pedocs
Open Access Erziehungswissenschaften

DIPF

Reisse, Wilfried

Zweck und Wirkung der Leistungsbeurteilung. Wie man Risiken und Fehlentwicklungen vermeiden kann

Die Deutsche Schule 96 (2004) 2, S. 184-200

Quellenangabe/ Reference:
Reisse, Wilfried: Zweck und Wirkung der Leistungsbeurteilung. Wie man Risiken und Fehlentwicklungen vermeiden kann - In: Die Deutsche Schule 96 (2004) 2, S. 184-200 - URN: urn:nbn:de:0111-pedocs-274114 - DOI: 10.25656/01:27411

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-274114
https://doi.org/10.25656/01:27411

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

pedocs
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der Leibniz-Gemeinschaft

Wilfried Reisse

Zweck und Wirkung der Leistungsbeurteilung

Wie man Risiken und Fehlentwicklungen vermeiden kann

„Wer andere Menschen einer Prüfung oder sonstigen Leistungskontrolle unterzieht, muss begründen, welchen Zweck er damit verfolgt."[1]

Wenn den Schulen für die Leistungsbeurteilung *extern entwickelte Tests* vorgegeben werden, dann könnte dadurch – so wird häufig befürchtet – die Zielsetzung von Schule gesteuert und verengt werden (u.a. Boenicke 2000, Weiland 1999). Zudem könnten Leistungsbeurteilungen einen Umfang einnehmen, der Unterrichten und Erziehen als eigentliche Kernaufgabe beeinträchtigt und an den Rand drängt.

Das Risiko solcher *Fehlentwicklungen* hat sich vergrößert: Das schon vorhandene weite Spektrum an schulischen Leistungsbeurteilungen wird zurzeit ergänzt durch neue Formen von Leistungsmessungen, die als Reaktion auf die Ergebnisse internationaler Vergleichsuntersuchungen von Schulleistungen zu sehen sind. So wurde die Öffentlichkeit informiert über die anstehenden Wiederholungen der verschiedenen internationalen Vergleichsuntersuchungen und über die Absicht der Länder, neue Formen externer Leistungserhebungen einzuführen. Bezeichnungen dafür sind „Diagnosearbeiten, Jahrgangsarbeiten, Orientierungsarbeiten, Jahrgangsstufentests, Vergleichsarbeiten, Lernstandserhebungen, Kompetenztests" (KMK 2003, Teil D, S. 267). Zudem wird noch in einigen Ländern darüber diskutiert, ob man vorhandene Beurteilungsformen durch andere ersetzen sollte, zum Beispiel die dezentrale Form des Abiturs durch ein Zentralabitur. Schließlich gibt es Überlegungen, bisher nicht praktizierte Abschlussprüfungen für Haupt- und Realschulen einzuführen.

1. Leistungsbeurteilung – Herr oder Diener?

Externe Leistungsmessungen werden in vielen Ländern verwendet. Prinzipiell ist es eine der sinnvollen Reaktionen auf die schlechten deutschen PISA-Ergebnisse, sie auch verstärkt in Deutschland einzusetzen. Dennoch sollte das Risiko nicht vernachlässigt werden, dass Fehlentwicklungen anstelle der erhofften positiven Effekte eintreten können. Dabei muss man auch die Akzeptanz solcher Leistungsbeurteilungen durch das Lehrpersonal berücksichtigen. So haben in einer Erhebung von Ditton (2202, S. 29) etwa ein Drittel der befragten Lehrer in Bayern angegeben, dass die Ergebnisse externer Leistungskontrollen – wie die der internationalen Vergleichsuntersuchungen – keine Hilfe, sondern eher ein Problem für die alltägliche Unterrichtspraxis darstellen.

1 Mit dieser Aussage wurden die Teilnehmer eines Workshops aufgefordert, die Zwecke zu beschreiben, die mit einer Revision von Prüfungssystemen erreicht werden sollten. Die Teilnehmer waren leitende Mitarbeiter von Ministerien und Testinstituten afrikanischer Länder.

Dieses Problem einer unerwünschten *Dominanz des Prüfungs- oder Kontrollbereichs* wird in vielen Ländern diskutiert. Beispiele sind Südafrika (Lolwana 2001), Hongkong (Scott 2001) und England, wo dies zu einer ausgesprochenen Krise der Prüfungen führte (Kingdon 2001, Jungclaussen 2003, McCaig 2003). In England sind zwei mögliche negative Effekte (quantitatives Übergewicht der Prüfungen und zu große Steuerungswirkung) eingetreten.

Solche Probleme werden nicht nur durch ein Übermaß an externen Leistungstests bewirkt, als „Testeritis" gescholten. Auch aufwendige *Beurteilungen zur Diagnose und Lernförderung* auf Klassenebene können den Unterricht beeinträchtigen, wie sich u. a. in einer empirischen Untersuchung zu Lern- und Leistungsberichten (Jürgens 2001) zeigte. – Wie kann man solche Fehlentwicklungen vermeiden oder zumindest reduzieren?

