehen daher diagnostisches Handeln dann als *pädagogische* Diagnostik, wenn es dazu dient, professionelles pädagogisches Handeln zu begründen und zu unterstützen. Im schulischen Kontext umfasst dies neben der Vorbereitung von pädagogischen Entscheidungen auch die Berichtlegung und Informierung von Schülern und Eltern sowie die Vergabe von Qualifikationen.

2.2 Merkmale von Kriterien und Methoden der Pädagogischen Diagnostik

Die Fokussierung auf professionelles pädagogisches Handeln als Zielsetzung impliziert, dass im Rahmen einer pädagogischen Diagnostik andere Informationen, das heißt Kriterien, mit Hilfe anderer Methoden erfasst werden als beispielsweise in der Medizin. Allerdings finden sich auch Überschneidungen: Ähnlich wie eine Ärztin ein Anamnesegespräch führt, wird zum Beispiel auch eine Lehrerin das Gespräch mit einem Schüler oder seinen Eltern suchen um zu ergründen, woher dessen plötzliche Schulschwierigkeiten resultieren. Beide, Ärztin und Lehrerin, können sich auch

für den Leistungsstand, das Sozialverhalten, die Leistungsangst etc. eines Schülers interessieren.

Davon abgesehen finden sich aber auch deutliche Unterschiede. Selbst wenn Arzt und Lehrkraft bisweilen dieselben Inhalte (*Kriterien*) erfassen, wird die Lehrkraft in der Regel mit *Merkmalsausprägungen* im Normalbereich beschäftigt sein (z. B. bei Nervosität vor einem Referat), während der Arzt sich für außergewöhnliche Ausprägungen (z. B. bei Sozialphobie) interessiert. Mit der Unterscheidung nach der Merkmalsausprägung geht die Unterscheidung nach der *Verbindlichkeit und Spezifität* von Kriterien einher. Für Diagnosen, die auf außergewöhnlichen Ausprägungen basieren, existieren feste, sehr spezifische Kriterienlisten, nicht jedoch für Diagnosen, die auf gewöhnlichen Ausprägungen basieren. Beispielsweise regeln Klassifikationswerke sehr genau, welche Kriterien für eine Dyslexie-Diagnose erfüllt sein müssen (vgl. ICD-10-GM). Im Vergleich dazu haben Bildungsstandards zur Beurteilung der Lesekompetenz eines in seiner Entwicklung unbeeinträchtigten Schülers vielleicht eine ähnliche Verbindlichkeit, erreichen aber nicht die Spezifität von Klassifikationswerken der Medizin.

Die Unterschiede in der Verbindlichkeit und Spezifität von Kriterien bilden sich schließlich auch in den *Methoden* ab. Zwar können grundsätzlich dieselben Methoden eingesetzt werden (z. B. Befragung), sie werden sich aber im *Grad ihrer Standardisierung* unterscheiden. Verbindliche und hoch spezifische Kriterien erlauben den Einsatz standardisierter Methoden. Jeder Patient beantwortet weitgehend dieselben Anamnesefragen oder nimmt am gleichen Untersuchungsverfahren teil. Dagegen existieren für Unterrichtsgespräche keine Gesprächsleitfäden, in denen die zu erfragenden Inhalte festgelegt sind, und auch die Gestaltung von Klassenarbeiten variiert zum Beispiel je nach vorausgegangener Unterrichtsreihe, Leistungsniveau der Klasse oder curricularen Vorgaben. Die hohe Kontextabhängigkeit der Diagnostik führt zur Nutzung von vornehmlich informellen Untersuchungsinstrumenten, die für die spezifische Situation adhoc generiert werden. Die Verfahren orientieren sich in der Regel am sozialwissenschaftlichen Instrumentarium der schriftlichen und mündlichen Befragung sowie der Beobachtung.

2.3 Eine Definition pädagogischer Diagnostik

Zusammenfassend ist zu konstatieren, dass für die Definition von Diagnostik als *pädagogischer Diagnostik* deren Zielsetzung entscheidend ist. Die Zielsetzung pädagogischer Diagnostik besteht darin, das professionelle Handeln eines Pädagogen zu begründen und zu unterstützen. Inhalte und Methoden des diagnostischen Handelns sind zwar durch die Aufgaben des Pädagogen weitgehend mitbestimmt, sie sind aber dadurch weder exklusiv noch eindeutig festgelegt. Damit definieren sie nicht das Pädagogische der Diagnostik, dienen aber der genaueren Beschreibung des diagnostischen Handelns. Die bisher angestellten Überlegungen lassen sich in der folgenden Definition pädagogischer Diagnostik zusammenfassen:

Pädagogische Diagnostik meint das systematische Sammeln von Informationen mittels vornehmlich sozialwissenschaftlicher, wenngleich unstandardisierter Instrumente sowie deren Analyse und Integration, um professionelles pädagogisches Handeln vorzubereiten und zu unterstützen. Im schulischen Kontext sind neben pädagogischen Entscheidungen im Sinne von Modifikation oder Zuweisung/Selektion auch die Informierung und Berichtlegung sowie die Vergabe von Qualifikationen als professionelle Handlungsweisen zu fassen.

3. Der diagnostische Prozess

Die obige Definition impliziert, dass Diagnostik einerseits eine wichtige Voraussetzung für die Erfüllung der unmittelbar auf den Schüler bezogenen Aufgaben von Lehrerinnen und Lehrern bedeutet, die von der KMK (2004) durch die Tätigkeiten Unterrichten, Erziehen, Beurteilen und Beraten beschrieben werden. Gleichzeitig ist Diagnostik selbst ein Teil des professionellen Handelns und sollte entsprechend bestimmten Standards genügen. Sie sollte systematisch und zielgerichtet erfolgen sowie auf wissenschaftlichen Erkenntnissen und auf eigenen, reflektierten Erfahrungen basieren (Baumert & Kunter, 2006; Schön, 1983). Diese Forderungen werden durch den diagnostischen Prozess, wie Jäger (1986) ihn darlegt, erfüllt. In seinem Prozessmodell beschreibt er das diagnostische Handeln in enger Analogie zum wissenschaftlichen Forschungsprozess: Ausgehend von einer konkreten Problemlage gilt es als erstes, eine Annahme über die Ursache beziehungsweise die dahinter liegende Störung zu formulieren. Darauf basierend ist zu klären, wie die jeweils relevanten Kriterien messbar/erfassbar gemacht werden können. Es ist die Frage nach den für die Zielgruppe angemessenen Erhebungsinstrumenten zu klären. Nach der Durchführung der Informationssammlung werden die Daten zusammenfassend analysiert. Es ist zu überprüfen, ob mit den Befunden die eingangs aufgestellte Annahme über die Problemursache belegt werden kann. Ist dies nicht der Fall, werden sukzessive weitere Untersuchungen geplant und durchgeführt, bis eine abschließende Beurteilung der Problemlage und ihrer Ursachen möglich ist.

Dieses Modell ist für das diagnostische Handeln eines Mediziners oder eines klinischen Psychologen passend. Dessen Diagnosen zielen in der Regel darauf ab, aufgrund des Vorliegens bestimmter Symptome einen Krankheitsbegriff zuzuordnen. Basierend auf einer ersten Annahme über eine mögliche Krankheit wird entschieden, welche Merkmale untersucht werden müssen. Die Zusammenführung der Einzelinformationen ermöglicht dann entweder direkt eine schlüssige Antwort oder es werden weitere Informationen eingeholt.