2. Planung von Leistungsbeurteilungen

Wenn Fehlentwicklungen bei Leistungsbeurteilungen verhindert werden sollen, muss dies *mit der Planung beginnen*. Zu klären, welche Zwecke durch vorgesehene Leistungsbeurteilung erreicht werden sollen, ist dabei ein sehr wichtiger, allerdings nicht der einzige erforderliche Planungsschritt, der dazu beitragen soll, Fehlentwicklungen zu vermeiden.

Nehmen wir ein *fiktives Beispiel*: In den Schulen eines Landes sollen neben den üblichen Routine-Leistungsbeurteilungen neue externe Leistungsbeurteilungen eingeführt werden, die jährlich weitgehend flächendeckend einzusetzen sind. Geplant werden zusätzliche Untersuchungen bei der Einschulung und Leistungsmessungen in mehreren Jahrgangsstufen in den wichtigsten Fächern. Außerdem sind vorgesehen mehrere wissenschaftliche Untersuchungen zur Entwicklung von Schulleistungen sowie die weitere Beteiligung an internationalen Vergleichsuntersuchungen. Schließlich soll das Abitur in Zukunft nicht mehr schulspezifisch, sondern als Zentralabitur durchgeführt werden.

Nach meinen Erfahrungen sind dazu folgende *Planungsschritte* erforderlich, die hier kurz erläutert werden:

- *Beschreibung des pädagogischen Umfeldes:* Zunächst sind alle Informationen zusammenzustellen, die für die weitere Planung erforderlich sind und für ein einheitliches Verständnis aller Beteiligten sorgen (Struktur des jeweiligen Teils des Bildungssystems, Zielvorstellungen für eine Weiterentwicklung, Schülerzahlen und deren Entwicklung, bisher praktizierte Formen der Leistungsbeurteilung, Lehrpläne, Infrastruktur für neue Leistungsbeurteilungen usw.).
- *Klärung und Definition der Zwecke, die mit den neuen Leistungsbeurteilungen erreicht werden sollen:* Es versteht sich, dass dieser Planungsschritte, der im vorliegenden Beitrag im Vordergrund steht, sowohl enge Bezüge zum pädagogischen Umfeld aufweist als auch die weiteren Planungsschritte determiniert. Diese Bezüge zu beschreiben, ist im Rahmen des vorliegenden Beitrags nur andeutungsweise möglich.
- *Analyse der Bildungsstandards, an denen sich die Leistungsbeurteilung orientieren soll:* Diese Bildungsstandards, mit denen auch Rahmenlehrpläne gemeint sein können, müssen mehrere Bedingungen erfüllen, um als geeignete Grundlage auch für externe Leistungsbeurteilungen verwendbar zu sein. Solche Bedingungen lassen sich mit den Stichworten „Hohe Qualität, Operationalisierbarkeit, klare Struktur" umschreiben.

Diese drei Arbeitsschritte bilden die Voraussetzung, um das umfassende Know-How anzuwenden, das die Pädagogische Diagnostik zu Verfügung stellt. So wird man beispielsweise viele Anregungen für Vergleichsarbeiten unter Termini wie „(informelle) Schulleistungstests" und „Lernzielorientierte Tests" in der Literatur finden. Bei den folgenden Arbeitsschritten wird diese Wissensbasis genutzt:

- *Definition von Qualitätskriterien für die Leistungsbeurteilung:* Für kein Gebiet der Pädagogik stehen seit mehr als einem halben Jahrhundert so eindeutig definierte und weltweit anerkannte Qualitätskriterien zu Verfügung wie für schulische Leistungsbeurteilungen. Große Testinstitute wie der Educational Testing Service in Princeton haben längst erkannt, wie entscheidend eine Qualitätskontrolle ihre Produkte für die Akzeptanz ist und entsprechende Systeme zur Qualitätssicherung (der Tests!) entwickelt (Shaw 1996). Je nach Beurteilungszweck sind die Qualitätskriterien anders zu gewichten. Wenn beispielsweise eine Leistungsmessung für exakte Vergleiche beabsichtigt ist, wird das Kriterium der Objektivität wichtiger als bei Beurteilungen, die beim Erlernen komplexer Kompetenzen eingesetzt werden sollen.
- *Entscheidungen über grundlegende Vorgehensweisen bei der Leistungsbeurteilung:* Hier ist zu klären, ob eine externe Leistungsbeurteilung mit einheitlichen Aufgaben, eine interne Leistungsbeurteilung auf Schul- oder Klassenebene oder eine Mischform praktiziert werden soll. Eine weitere Entscheidung umfasst die Bezugsnormen, mit denen die einzelnen Leistungen verglichen werden sollen: einheitliche statistische oder sachliche curriculare Normen (eben die Bildungsstandards) oder – was ungewöhnlich ist – individuelle Vergleichswerte, die sich an der Ausgangslage des Beurteilten orientieren. Diese Entscheidungen sind ebenfalls im hohen Maße von den angestrebten Zwecken abhängig.
- *Auswahl von Beurteilungsmethoden und Aufgabenformen:* Mit diesem Schritt ist u. a. festzulegen, wie die Einzelergebnisse bei komplexen Aufgabenformen zu bewerten sind.

Die weitere Planung bezieht sich auf Aspekte von Leistungsbeurteilungen, für welche die Informationsgrundlagen auch außerhalb der Pädagogischen Diagnostik zu suchen sind. Wie schon angedeutet, sind vielfache Wechselwirkungen zu beachten zwischen dem pädagogischen Umfeld, den Zwecken, den Bildungsstandards, den Qualitätskriterien und den einzelnen Methoden und Aufgabenformen bei der Leistungsbeurteilung. Der Verfasser verwendet daher ein Planungsmodell, bei dem die Gestaltung von Leistungsbeurteilungen als Entwicklung von Systemen verstanden wird, deren Komponenten sich gegenseitig beeinflussen (Reisse 1977, 2001). Solche Systemen enthalten u. a. noch die folgenden Komponenten:

- *Rechtsgrundlagen:* In Deutschland sind schulische Leistungsbeurteilungen in einen rechtlichen Rahmen eingebunden. Dieser ist zu beachten und – falls erforderlich – anzupassen. Dabei ist nach meinen Erfahrungen eine unzweckmäßige Festlegung schädlicher als eine fehlende.
- *Diagnostische Kompetenz des Lehrpersonals:* Dies bedeutet, die beteiligten Personen auf spezielle Tätigkeiten bei der Leistungsbeurteilung vorzubereiten. Ein typisches Beispiel ist das Training der „Aufgabenersteller". Die Erfahrung zeigt, dass es effizienter ist, diese Personen mit Hilfe von Anleitungen zu schulen als mangelhaft formulierte Aufgaben zu korrigieren.

• *Infrastruktur:* Besonders wenn flächendeckend jährlich externe Leistungsbeurteilungen mit schneller Ergebnisrückmeldung realisiert werden sollen, benötigt man dafür eine ausgebaute Infrastruktur (Stichworte sind Aufgabenbanken, Datenverarbeitungsressourcen, Ablauforganisation).
• *Verwendung der Ergebnisse:* Wer erhält wann welche Ergebnisse? Dass dies ein brisanter Aspekt ist, zeigen Erfahrungen aus anderen Ländern (Robinson 2001). Auch bei Vergleichsuntersuchungen in Deutschland wird schon erkennbar, dass Informationen über die Leistungen einzelner Schulen schnell in den Brennpunkt unterschiedlicher Interessen geraten können (wie z. B. der Schüler und Eltern, Lehrerinnen und Lehrer, der Schulverwaltung sowie der Medien).

Zusammenfassend kann man festhalten:
• Wie man die hier erwähnten Komponenten von Beurteilungssystemen im Einzelnen gestaltet, ist in hohem Maße abhängig von den Zwecken, die mit der jeweiligen Leistungsbeurteilung erreicht werden sollen.
• Mängel bei der Entscheidung über die Zwecke von Leistungsbeurteilungen sowie fehlende Orientierung der anderen Systemkomponenten an diesen Zwecken können leicht zu Fehlentwicklungen führen, die eine Leistungsbeurteilung in Misskredit bringen.
• Schon an dieser Stelle der Erörterung ist zu erkennen, dass die Gestaltung von umfassenden Leistungsbeurteilungen (large scale assessment) ein hochkomplexes Feld ist – was einfache und schnelle Lösungen ausschließt.

Die folgende Darstellung beschränkt sich auf den zentralen Aspekt der Zwecke, Wirkungen und Nebenwirkungen.