Im Schulalltag erscheint diese Vorstellung vom diagnostischen Prozess allerdings weniger zutreffend. Der Ausgangspunkt schulischer Diagnostik ist in der Regel keine spezifische Problemlage. Vielmehr gilt es für die Lehrkraft, über alle von ihr unterrichteten Schülerinnen und Schüler eine breite Basis an Informationen zu sammeln. Diese müssen ihr beispielsweise ermöglichen, alltägliche Entscheidungen über die

Unterrichtsgestaltung zu treffen, in einem Beratungsgespräch die Eltern über Leistung und Verhalten ihres Kindes in Kenntnis zu setzen und ggf. Empfehlungen zu geben, wie dies verbessert werden kann. Auch die Vergabe von Zeugnisnoten ist eine Kernaufgabe von Lehrerinnen und Lehrern, die mit der oben beschriebenen Idee einer Sammlung von Informationen zur Lösung eines spezifischen Problems nicht kompatibel erscheint.

Die Art der Diagnostik, die eine Lehrkraft in der Schule anwendet, ist besser mit der Durchführung von Vorsorgeuntersuchungen im medizinischen Bereich zu vergleichen. Bei Vorsorgeuntersuchungen werden Informationen ohne eine spezifische Vorannahme oder Problemlage gesammelt. Die Informationen dienen der Einschätzung des Gesamtbefindens des Patienten und ggf. der Ableitung von Empfehlungen für verbessertes Gesundheitsverhalten, ohne dass eine akute Krankheit vorliegt. Auf Basis der erhobenen Informationen können aber auch Indizien für Krankheiten aufgedeckt werden, die dann weitere Untersuchungen nach sich ziehen. Der diagnostische Prozess beginnt in diesem Fall nicht durch die Entscheidung, welche Merkmale zu erfassen sind, da diese in der Regel durch entsprechende Checklisten bereits vorgegeben sind. Erst bei Abweichungen von vorliegenden Normwerten tritt der problembezogene Diagnosetyp auf, bei dem die Entscheidung über weitere Untersuchungen auf Basis der konkreten Problemlage individuell erfolgen muss.

Die Analogie zum Schulkontext ist offensichtlich. In der Regel erfasst eine Lehrkraft für alle Schüler systematisch eine gewisse Anzahl an Basisinformationen. Erst wenn sie spezielle Probleme erkennt, wird sie über spezifische diagnostische Maßnahmen jenseits einer alltäglichen Basisdiagnostik nachdenken. Im Folgenden wird ein heuristisches Prozessmodell entwickelt (vgl. Abbildung 1), das den (wünschenswerten) Ablauf einer schulischen Basisdiagnostik abbildet.

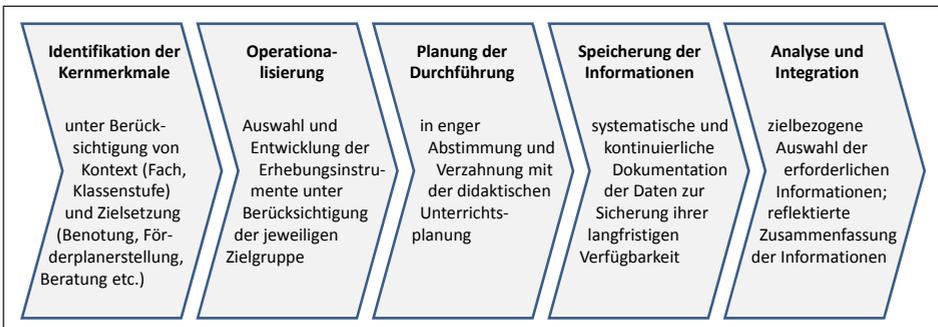


Abbildung 1: Heuristisches Prozessmodell

Schritt 1 (Identifikation der Kernmerkmale). Im ersten Schritt gilt es, die erforderlichen beziehungsweise sinnvollen Basisinformationen zu identifizieren, denn im Vergleich zur Medizin liegen kaum verbindliche Standards vor. Mit Ausnahme der Anzahl und Dauer der Klassenarbeiten gibt es hinsichtlich der Leistungsbeurteilung keine eindeutigen Vorgaben, welche Informationen in welcher Art und welcher Häufigkeit zu erfassen sind. Bevor eine unterrichtsbezogene Basisdiagnostik stattfinden

kann, muss eine Lehrkraft sich also Klarheit über die relevanten Inhaltsbereiche verschaffen. In Abhängigkeit von den Kontextbedingungen (z. B. Alter der Schülerinnen und Schüler, Unterrichtsfach, Schulform etc.) sollten Kernmerkmale identifiziert werden, zu denen systematisch Informationen im Unterrichtsalltag erfasst werden. Dabei ist zu bedenken, dass diese Informationen nicht nur der Leistungsbeurteilung, sondern auch der Information und Beratung von Schülerinnen und Schülern und Eltern sowie der Unterrichtsgestaltung dienen sollen.

Schritt 2 (Operationalisierung). Wenn Klarheit über die zu erfassenden Informationen herrscht, ist zu überlegen, auf welche Weise diese Informationen erhoben werden können. Die Auswahl beziehungsweise Entwicklung geeigneter Instrumente (Klassenarbeiten, Tests, Beobachtungsbögen etc.) stellt hier eine wichtige Aufgabe dar, da schulische Diagnostik in der Regel mit engem Kontextbezug erfolgt und der Einsatz standardisierter Verfahren nur in Ausnahmefällen möglich ist. Eine Vielzahl praxisorientierter Ratgeber leistet Hilfe bei der Entwicklung oder Anpassung von Instrumenten an die individuellen Bedarfe. Grundsätzliches Wissen über Qualitätsmerkmale von Instrumenten und Methoden ist jedoch erforderlich, um eine angemessene Auswahl treffen zu können.

Schritt 3 (Planung). Diagnostikplanung ist ein kaum thematisierter Aspekt in der bisherigen Literatur. Auch in der Praxisliteratur finden sich nach unserem Kenntnisstand keine Informationen darüber, wie die Diagnostik sinnvoll in den Unterrichtsalltag integriert werden kann. Da die Sammlung von Informationen über Schülerinnen und Schüler jedoch über einen längeren Zeitraum hinweg erfolgt und eine Vielzahl an Informationen zu sammeln ist, erscheint eine systematische Planung, wann welche Informationen erfasst werden sollen, dringend erforderlich, um einen effizienten Diagnostikprozess gewährleisten zu können. Eine solche Planung ermöglicht es abzusichern, dass die jeweils erforderlichen/gewünschten Informationen zum richtigen Zeitpunkt vorliegen. Der immer wieder berichtete „Diagnosestress“ von Lehrerinnen und Lehrern, die kurz vor der Vergabe der Zeugnisnoten auf Informationsjagd sind, sprich Heftkontrollen durchführen, Referate und Hausarbeiten verteilen, Schülerinnen und Schüler systematisch im Unterrichtsgespräch prüfen etc., könnte vermieden werden, wenn entsprechende Informationen regelmäßig über das Schuljahr verteilt erfasst werden.