3. Notwendig: Zwecke definieren und Wirkungen analysieren

Zwecke von Leistungsbeurteilungen zu beschreiben, ist nicht nur angebracht, weil Prüfungen und Leistungskontrollen Persönlichkeitsrechte tangieren. Es gibt hierfür auch pragmatische Gründe:
• Wie schon erläutert, sind Leistungsbeurteilungen je nach angestrebtem Zweck anders zu gestalten. Leistungsbeurteilungen erfüllen dann am ehesten ihren Zweck, wenn sie für diesen „maßgeschneidert" sind. Beispielsweise ist je nach Zweck zu entscheiden, welcher Vergleichsmaßstab (normorientiert oder sachorientiert) und welche Form der Ergebnisdarstellung gewählt werden soll.
• Bei einer unklaren Koppelung divergierender Zwecke sind erhebliche Probleme in der Beurteilungspraxis zu erwarten, deren Hintergrund oftmals nicht deutlich wird (vgl. die Evaluationsuntersuchung von Lennartz/Klähn 1987). Wenn Zwecke nicht präzisiert und nach ihrer Bedeutung gewichtet werden, geht das zu Lasten derjenigen, die Leistungsbeurteilungen durchführen. Typisch dafür ist das Dilemma, ausgelöst durch eine auf individuelle Lernförderung angelegten Beurteilung, die aber gleichzeitig als Grundlage für allgemein vergleichbare Abschlusszeugnisse dienen soll – zwei sehr widersprüchliche Zwecke.
• Wenn keine Klarheit über die Zwecke einer Beurteilungsform hergestellt wird, ist häufig nicht erkennbar, dass diese Beurteilungsform für den erhofften oder angenommenen Zweck ungeeignet ist. Ein aktuelles Beispiel: Die Ergebnisse stichprobenartiger nationaler oder internationaler Vergleichsuntersuchungen von Schulleistungen dienen in erster Linie dem Zweck, das Gesamtsystem zu

evaluieren. Als unmittelbare Hilfe für den Klassenunterricht sind sie nach Auffassung von Rolff (2002, S. 337) wenig geeignet. Daher empfiehlt es sich auch, die Zwecke zu ermitteln, für die ein Beurteilungsverfahren wenig hilfreich ist.

- Diskussionen über Beurteilungsverfahren entzünden sich oft an den Verfahren, es geht aber eigentlich mehr um die Zwecke, die unterschiedlich wichtig erscheinen. Dies wird veranschaulicht in einer fiktiven Diskussion über Schulleistungsmessungen in Deutschland (Brügelmann 2002). Eine Zweckdiskussion auch aus verschiedenen Interessenlagen wäre sicherlich effizienter als ein Streit über die Beurteilungsformen und -methoden.
- Nur wenn die Zwecke eindeutig definiert sind, ist eine Kontrolle möglich, ob der angestrebte Zweck erreicht oder verfehlt wird oder ob sogar nicht beabsichtigte Nebenwirkungen überwiegen. Damit wird betont, wie wichtig eine Evaluation besonders von neuen Beurteilungsformen ist (vgl. Abschn. 5), für die, wie bei jeder Evaluation, eine Explikation der Ziele oder Zwecke erfolgen muss (Wottawa 2001, S. 660f).
- Zu Recht wird gefordert, dass Lehrerinnen und Lehrer als Teil ihrer Professionalität maßgeblich mitwirken sollen bei Entscheidungen über Formen und Verfahren der Leistungsbeurteilung und über deren Entwicklung und Anwendung (Schlömerkemper 2002). Erfolgreiche Projekte setzen in der Tat voraus, dass Pädagogen und Psychologen zusammenarbeiten – mit unterschiedlichen Schwerpunkten. Zwecke für Beurteilungen festzulegen ist eindeutig ein Arbeitsfeld, für das die Pädagogen „zuständig" sind und hier ihren Sachverstand einbringen müssen (s. Abschn. 7).

Dies setzt allerdings voraus, dass es möglich ist, die Zwecke der Leistungsbeurteilung genau zu klären und eindeutige Entscheidungen darüber zu treffen. Nach meinen Erfahrungen bei sehr unterschiedlichen Projekten der Leistungsbeurteilung ist eine solche Klärung nicht immer einfach: Die Entscheidungsträger überschauen zunächst nicht das Spektrum der Zwecke von Leistungsbeurteilungen. Unterschiedliche bildungspolitisch begründete Auffassungen sollen alle zumindest auf der Ebene von Ankündigungen berücksichtigt werden, eventuell gibt es sogar im Hintergrund stehenden Zwecke, die ungenannt bleiben sollen. Trotz dieser Schwierigkeiten ist viel erreichbar: Dass erkannt wird, wie wichtig Zweckentscheidungen sind, dass die in Frage kommenden Zwecke aufgelistet werden und dass eine Gewichtung versucht oder auch realisiert wird.

4. Welche Zwecke können mit Leistungsbeurteilungen erreicht werden?

4.1. Zwecke als Kategorie, um Leistungsbeurteilungen zu beschreiben, zu analysieren und zu planen

Funktionen, Ziele, diagnostische Zielsetzungen, Anwendungsmöglichkeiten, Aufgabenfelder, möglicher Nutzen von Leistungsbeurteilungen... das sind Formulierungen, die häufig verwendet werden, wenn es um die hier behandelte Thematik geht. Stattdessen wird in diesem Beitrag der *Begriff des Zwecks* in den Mittelpunkt gestellt: Mit „Zielen" können auch Bildungs- oder Lernziele gemeint sein, Formulierungen wie Aufgabenfelder, Funktionen, Anwendungsmöglichkeiten sind mehr auf Möglichkeiten bezogen. Man kann visionäre Ziele verfolgen, aber keine visionären Zwecke. Zwecke werden dagegen

als eindeutige Absichten verstanden, auch wenn sie nicht explizit formuliert, aber durchaus vorgesehen sind und mit einer speziellen Form der Leistungsbeurteilung erreicht werden sollen.

Zwecke sind abzugrenzen von den eingetretenen *Wirkungen* einer Leistungsbeurteilung, die keineswegs den Zwecken entsprechen müssen. Es können – neben eventuell erreichten Zwecken – auch darüber hinausgehende Wirkungen positiver oder negativer Art auftreten. Man denke an die kaum erwartete Bildungsdiskussion, welche die PISA-Untersuchung in Deutschland bewirkt hat – negativ wäre dagegen eine Einengung des in der Schulpraxis realisierten Zielspektrums wie es häufig als Folge einer externen Leistungsmessung befürchtet wird (Weiland 1999, Boenicke 2000). Diese unerwünschten negativen Wirkungen werden hier entsprechend einem üblichen Sprachgebrauch als Nebenwirkungen bezeichnet. Damit werden *drei verschiedene Wirkungen* differenziert: die beabsichtigten Wirkungen, die damit den Zwecken entsprechen, sonstige positive Wirkungen, die nicht eingeplant waren, und negative Wirkungen als Nebenwirkungen.

4.2. Vier Kategorien, um Zweckbeschreibungen einzuordnen

Beschreibungen von Beurteilungszwecken – soweit sie überhaupt vorliegen – umfassen häufig mehrere Zwecke, sind nicht eindeutig und beziehen sich auf unterschiedliche Abstraktionsebenen. Das erschwert eine systematische Darstellung. Daher wird eine Grobklassifizierung nach vier Kategorien verwendet und eine Zuordnung zu mehreren Kategorien praktiziert.

Ich schlage die folgende *Klassifizierung von Zweckbeschreibungen* vor, die im Folgenden erläutert wird:

a) Förderung von Lehr-/Lernvorgängen
b) Grundlage für Kompetenzzertifikate
c) Hilfen für Laufbahnentscheidungen
d) Qualitätssicherung und Bildungsplanung

Diese Kategorien wurden gewählt, weil sie auch für die pädagogische Praxis deutlich unterscheidbare Anwendungsfelder beschreiben und zugleich unterschiedliche Formen der Leistungsbeurteilung erfordern. Zum besseren Verständnis werden diese Kategorien zunächst voneinander abgegrenzt beschrieben, obwohl sie häufig miteinander kombiniert sind.

4.2.1 Förderung von Lehr-/Lernvorgängen

Ein Hauptzweck der schulischen Leistungsbeurteilung ist, Lehren und Lernen unmittelbar zu fördern, besonders durch *Rückmeldungen im Klassenverband oder in einer anderen Lerngruppe*. Eine Variante sind Beurteilungszwecke, die mittelbar auf eine Förderung von Lehren und Lernen zielen (z. B. Untersuchungen in einer Region über Lernvoraussetzungen zur Information der Lehrer und Lehrerinnen).

Das Spektrum von *Beurteilungsverfahren*, die in diese Zweckkategorie fallen, ist sehr weit. Es umfasst die schon immer üblichen Formen der Leistungsbeurteilung in Schulen, wie mündliche Leistungskontrollen und Klassenarbeiten. Dazu gehören ebenso alle Beurteilungsformen, die im Rahmen von Re-

formbestrebungen entwickelt wurden wie Lern- oder Leistungsberichte, verbale Beurteilungen, Lern- und Entwicklungspläne, Diagnosebögen usw. Als neuere Beurteilungsformen sind Orientierungs- oder Vergleichsarbeiten anzuführen, die auch noch weitere Zwecke erfüllen sollen (d: Qualitätssicherung). Für Deutschland relativ neu sind Portfolios.

Ein besonderer Fall ist die sogenannte *„Zwischenprüfung" in der Berufsausbildung*, die – als Feststellung des Ausbildungsstandes – primär unter die Zweckkategorie „Lernförderung" einzuordnen ist. Gleichzeitig dient diese Leistungsbeurteilung auch (als Nebenzweck) der Qualitätssicherung und ist damit auch als Zweckkombination anzusehen: Ausbildungsberater zum Beispiel einer zuständigen Handwerkskammer können mit einem Ausbildungsbetrieb Kontakt aufnehmen, dessen Auszubildende durch schlechte Ergebnisse aufgefallen sind (Lennartz/Klähn 1987).