Schritt 4 (Durchführung und Dokumentation). Die konkrete Durchführung der Informationserfassung ist ein zentraler Bestandteil der Diagnostik. Je nach Methode werden hier an die Lehrkräfte unterschiedliche Anforderungen gestellt. Ziel ist es, objektive und valide Daten zu erhalten. Die Lehrkraft muss also dafür sorgen, dass sie nicht durch ihr eigenes Verhalten das Ergebnis verfälscht.

Die Frage, wie die erfassten Daten anschließend dokumentiert und gespeichert werden, ist ebenfalls ein wichtiger Aspekt der Diagnostik. Ziel sollte es sein, die Daten langfristig verfügbar zu halten. Lehrerkalender und Computerprogramme bieten hier Unterstützungsmöglichkeiten. Konkrete Empfehlungen oder Standards über die Informationsspeicherung liegen jedoch nicht vor.

Schritt 5 (Analyse und Integration). In der Regel sind die Aufgaben, denen die pädagogische Diagnostik im Schulkontext dienen soll, hoch komplex. Dies impliziert, dass auch umfangreiche Informationen berücksichtigt werden müssen. Ein diagnostischer Prozess schließt entsprechend mit der Analyse und Integration der verfügbaren Informationen ab. Dabei sind sowohl Entwicklungen über die Zeit zu berücksichtigen als auch die Kombination unterschiedlicher Merkmale zu einem Zeitpunkt. Welche Informationen im Einzelnen berücksichtigt und wie sie relativ zueinander gewichtet werden, ist bislang in der schulischen Diagnostik wenig einheitlich und verbindlich geregelt. Dabei ist jedoch auch fraglich, ob eine starre Vorgabe von Regeln überhaupt möglich und wünschenswert wäre. So zeigt die Expertise-Forschung, dass die Informationsintegration von Experten nicht nach starren Regeln erfolgt, sondern hoch situationsspezifisch und intuitiv gelingt.

4. Qualität Pädagogischer Diagnostik in der Schule

Das vorgestellte Prozessmodell macht deutlich, dass pädagogische Diagnostik eine komplexe Aufgabe ist, die mit hohen Anforderungen an die handelnden Personen einhergeht. Es stellt sich die Frage, wie gut diese Aufgabe im Schulalltag von Lehrerinnen und Lehrern erfüllt wird. Das Modell erlaubt es, die Antworten auf diese Frage zu systematisieren. Nachfolgend werden entsprechende Forschungsbefunde in das Modell eingeordnet. Gleichzeitig wird aber auch aufgezeigt, welche Leerstellen die bisherige Forschung aufweist.

4.1 Akkuratheit des Lehrerurteils

Die Frage nach der Qualität des diagnostischen Handelns verlangt zunächst nach einem Bewertungsmaßstab. In der psychologischen Forschung wird als Maßstab der Diagnosequalität häufig die empirisch gut abbildbare Urteilsakkuratheit herangezogen. Damit ist typischerweise die zutreffende Einschätzung von Personenmerkmalen gemeint (Schrader, 2006). Zahlreiche Forschungsarbeiten widmen sich der Frage nach der Akkuratheit von Lehrerurteilen. Die Forschung auf diesem Gebiet hat zwei zentrale Ergebnisse zu verzeichnen.

1. Durchschnittlich gesehen gelingt es Lehrern sehr gut, die Schüler ihrer Klasse hinsichtlich ihrer Fachleistung in eine Rangreihe zu bringen (z. B. Begeny, Eckert, Montarello & Storie, 2008; Demaray & Elliott, 1998; Hosenfeld, Helmke & Schrader, 2002; Lorenz & Artelt, 2009). Dabei besteht eine leichte Tendenz zur Überschätzung der Fähigkeiten (Rjosk, McElvany, Anders & Becker, 2011; Schrader & Helmke, 1987). Die Beurteilung leistungsferner Merkmale ist bisher seltener untersucht. Hier deutet sich eine geringere Akkuratheit der Urteile an (Karing, 2009; Spinath, 2005).

2. Es zeichnet sich eine hohe interindividuelle Variabilität in der Akkuratheit der Leistungsurteile ab (z. B. Anders et al., 2010; Lorenz & Artelt, 2009). Die Suche nach Lehrer-, Klassen- oder Unterrichtsmerkmalen, die diese Unterschiede erklären können (z. B. Karing, Matthäi & Artelt, 2011; McElvany et al., 2009), oder nach Möglichkeiten zur Förderung der Urteilsakkuratheit (z. B. Bos & Hovenga, 2010; Helmke et al., 2004) ist aber erst in den Anfängen und hat daher bislang noch keine wirklich überzeugenden Antworten geliefert.

Mit Blick auf eine Unterstützung der pädagogischen Praxis ist diese Forschung wenig hilfreich, da zu selten untersucht wird, unter welchen Bedingungen eine gute Diagnose zustande kommt oder wie sie gefördert werden kann. Zudem ist die Erfassung der Urteilsakkuratheit wenig praxisnah, da meist nur punktuelle Lehrerurteile erhoben werden, die ihrerseits mit teils wenig messgenauen Vergleichswerten abgeglichen werden. Diagnosen im Schulalltag sind aber insbesondere durch das Sammeln vielfältiger Informationen über einen längeren Zeitraum gekennzeichnet, die erst zu einem späteren Zeitpunkt zu einem komplexen Urteil verdichtet werden. Doch selbst wenn die Forschung solche komplexen Urteile in den Blick nimmt, bleibt die Aussagekraft eingeschränkt. So ist es bei komplexen Urteilen kaum möglich, den wahren Referenzwert zu finden, gegen den das Lehrerurteil abgeglichen werden kann. Beispielsweise kann bei der Schullaufbahneempfehlung die „richtige“ Empfehlung allenfalls im Rückblick am Erfolg oder Misserfolg auf der empfohlenen weiterführenden Schule erkannt werden (vgl. ausführlich und kritisch dazu van Ophuysen, 2006b).

4.2 Kognitive Prozesse bei der Urteilsbildung

Während die Akkuratheitsforschung eine direkte Bewertung des Urteils anstrebt, kann die Urteilsqualität alternativ auch indirekt erfasst werden (vgl. van Ophuysen, 2010). Dazu wird anstelle der Ergebnisqualität die Qualität der vorauslaufenden kognitiven Prozesse betrachtet. Einschlägige Befunde stammen beispielsweise aus Forschungsarbeiten im Rahmen des Novizen-Experten-Paradigmas. Dort werden Prozesse der *Informationsverarbeitung* bei der Bildung komplexer Urteile zwischen Lehrerinnen und Lehrern mit hoher vs. geringer Berufserfahrung beziehungsweise hohem vs. geringem domänenspezifischen Wissen/Können in quasiexperimentellen Studien verglichen. Als zentraler Befund kann gelten, dass sich die Informationsverarbeitung von Lehrerexperten durch eine hohe Flexibilität auszeichnet. Lehrerexperten können zum Beispiel in Abhängigkeit von der Bedeutsamkeit des Urteils zwischen einer ressourcenintensiven, genauen und einer sparsameren, ungenaueren Entscheidungsstrategie wechseln, während Laien nur die ressourcenintensive Strategie zur Verfügung steht (Krolak-Schwerdt & Rummer, 2005; Krolak-Schwerdt, Böhmer & Gräsel, 2009). Außerdem differenzieren Experten stärker bei der Gewichtung unterschiedlicher Inhalte als Laien (Carter, Cushing, Sabers, Stein & Berliner, 1988). Des Weiteren sind sie besser in der Lage, ein bereits getroffenes Urteil zu revidieren,

wenn sie neue, mit dem Urteil konfligierende Informationen erhalten (van Ophuysen, 2006a). Wenngleich experimentelle Studien mit Lehrkräften durchaus komplexe Stimuli verwenden (z. B. Krolak-Schwerdt, Böhmer & Gräsel, 2009), bleibt die Übertragbarkeit der Ergebnisse auf die Praxis eingeschränkt, da der experimentelle Kontext die reale Situation immer nur stark vereinfacht abbilden kann.