In der Zweckkategorie „Förderung von Lehr/Lernvorgängen" werden sehr unterschiedliche spezielle Zwecke zusammengefasst, die zudem mit anderen Zweckkategorien kombiniert werden. Hauptzweck ist aber, *Lehr- und Lernvorgänge unmittelbar zu fördern*. Aus diesem Zweck der Lernförderung, der häufig im Vordergrund steht, leiten sich spezielle Formen von Leistungsbeurteilungen ab (Winter 2004). Dabei sollte aber das naheliegende Missverständnis vermieden werden, dass sich die schulische Leistungsbeurteilung nur an diesem Zweck zu orientieren hat – es gibt auch andere notwendige Zwecke, die andere Formen der Leistungsbeurteilung bedingen.

4.2.2 Grundlage für Kompetenzzertifikate

Isoliert betrachtet, steht eine zweite Gruppe von Beurteilungszwecken völlig im Kontrast zu dieser Kategorie der Lehr/Lernförderung. Weitgehend unbemerkt von der Bildungsdiskussion in Deutschland ist in Europa und auch darüber hinaus das *Konzept des „Kompetenzzertifikats"* entwickelt worden, das in einer EU-Norm kodifiziert wurde (Deutscher Zertifizierungsrat 1989). Stark vereinfacht ist ein solches Zertifikat der Nachweis dafür, dass die zertifizierte Person einen definierten Qualifikationsstandard erfüllt hat. Dabei wird das bekannte Verfahren der Zertifizierung von Produkten, Serviceleistungen und Organisationen auf den Nachweis der Qualifikation von Personen angewandt – ein sehr technokratisches, aber zunehmend praktiziertes Vorgehen. Kompetenzzertifikate spielen im europäischen Raum eine zunehmende Rolle, besonders in der beruflichen Bildung, in der Fortbildung und im Hochschulbereich. Beispielsweise ist die Akkreditierung neuer Studiengänge ein Bestandteil dieses Konzeptes.

Beispiele für Kompetenzzertifikate sind *berufliche Zeugnisse*, zu denen auch Hochschul-Diplome gezählt werden. Beurteilungen – als laufende Beurteilungen, Prüfungen oder Credits – bilden die Grundlage für solche Kompetenzzertifikate. In diese Kategorie muss man aber nicht nur berufliche Zertifikate einordnen, sondern auch Abschlusszeugnisse allgemeinbildender Schulen wie das Abiturzeugnis, wenn dieser Aspekt des Kompetenznachweises im Vordergrund steht. Die Diskussion in einzelnen Ländern über neue Abschlussprüfungen an Haupt- und Realschulen sowie über ein Zentralabitur sind in diesem Zusammenhang zu sehen. Dahinter steht die Auffassung, dass Abschlusszeugnisse verlässlich und glaubwürdig über den erreichten Bildungsstand informieren sollen.

Auch bei den Orientierungs- oder Vergleichsarbeiten ist zu fragen, ob hier nicht durch eine Leistungsbeurteilung nach landesweiten Maßstäben bewirkt werden soll, dass Zeugnisse auf einer einheitlichen Beurteilungsgrundlage beruhen und damit dem Konzept der Kompetenzzertifikate nahe kommen.

Bei meiner Beratungstätigkeit in anderen Ländern wurde deutlich, dass mit der Einführung von Kompetenzzertifikaten auch noch weitergehende Zwecke verfolgt werden, die auf eine generelle Qualitätsverbesserung von Bildung abzielen (d: Qualitätssicherung):

- Anerkennung von allgemeinen oder beruflichen Qualifikationen, die ohne entsprechende formale Bildung erworben wurden.
- Damit verbunden Bildungsexpansion durch ein umfassendes Angebot an Kompetenzzertifikaten.
- Bessere Verbindung zwischen Bildungssystem und Arbeitsmarkt auch als Vorteil für den Zertifikatsinhaber.
- Hohes (und relativ einheitliches) Bildungsniveau eines Landes.

Die Beschreibung der Zweckkategorie „Kompetenzzertifikate" zeigt, dass diese ihren besonderen Schwerpunkt in der Berufsbildung und in der Fortbildung hat, ebenso im Hochschulbereich. Dies ist unter dem Aspekt einer zunehmenden Globalisierung zu sehen, bei der international verwertbare Zertifikate immer wichtiger werden – auch im Interesse der Zertifikatsinhaber. Hier wird offensichtlich, dass den vier Zweckkategorien je nach Bildungsstufe unterschiedliche Wertigkeit zukommt: In der Primarstufe wird es um Lernförderung gehen – in der Sekundarstufe II und in der beruflichen Bildung muss man sich mit Kompetenzzertifikaten auseinander setzen.

4.2.3 Hilfen für Laufbahnentscheidungen

Es ist kein Bildungssystem denkbar, das nicht Entscheidungen über Laufbahnen erfordert, die durch die Struktur des jeweiligen Bildungssystems und die Vorgaben für Übergänge festgelegt sind.