Daneben liefert auch die sozialpsychologische Forschung zur Personenwahrnehmung und zur sozialen Urteilsbildung relevante Erkenntnisse. Sie benennt eine Reihe *typischer Fehler* bei der Sammlung, Speicherung und Verarbeitung von Informationen, die zu systematischen Verzerrungen bei der Urteilsbildung führen. Abbildung 2 stellt die Durchführungsphase der Diagnostik als schematischen Ablauf dar und kennzeichnet, welche typischen Mechanismen bei den einzelnen Schritten zu diesen Urteilsfehlern führen können.

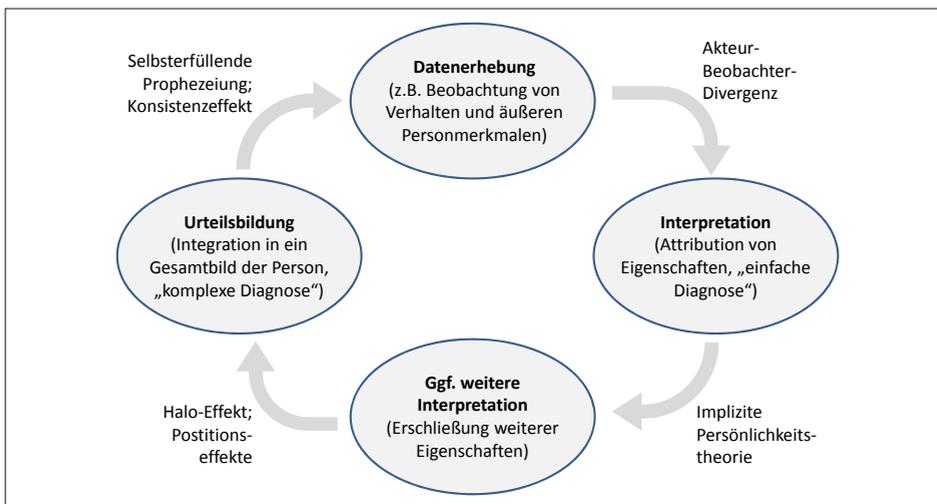


Abbildung 2: Durchführung der Diagnostik und Urteilsfehler

Die *Akteur-Beobachter-Divergenz* kann bei der Interpretation von Beobachtungsdaten auftreten. Sie bezeichnet die Tendenz, eigenes Verhalten situativ zu erklären, also als variabel anzusehen, während das Verhalten anderer Personen eher auf Persönlichkeitseigenschaften zurückgeführt wird und als relativ stabil betrachtet wird (Jones & Nisbett, 1972). Eine fälschliche Attribution des Verhaltens auf stabile Personenmerkmale kann erschwert werden, wenn in der diagnostischen Situation bewusst die Perspektive des Beobachteten eingenommen wird (Frank & Gilovich, 1989) oder wenn Erklärungen mit anderen diskutiert werden (Wells, Petty, Harkins, Kagehiro & Harvey, 1977). So ist zum Beispiel als Erklärung für die morgendliche Unpünktlichkeit eines Schülers nicht mehr nur mangelndes Zeitmanagement denkbar, sondern auch eine Versorgungspflicht für jüngere Geschwister. Zudem sollte die Diagnostik so geplant sein, dass ein entsprechendes Verhalten in verschiedenen Kontexten beobachtet wird. Erst wenn sich die Unpünktlichkeit in mehreren Situationen und wie-

derholt zeigt, zum Beispiel bei der Abgabe von Projektarbeiten, bei der Rückkehr aus der Pause etc., kann zuverlässig auf eine Eigenschaft der Person geschlossen werden.

Ein weiterer Urteilsfehler kann auftreten, wenn aus den so ermittelten Eigenschaften vorschnell auf weitere Eigenschaften geschlossen wird. Grundlage für solche unzulässigen Schlussfolgerungen sind *implizite Persönlichkeitstheorien*. Sie beinhalten stabile Annahmen über den Zusammenhang von Persönlichkeitsmerkmalen (Hofer, 1981; Kleber, 1976). So könnte von mangelndem Zeitmanagement auf Unordentlichkeit oder Faulheit geschlossen werden, obwohl zu letzterem gar keine Daten vorliegen. Ein erster Schritt zur Vermeidung solcher Schlussfolgerungen besteht in der konsequenten Dokumentation aller erhobenen Daten, sodass stets nachvollziehbar ist, zu welchen Personenmerkmalen keine Informationen vorliegen. Ein weiterer Schritt ist die kritische Reflexion der eigenen impliziten Persönlichkeitstheorien. Dies ist allerdings schwierig, denn solche Theorien sind aufgrund ihres impliziten Charakters der Selbstreflexion nur schwer zugänglich. Eine höhere Chance auf Aufdeckung dieser Theorien besteht, wenn im kollegialen Austausch Interpretationen und Bewertungen verglichen werden.

Nach der Interpretation der Daten erfolgt die Urteilsbildung, bei der eine Integration aller Informationen zu einem Gesamtbild von der zu beurteilenden Person stattfindet. Diese Integration kann fehlerbehaftet sein, wenn sich Positionseffekte oder ein Halo-Effekt einstellen.

Positions- oder Reihenfolgeeffekte liegen vor, wenn in eine Bewertung der erste (Primäreffekt) und/oder der letzte Eindruck (Rezenzeffekt) mit einer höheren Gewichtung eingeht. Diese Effekte können sich über ein ganzes Schuljahr erstrecken – etwa, wenn die mündliche Mitarbeit zu Beginn und/oder zum Schluss des Schuljahrs ein besonders hohes Gewicht bei der Urteilsbildung erhält. Ebenso treten sie auch bei kürzeren Zeiträumen auf, zum Beispiel wenn man die Zusammenarbeit einer Gruppe im Verlauf einer Schulstunde beobachtet. Sofern solche Effekte durch einfache Erinnerungsdefizite begründet sind, kann schon eine kontinuierliche Informationssammlung und -dokumentation über den gesamten relevanten Zeitraum ein wirksames Hilfsmittel sein. Darüber hinaus kann man ihnen durch die Vermeidung von Zeitdruck und durch eine hohe Motivation zur Akkuratheit entgegenwirken (Richter & Kruglanski, 1998).

Die Dokumentation der gesammelten Informationen, aber auch der darauf basierenden Interpretationen, ist ebenso hilfreich, um den *Halo-Effekt* zu reduzieren. Ein Halo-Effekt besteht, wenn die starke Ausprägung eines zentralen Merkmals den Gesamteindruck über eine Person bestimmt. In diesem Kontext verdeutlicht die Dokumentation, dass auch andere Informationen über die Person vorliegen, deren Gewichtung es in Bezug auf eine bestimmte Diagnose zumindest bewusst festzulegen gilt. Weitere Gegenmaßnahmen sind die Vermeidung von Zeitdruck (Klauer & Schmeling, 1990) und eine Erhöhung der Anzahl an Datenerhebungen pro Schüler (Cooper, 1981). Wenn mehrere Personen hinsichtlich mehrerer Kriterien beurteilt werden, sollte zudem nicht personen- sondern kriterienweise vorgegangen werden (Johnson & Vidulich, 1956). Bei einem Aufsatz könnten alle Schülerinnen und Schü-

ler zum Beispiel zuerst hinsichtlich ihres sprachlichen Ausdrucks bewertet werden, bevor als weiteres Kriterium die Argumentationslinie bewertet wird.

Nach erfolgter Urteilsbildung ist schließlich zu beachten, dass das gefällte Urteil eine weitere Datenerhebung beeinflussen kann. Als Mechanismen sind hier die selbst erfüllende Prophezeiung und der Konsistenzeffekt denkbar. Während Letzterer eine Beeinflussung der diagnostizierenden Person annimmt, beschreibt Ersterer auch eine Beeinflussung der diagnostizierten Person.

Eine *selbst erfüllende Prophezeiung* tritt ein, wenn Erwartungen der diagnostizierenden Person das Verhalten der diagnostizierten Person in Richtung der Erwartung beeinflussen. Beispielsweise könnten Lehrkräfte an Schülerinnen und Schüler mit Migrationshintergrund geringere Leistungserwartungen als an Schülerinnen und Schüler ohne Migrationshintergrund haben. Dies wiederum könnte dazu führen, dass sie für Erstere eine Lernumgebung mit weniger Anregungen anbieten, was wiederum zu einer schlechteren Leistungsentwicklung als bei den Schülerinnen und Schülern ohne Migrationshintergrund führen könnte. Zur Effektivität von Maßnahmen gegen die negativen Wirkungen von Erwartungseffekten gibt es kaum Forschungsergebnisse. Die Aufdeckung und Vermeidung solcher negativen Erwartungseffekte ist in der Selbstreflexion meist nur schwer möglich. Als erfolgreich hat sich allerdings das Feedback von Kollegen und neutralen Beobachtern erwiesen (vgl. zusammenfassend Alexander & Schofield, 2006).

Im Gegensatz dazu ist eine kritische Selbstreflexion bei der Eingrenzung des *Konsistenzeffekts* allerdings sehr hilfreich. Der Konsistenzeffekt beschreibt die Tendenz der menschlichen Informationsverarbeitung, ein möglichst konsistentes Bild von anderen Personen zu konstruieren. In der Folge finden Informationen, die gut in das bisherige Gesamtbild passen, bei der Suche und Verarbeitung neuer Informationen mehr Berücksichtigung. Eine verzerrte Urteilsbildung aufgrund des Konsistenzeffektes lässt sich verringern, indem Vorabentscheidungen vermieden werden und stattdessen die vorliegenden Informationen fokussiert werden (Jonas, Traut-Mattausch, Frey & Greenberg, 2008). Dabei ist es von Vorteil, wenn die Informationen nicht nach und nach, sondern auf einen Blick verfügbar sind (Jonas, Schulz-Hardt, Frey & Thelen, 2001). Beides wird durch eine gezielte Planung und Dokumentation unterstützt.

4.3 Diagnostisches Handeln im Schulalltag

Die Ausführungen zu den Urteilsfehlern machen deutlich, dass mit (1) der Entwicklung geeigneter Erhebungsinstrumente, (2) der Schaffung vielfältiger diagnostischer Anlässe im Unterrichtsalltag und (3) der Dokumentation/Speicherung der Informationen mindestens drei Aspekte des diagnostischen Handelns als zentrale Vorläuferbedingungen für eine unverzerrte Informationsverarbeitung und damit für ein qualitativ hochwertiges diagnostisches Urteil gelten können. Darüber hinaus ist schließlich auch (4) die Auswahl der relevanten Merkmale, die zu einem Gesamturteil verdichtet werden, entscheidend für dessen Qualität. Welche Befunde liefert die

empirische Forschung bislang zu diesen Facetten des diagnostischen Prozesses bei Lehrerinnen und Lehrern?

Aspekt 1: Entwicklung/Auswahl geeigneter Erhebungsinstrumente. Aus Lehrerbefragungen geht hervor, dass in der Schule ein breites Repertoire an Erhebungsmethoden eingesetzt wird (z. B. Projektarbeiten, Vorträge, Klassenarbeiten etc., Martínez, Stecher & Borko, 2009; McMillan, Myran & Workman, 2002), wobei bisher keine Erkenntnisse über die Qualität der eingesetzten Instrumente vorliegen. Unklar ist beispielsweise, ob es Lehrkräften gelingt, die Schwierigkeit einer Aufgabenstellung an das Niveau einer Klasse oder eines Schülers anzupassen. Ebenso wenig ist bekannt, ob und in welchem Umfang Lehrkräfte die umfangreiche Praxisliteratur (z. B. Bohl, 2009; Paradies, Wester & Greving, 2005) zur Leistungsmessung und Leistungsbewertung nutzen.

Die psychologische Diagnostik fordert grundsätzlich die Verwendung vielfältiger Methoden zur Erhöhung der *Validität* eines (komplexen) Urteils. Amelang und Schmidt-Atzert (2006) verweisen diesbezüglich auf den Leitsatz, dass jedes Merkmal/jede Eigenschaft mit mindestens zwei verschiedenen Methoden erhoben werden sollte. Dies ist auch im Schulkontext eine plausible Forderung. So kann aus Klassenarbeiten, die mit Zeit- und Leistungsdruck verbunden sind, nur ein unvollständiges Bild über die Leistungsfähigkeit eines Schülers gewonnen werden. Dieses sollte also durch Informationen über die Qualität der Bearbeitung von Arbeitsblättern, von Projektarbeiten oder der kontinuierlichen mündlichen Mitarbeit erweitert werden.

Darüber hinaus ist auch die Forderung nach *Objektivität* für den schulischen Kontext unbestritten. Dazu sollte die Auswertung möglichst nach vorab festgelegten und eindeutig beschriebenen Kriterien erfolgen, denn Kriterienkataloge können Urteilern zu einer einheitlicheren Beurteilung verhelfen (vgl. zusammenfassend dazu Ingenkamp & Lissmann, 2008, Kapitel 3). Während Birkel und Birkel (2002) noch davon ausgingen, dass die Anwendung von Kriterienkatalogen kaum in die Lehrerbildung und noch weniger in die Praxis vorgedrungen ist, ist mit der Einführung der landesweit verbindlichen Bildungsstandards zum Schuljahresbeginn 2004/05 beziehungsweise 2005/06 dementsprechend ein Wandel zu erwarten. Aktuelle empirische Befunde zur Validität und Objektivität von Erhebungsinstrumenten in der alltäglichen schulischen Diagnostik liegen jedoch kaum vor.