Bei diesen Entscheidungen fließen *viele Gesichtspunkt* ein. Hier interessieren nur die Leistungsbeurteilungen, die dabei von den Entscheidungsträgern – Eltern, Lehrerinnen und Lehrer – mehr oder minder mit einbezogen werden. Laufbahnentscheidungen in der Bildung sind ein sehr umfassender Anwendungsbereich der Pädagogischen Diagnostik. In diesem Beitrag ist nach dem möglichen Zweck einer Leistungsbeurteilung oder -einschätzung als eine Grundlage dafür zu fragen. Dabei müssen Varianten des Gesamtzwecks „Pädagogisch-diagnostische Hilfen für Laufbahnentscheidungen bereitzustellen" voneinander abgegrenzt werden. Einmal ist beabsichtigt, eine Voraussage über den zu erwartenden Schul- oder Studienerfolg des Beurteilten zu treffen. Eine andere Zweckvariante besteht darin, lehr- und lerndiagnostische Informationen den Beurteilten, den Eltern sowie den Lehrerinnen und Lehrern der aufnehmenden Bildungsinstitution zur Verfügung zu stellen. Dies hängt davon ab, ob eine mehr förderungs- oder eine mehr ausleseorientierte Diagnostik praktiziert werden soll. Förderungsorientierte Diagnostik setzt voraus, dass das jeweilige Bildungssystem auch eine Förderung zulässt.

Auch für diese Zweckkategorie gilt, dass die verschiedenen Zweckvarianten auch sehr unterschiedliche Beurteilungsformen und -methoden bedingen, wenn der beabsichtigte Zweck erreicht werden soll.

4.2.4 Qualitätssicherung und Bildungsplanung

Diese Zweckkategorie ist im Zusammenhang mit den internationalen Vergleichsuntersuchungen in den Vordergrund gerückt. Hier werden zwei Richtungen zusammengefasst: So kann der Zweck mehr in einer Qualitätssicherung gesehen werden, ohne dass Systemänderungen geplant sind, oder darin, Informationen für eine notwendig erscheinende Modifikation oder grundlegende Änderung des System zu erhalten.

Bei dem ersten Teilzweck (*Systemmonitoring*) werden Leistungsbeurteilungen als Evaluationsinstrumente eingesetzt, wobei dieser Ansatz einzuordnen ist in die schon bisher vorliegenden Konzepte und Verfahren der Evaluation von Bildungseinrichtungen. Die teilweise in anderen Ländern übliche, aber sehr problematische Praxis, Schulen nur nach den bei vergleichenden Leistungsuntersuchungen ermittelten Ergebnissen (oder nach Prüfungsresultaten) zu evaluieren, hat in Deutschland hoffentlich keine Realisierungschancen.

Für den Zweckaspekt der *Bildungsplanung* stehen Untersuchungen von Schulleistungen, aus denen sich Folgerungen für die Bildungsplanung ableiten lassen. Ein Beispiel dafür sind Leistungsmessungen für Schulsystemvergleiche (Hanisch 2001).

Größere Bedeutung für diesen Zweckaspekt der Bildungsplanung haben neuerdings internationale Vergleichsuntersuchungen gewonnen (u. a. Deutsches PISA-Konsortium 2001). Um den umrissenen Zweck zu erreichen, sind noch zusätzliche Untersuchungen (Arbeitsgruppe 2003) erforderlich.

4.3 Kombinationen von Zwecken

Wie die Beispiele zeigen, sind in der Praxis der Leistungsbeurteilung im Regelfall mehrere Zwecke bei einer Beurteilungsform miteinander kombiniert. Das ist auch prinzipiell sinnvoll – für jeden speziellen Zweck gesonderte Beurteilungsverfahren einzusetzen, könnte auf eine unerwünschte Häufung von Leistungsbeurteilungen hinauslaufen.

Aber kann man überhaupt diese völlig unterschiedlichen Zweckkategorien miteinander kombinieren? Dabei kommt noch hinzu, dass die Zwecke nicht immer genügend präzise spezifiziert werden und eine Gewichtung fehlt. Jeder Zweck erfordert eigentlich eine spezielle Beurteilungsform oder -methode wie beispielsweise Winter (2004) überzeugend für die Lernförderung belegt. Daraus folgert, dass es *keine Beurteilungsform geben kann, mit der man in einer Art „Omnibus-Verfahren" alle Zwecke erreicht.* Insofern ist immer Skepsis angebracht, wenn behauptet wird, endlich das „richtige" Beurteilungsmodell gefunden zu haben, ob es sich nun um verbale Lernberichte, Portfolios oder landesweite Vergleichsarbeiten handelt.