Aspekt 2: Schaffung und Gestaltung diagnostischer Anlässe im Schulalltag. In der englischsprachigen Literatur wird seit einiger Zeit kritisiert, dass eine Messtheorie, die im Kontext großer Schulleistungsstudien beziehungsweise standardisierter, testpsychologischer Diagnostik entwickelt wurde, nun auf die Diagnostik in der Schule übertragen wird. Ausgehend von der Annahme, dass beide Formen der Diagnostik nur bedingt vergleichbar sind, wird eine eigene Messtheorie gefordert, die Qualitätsindikatoren guter Diagnosen speziell im Kontext des Unterrichts benennt (Brookhart, 2003). So ist zu hinterfragen, ob hoch standardisierte Untersuchungsbedingungen, die im Kontext von Studien wie *Programme for International Student Assessment* (PISA) oder *Trends in International Mathematics and Science Study* (TIMSS) optimal

sind, im Schulkontext nicht sinnvoller durch kontextspezifische, repräsentative Erhebungssituationen ersetzt werden sollten (Moss, 2003). In diesem Zuge wird auf die Notwendigkeit einer Integration der Diagnostik in den Unterricht hingewiesen (vgl. McMillan, 2003). Demnach sollten Unterrichten und Diagnostizieren nicht als getrennte Tätigkeiten aufgefasst werden. Vielmehr sollte der Unterricht gleichzeitig sowohl dem Lehren beziehungsweise Lernen dienen als auch Diagnosemöglichkeiten bieten. Beispielsweise könnten kooperative Lernformen nicht nur zur Vermittlung von Lernzielen – zum Beispiel zur Verbesserung sozialer Kompetenzen und zur Erarbeitung neuen Wissens – eingesetzt werden. Vielmehr entstehen durch die – im Gegensatz zum lehrerzentrierten Frontalunterricht – geringere unmittelbare Beteiligung der Lehrkraft Freiräume, in denen sie systematisch das Arbeits- und Sozialverhalten der Schülerinnen und Schüler beobachten und beurteilen kann. Die gezielte Koppelung von Unterricht und Diagnose erhöht insbesondere die Anzahl an Gelegenheiten, zu denen diagnostische Informationen erfasst werden können. Somit wird neben der Forderung nach Repräsentativität auch der Forderung nach zuverlässigen, messgenauen Urteilen Rechnung getragen, die auf möglichst vielen (verschiedenen) Messungen beruhen (Smith, 2003).

Wie Lehrkräfte diagnostisches Handeln planen, ob sie der Forderung nach einer Verzahnung mit der didaktischen Unterrichtsplanung nachkommen, wie viele und welche Anlässe zur Erfassung von Schülerinformationen genutzt werden – über all dies liegen bislang jedoch keinerlei empirische Ergebnisse vor.

Aspekt 3: Dokumentation und Speicherung diagnostischer Informationen. Während die beiden bereits genannten Aspekte auf die Datengewinnung (Schritt 1 bis 3 des Prozessmodells, vgl. Abschnitt 3) fokussieren, wird nun der Blick auf die weitere Datenverarbeitung gerichtet. In vielen Situationen liegt zwischen der Erfassung der Informationen und ihrer zusammenfassenden Auswertung ein längerer Zeitraum. So müssen insbesondere bei der Festlegung von Zeugnisnoten vielfältige Informationen integriert werden, die über das komplette Schulhalbjahr hinweg gesammelt wurden. Auch in Beratungssituationen wird in der Regel nicht nur die aktuelle Situation, sondern die Entwicklung über einen längeren Zeitraum berücksichtigt.

Die sozialpsychologische Forschung macht deutlich, dass komplexe Urteile, die aus der Erinnerung gefällt werden, Verzerrungen unterliegen können (vgl. Abschnitt 4.2). Um solchen Erinnerungseffekten entgegenzuwirken, ist es erforderlich, die Informationen systematisch und kontinuierlich zu speichern. Eine detaillierte Datendokumentation gewährleistet, dass Informationen unverfälscht und vollständig zum Zeitpunkt der Urteilsbildung verfügbar sind. Die Bedeutung der Dokumentation wird insbesondere deutlich, wenn man bedenkt, dass Fachlehrer in der Sekundarstufe oft mehr als 200 Schülerinnen und Schüler in einem Halbjahr unterrichten. Ohne eine systematische Aufzeichnung ist es kaum möglich, die Zuordnung von Informationen zu den jeweiligen Schülerinnen und Schülern korrekt zu leisten. Trotz der hohen Relevanz dieser Facette des diagnostischen Lehrerhandelns gibt es keinerlei empirische Erkenntnisse, wie Lehrkräfte diese Aufgabe bewältigen. Auch in der Lehrerausbildung wird dieses Thema nicht behandelt. Wenngleich Lehrerkalender

verschiedene (Listen-)Formate zur Informationsspeicherung anbieten, weisen erste eigene Lehrerbefragungen darauf hin, dass die Dokumentation hoch individualisiert erfolgt. Dabei gewinnt die Nutzung von Computern an Bedeutung, zum Teil unter Einsatz spezieller Software.

Die empirische Erforschung der Datenspeicherung im Schul- beziehungsweise Unterrichtskontext und die darauf gründende Identifizierung von *best-practice*-Beispielen erscheint damit ein zentraler, aber bislang nicht beachteter Weg zur Verbesserung des diagnostischen Lehrerhandelns.

Aspekt 4: Auswahl relevanter Informationen. Je nach Diagnoseziel sind unterschiedliche Informationen aus den verfügbaren Daten über den Schüler relevant. Das Sozialverhalten kann bei einem Beratungsgespräch mit den Eltern eines Schülers eine relevante Information sein, nicht jedoch bei der Vergabe einer Schulnote.

Welche Merkmale Lehrkräfte im Rahmen ihrer Diagnostik nutzen, wird indirekt durch die groß angelegten, *standardisierten Schulleistungstudien* (z. B. PISA, TIMSS, IGLU) analysiert. Dort wird anhand großer Stichproben statistisch untersucht, welche Informationen die Schulnote oder Schulformempfehlung vorhersagen können. Vorhersagekraft hat dabei nicht nur die Fachleistung (i. d. R. operationalisiert über einen standardisierten Schulleistungstest) bewiesen. Darüber hinaus finden sich Zusammenhänge zum Beispiel zwischen der Motivation der Schülerinnen und Schüler und den Schulnoten (z. B. Frenzel, Pekrun & Zimmer, 2006) oder dem sozioökonomischen Hintergrund und der Schulformempfehlung (z. B. Arnold, Bos, Richert & Stubbe, 2007).

Da solche Studien Lehrkräfte nicht direkt zur Urteilsbildung befragen, sondern das schlussendlich getroffene Urteil mit Schülerdaten (z. B. Testergebnis, Fragebogenangaben zur Motivation) verbinden, können sie keine Aussage über die Einflussrichtung machen. So bleibt unklar, ob ein besonders positives Selbstkonzept eines Schülers Ursache oder aber Folge einer guten Note ist. Problematisch an diesen Analysen ist weiterhin, dass die Vorhersagekraft eines Merkmals nur dann erfasst werden kann, wenn es erfragt und im statistischen Modell berücksichtigt wird. Bei der Vielzahl möglicher Einflussgrößen oder Wechselwirkungen kann letztlich immer nur ein Ausschnitt aus einem Universum von Prädiktoren betrachtet werden. Ein weiterer Nachteil besteht darin, dass statistische Zusammenhänge kein Nachweis für bewusste Urteilsprozesse sind. Zeigt sich zum Beispiel auch nach Kontrolle von Unterschieden in der Testleistung ein Zusammenhang zwischen der Note und dem sozialen Hintergrund eines Schülers, ist dies nicht unbedingt Ausdruck einer expliziten Berücksichtigung des sozialen Hintergrunds. So erfassen Noten mehr Leistungsaspekte (z. B. mündliche Mitarbeit, Güte der Hausaufgabenbearbeitung) als ein standardisierter Leistungstest und diese wiederum könnten durch Unterschiede im sozialen Hintergrund des Schülers beeinflusst sein (Maaz et al., 2008).

Dieser Nachteil wird in *direkten Lehrerbefragungen* ausgeräumt. Im Rahmen solcher Befragungen werden Lehrkräfte gebeten, ihre Bewertungskriterien und die Gründe für ihre Urteile offenzulegen. Im deutschsprachigen Raum existieren allerdings kaum Studien, die die Urteilsbildung von Lehrerinnen und Lehrern mit diesem

methodischen Zugang untersuchen. Die wenigen verfügbaren Studien konzentrieren sich auf die Urteilsbildung bei der Übergangsempfehlung. Sie belegen, dass Lehrkräfte bei der Übergangsempfehlung nicht nur „Eignung, Neigung und Wille des Kindes zu geistiger Arbeit“ (KMK, 2010, S. 5) berücksichtigen, sondern zum Beispiel auch das Ausmaß der elterlichen Unterstützung (Nölle, Hörstermann, Krolak-Schwerdt & Gräsel, 2009; Pohlmann, 2009). Dies gilt insbesondere bei Schülerinnen und Schülern, deren Leistungsniveau für die entsprechende Schulform nur als knapp ausreichend bewertet wird (Hollstein, 2007).

In den USA dagegen sind Lehrkräfte seit der Anregung von Stiggins, Frisbie und Griswold (1989) schon vielfach zu ihrer Urteilspraxis (*grading/assessment practice*) befragt worden. Auch aus diesen Studien geht hervor, dass die Lehrerinnen und Lehrer bei der Notenvergabe neben der lernzielbezogenen Leistung auch andere Kriterien berücksichtigen. Dies können zum Beispiel die Anstrengungsbereitschaft des Schülers, aber auch das Leistungsniveau der Klasse sein (z. B. Cizek, Fitzgerald & Rachor, 1995). Allerdings bestehen zwischen den Lehrkräften teils erhebliche Unterschiede in der Gewichtung der Kriterien (McMillan, 2001). Nachteilig an dieser direkten Erfassung der Urteilsprädiktoren ist jedoch, dass sie möglicherweise sozial erwünschte Antworten provozieren – wenngleich die Lehrerinnen und Lehrer in den amerikanischen Studien erstaunlicherweise auch Benotungspraktiken schildern, die nicht mit gängigen Empfehlungen in Einklang stehen (Stiggins et al., 1989).

Obwohl die Forschung zu den Prädiktoren von komplexen Lehrerurteilen zu wichtigen Erkenntnissen hinsichtlich pädagogischer Diagnostik im Schulalltag beigetragen hat, ermöglicht sie per se keine Aussagen über die Qualität des diagnostischen Prozesses. Um eine Bewertung der Informationsauswahl zu leisten, müssten Kenntnisse darüber vorliegen, welche Informationen aus objektiver Sicht relevant für die Urteilsbildung sind und wie unterschiedliche Informationen relativ gewichtet werden sollten. Für die Schulformempfehlung ist zum Beispiel empirisch nicht eindeutig geklärt, welche Eigenschaften zur Vorhersage des Sekundarschulerfolgs eines Schülers relevant sind (Sauer & Gamsjäger, 1996). Auch rein normative Vorgaben sind schwierig: Manche Forscher etwa sehen es als Diskriminierung an, wenn Informationen über das häusliche Umfeld berücksichtigt werden (z. B. mögliche Unterstützung durch die Eltern, Hollstein, 2007). Andere dagegen betrachten das Hintergrundwissen über einzelne Schülerinnen und Schüler als diagnostisch relevant (Ingenkamp, 1995, S. 35).

Im Gegensatz zu den anderen drei Aspekten liegt zur Frage, welche Informationen letztlich integriert werden, vergleichsweise viel Forschung vor. Dennoch gelingt es damit nicht, eindeutige, empirisch begründete Empfehlungen zur Informationsauswahl für verschiedene zentrale Diagnoseziele zu liefern.

5. Fazit

Der Beschreibung des Prozessmodells in Abschnitt 3 folgend setzt eine gute Diagnose voraus, dass Informationen vollständig, systematisch und mittels adäquater Methoden erfasst und gespeichert werden, um gegebenenfalls eine relevante Teilmenge zu einem späteren Zeitpunkt abrufen und integrieren zu können. Über die zentralen Prozesse der Informationserfassung, -speicherung und -integration im Kontext der schulischen Diagnostik liegen jedoch bislang wenig gesicherte empirische Befunde vor. Ziel weiterer Forschungsaktivitäten sollte es daher sein, Wissen über das alltägliche Diagnostikhandeln von Lehrkräften zu generieren und den Zusammenhang mit kognitiven Prozessen der Urteilsbildung sowie mit der Urteilsqualität zu analysieren. Erst dann wird es möglich sein, empirisch gestützte, adäquate und praktikable Empfehlungen zur Optimierung des diagnostischen Handelns von Lehrkräften im Schulalltag zu entwickeln.

Literatur

- Alexander, K. & Schofield, J.W. (2006). Erwartungseffekte: Wie Lehrerverhalten schulische Leistungen beeinflusst. In J.W. Schofield (Hrsg.), *Migrationshintergrund, Minderheitenzugehörigkeit und Bildungserfolg. Forschungsergebnisse der pädagogischen, Entwicklungs- und Sozialpsychologie. AKI-Forschungsbilanz 5* (S. 47–69). Berlin: WZB.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- 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.
- Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T.C. (2007). Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 271–297). Münster: Waxmann.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Begeny, J.C., Eckert, T.L., Montarello, S.A. & Storie, M.S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23 (1), 43–55.
- Birkel, P. & Birkel, C. (2002). Wie einig sind sich die Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49 (3), 219–224.
- Bohl, T. (2009). *Prüfen und Bewerten im Offenen Unterricht* (4. Aufl.). Weinheim und Basel: Beltz.

- Bos, W. & Hovenga, N. (2010). Diagnostische Kompetenzen. Besser individuell fördern. *SchuleNRW*, 62 (8), 383–385.
- Bos, W., Lankes, E.-M., Schwippert, K., Valtin, R., Voss, A., Badel, I. & Plaßmeier, N. (2003). Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 69–142). Münster: Waxmann.
- Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Carter, K., Cushing, K., Sabers, D., Stein, P. & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, 39 (3), 25–31.
- Cizek, G.J., Fitzgerald, S.M. & Rachor, R.E. (1995). Teacher's assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3 (2), 159–179.
- Cooper, W.H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90 (2), 218–244.
- 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.
- DPK [Deutsches PISA-Konsortium] (Hrsg.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Frank, M.G. & Gilovich, T. (1989). Effect of memory perspective on retrospective causal attributions. *Journal of Personality and Social Psychology*, 57 (3), 399–403.
- Frenzel, A.C., Pekrun, R. & Zimmer, K. (2006). Selbstvertrauen, Engagement und Lernverhalten in Mathematik. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 195–208). Münster: Waxmann.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Grieser (Hrsg.), *Schulmanagement und Schulentwicklung* (S. 119–144). Baltmannsweiler: Schneider Verlag Hohengehren.
- Hofer, M. (1981). Schülergruppierungen im Urteil und Verhalten des Lehrers. In M. Hofer (Hrsg.), *Informationsverarbeitung und Entscheidungsverhalten von Lehrern. Beiträge zu einer Handlungstheorie des Unterrichtens* (S. 192–221). München: Urban & Schwarzenberg.
- Hollstein, B. (2007). Sozialkapital und Statuspassagen. Die Rolle von institutionellen Gatekeepern bei der Aktivierung von Netzwerkressourcen. In J. Lüdicke & M. Diewald (Hrsg.), *Soziale Netzwerke und soziale Ungleichheit. Zur Rolle von Sozialkapital in modernen Gesellschaften* (S. 53–83). Wiesbaden: VS.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen. Zeitschrift für Pädagogik*, 45. Beiheft (S. 65–82). Weinheim und Basel: Beltz.
- Ingenkamp, K. (Hrsg.). (1995). *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (9. Aufl.). Weinheim und Basel: Beltz.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* (6. Aufl.). Weinheim und Basel: Beltz.

- Jäger, R.S. (1986). *Der diagnostische Prozeß. Eine Diskussion psychologischer und methodischer Randbedingungen* (2. Aufl.). Göttingen: Hogrefe.
- Jäger, R.S. & Petermann, F. (1999). *Psychologische Diagnostik*. Weinheim: PVU.
- Jonas, E., Schulz-Hardt, S., Frey, D. & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80 (4), 557–571.
- Jonas, E., Traut-Mattausch, E., Frey, D. & Greenberg, J. (2008). The path or the goal? Decision vs. information focus in biased information seeking after preliminary decisions. *Journal of Experimental Social Psychology*, 44 (4), 1180–1186.
- Johnson, D.M. & Vidulich R.N. (1956). Experimental manipulation of the halo effect. *Journal of Applied Psychology*, 40 (2), 130–134.
- Jones, E.E. & Nisbett, R.E. (1972). The actor and the observer: Divergent perceptions of the causes of the behavior. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins & B. Weiner (Hrsg.), *Attribution: Perceiving the causes of behavior* (S. 79–94). Morristown, NJ: General Learning Press.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 197–209.
- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25 (3), 159–172.
- Klauer, K.J. (1982). Perspektiven Pädagogischer Diagnostik. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik* (Bd. 1, S. 3–14). Schwann: Düsseldorf.
- Klauer, K.J. & Schmeling, A. (1990). Sind Halo-Fehler Flüchtigkeitsfehler? *Zeitschrift für experimentelle und angewandte Psychologie*, 37 (4), 594–607.
- Kleber, E.W. (1976). Tendenzen, die das Urteil des Lehrers beeinflussen. In E.W. Kleber, H. Meister, C. Schwarzer & R. Schwarzer (Hrsg.), *Beurteilung und Beurteilungsprobleme. Eine Einführung in Beurteilungs- und Bewertungsfragen in der Schule* (S. 39–62). Weinheim und Basel: Beltz.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der KMK vom 16.12.2004*. Zugriff am 17.05.2013 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (2010). *Regelungen der Länder zum Übergang von der Grundschule in Schulen des Sekundarbereichs I und über die Gestaltung der Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe)*. Zugriff am 17.05.2013 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_10_18-Uebergang-Grundschule-S_eI1-Orientierungsstufe.pdf
- Krohne, H.W. & Hock, M. (2007). *Psychologische Diagnostik. Grundlagen und Anwendungsfelder*. Stuttgart: Kohlhammer.
- 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. & Rummer, R. (2005). Der Einfluss von Expertise auf den Prozess der schulischen Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37 (4), 205–213.
- 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.
- Maaz, K., Neumann, M., Trautwein, U., Wendt, W., Lehmann, R. & Baumert, J. (2008). Der Übergang von der Grundschule zur weiterführenden Schule: Die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30 (3), 519–548.
- Martínez, J.F., Stecher, B. & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, 14 (2), 78–102.
- 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*, 23 (3–4), 223–235.
- McMillan, J.H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20 (1), 20–32.
- McMillan, J.H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22 (4), 34–43.
- McMillan, J.H., Myran, S. & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95 (4), 203–213.
- Moss, P.A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22 (4), 13–25.
- Nölle, I., Hörstermann, T., Krolak-Schwerdt, S. & Gräsel, C. (2009). Relevante diagnostische Informationen bei der Übergangsempfehlung – die Perspektive der Lehrkräfte. *Unterrichtswissenschaft*, 37 (4), 294–310.
- Paradies, L., Wester, F. & Greving, J. (2005). *Leistungsmessung und -bewertung*. Berlin: Cornelsen.
- Pohlmann, S. (2009). *Der Übergang am Ende der Grundschulzeit. Zur Formation der Übergangsempfehlung aus der Sicht der Lehrkräfte*. Münster: Waxmann.
- Richter, L. & Kruglanski, A.W. (1998). Seizing on the latest: Motivationally driven recency effects in impression formation. *Journal of Experimental Social Psychology*, 34 (4), 313–329.
- Rjosk, C., McElvany, N., Anders, Y. & Becker, M. (2011). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung der basalen Lesefähigkeit ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 58 (2), 92–105.
- Rogalla, M. & Vogt, F. (2008). Förderung adaptiver Lehrkompetenz: Eine Interventionsstudie. *Unterrichtswissenschaft*, 36 (1), 17–36.
- Sauer, J. & Gamsjäger, E. (1996). *Ist Schulerfolg vorhersagbar? Die Determinanten der Grundschulleistung und ihr prognostischer Wert für den Sekundarschulerfolg*. Göttingen: Hogrefe.
- Schön, D.A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
- Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. h. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 95–100). Weinheim: PVU.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1 (1), 27–52.

- Smith, J.K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22 (4), 26–33.
- 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.
- Stiggins, R.J., Frisbie, D.A. & Griswold, P.A. (1989). Inside high school: Building a research agenda. *Educational Measurement: Issues and Practice*, 8 (2), 5–14.
- van Ophuysen, S. (2006a). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnenempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38 (4), 154–161.
- van Ophuysen, S. (2006b). Zur Problematik der Schulformempfehlung nach der Grundschulzeit und ihrer prognostischen Qualität. In W. Bos, H.G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 14, S. 49–79). Weinheim: Juventa.
- 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. Daten, Beispiele und Perspektiven* (Bd. 16, S. 203–234). Weinheim: Juventa.
- Wells, G.L., Petty, R.E., Harkins, S.G., Kagehiro, D. & Harvey, J.H. (1977). Anticipated discussion of interpretation eliminates actor-observer differences in the attribution of causality. *Sociometry* 40 (3), 247–253.