Welche Zweckkombinationen leichter möglich sind und welche auf erhebliche Probleme stoßen, wird an den folgenden Zweckkombinationen erläutert:

4.3.1 „Lernförderung (a)" kombiniert mit „Kompetenzzertifikaten (b)" und/ oder „Laufbahnentscheidungen (c)

Eine solche Kombination wird erwartet, wenn Noten als lernfördernde Informationen pädagogische Wirkungen entfalten und ebenso als vergleichbare Leistungsnachweise den erreichten Bildungsstand bestätigen sollen. Diese schwie-

rige Zweckkombination führt zu Problemen, die man mit Recht als *Dilemma der Leistungsbeurteilung* in der Bildung sehen kann. Hier sind nämlich sehr unterschiedliche Verfahrensansätze erforderlich. Vereinfacht beschrieben: Für die direkte Lernförderung benötigt man Verfahren auf Klassen- oder Schulebene (interne Beurteilung), die sich an der jeweiligen Situation und dem individuellen Leistungsstand des Lernenden orientieren, was notwendigerweise mit einem Verzicht auf Vergleichbarkeit der Ergebnisse erkauft werden muss. Vergleichbare Aussagen sind aber gerade für Kompetenzzertifikate und als prognostische Grundlage für Laufbahnentscheidungen erforderlich. Das Verfahren der Wahl ist dafür eher eine extern kontrollierte Beurteilung nach einheitlichen Maßstäben, die individuelle pädagogische Aspekte weniger berücksichtigen kann.

Die unbefriedigende Situation, dass pädagogische Aspekte bei solchen Kombinationen nur unvollständig realisiert werden können, führt nicht selten zu der Forderung, generell Prüfungen, Tests und andere auf Objektivität abzielende Beurteilungsformen durch *„pädagogische Beurteilungen"* zu ersetzen. Das wäre aber nur für den Preis möglich, dass die anderen Zwecke nicht realisiert werden. Wie man versucht, Lösungen für dies Dilemma zu erarbeiten, kann in diesem Rahmen nicht dargestellt werden. Wie oben erläutert, relativiert sich dieses Problem etwas, wenn man berücksichtigt, dass in der Primarstufe Kompetenzzertifikate und Laufbahnentscheidungen selektiver Art nicht die Rolle spielen sollten, wie auf höheren Stufen des Bildungssystems.

4.3.2 Vergleichende Leistungsmessungen zur „Qualitätssicherung/Bildungsplanung (d)" kombiniert mit „Lernförderung (a)"

Vergleichsuntersuchungen von Schulleistungen werden häufig mit diesem doppelten Zweck geplant und angekündigt. Ein Motiv dafür ist nachvollziehbar: Lehrerinnen und Lehrer sollen dadurch für eine *Mitarbeit* gewonnen werden, dass sie konkret in ihrem Unterricht von den Ergebnissen profitieren können. Das ist aber nicht selbstverständlich, Helmke/Schrader haben begründet, warum dies auf Schwierigkeiten stößt (S. 249). Probleme gibt es überall, wo umfassende Untersuchungen von Schulleistungen (large scale assessment) beiden Zwecken gerecht werden sollen. Bei den neuen Vergleichsuntersuchungen auf Länderebene in Deutschland wird versucht, Lösungen dafür zu finden. Interessante Erfahrungen für diese Koppelung von Lernförderung und Qualitätssicherung liegen über die schon erwähnte „Zwischenprüfung" in der Berufsbildung vor, die seit über 30 Jahren in Deutschland jährlich mit mehreren Hunderttausend Auszubildenden durchgeführt wird.

4.3.3 Abschlussprüfungen mit „Kompetenzzertifikaten (b)" als Grundlage für die „Qualitätssicherung (d)"

Könnte man nicht einfach die Prüfungsergebnisse der einzelnen Bildungseinrichtungen – Schulen oder Hochschulen – als Qualitätsnachweis heranziehen? Dann könnte man sich viel Aufwand ersparen. Dies wurde gelegentlich praktiziert, indem man beispielsweise Bestehensquoten oder Durchschnittsnoten von Gymnasien oder Hochschulen miteinander verglich. Das dies problematisch ist, hat man wohl erkannt.

In anderen Ländern wird eine solche Methode zur Qualitätssicherung nicht selten praktiziert, wobei aber zwei *Voraussetzungen* erfüllt sein müssen:

• Zunächst müssen die Prüfungsergebnisse miteinander vergleichbar sein. Mittel dazu sind zentralisierte externe Prüfungen oder Prüfungen auf Schulebene, die aber „moderiert" werden. Dies ist ein in angelsächsischen Ländern häufig praktiziertes Verfahren. Dabei werden die auf klassen- oder schulspezifischen Maßstäben begründeten Noten so geändert, dass sie vergleichbar sind – in Deutschland kaum vorstellbar. Um Noten vergleichbarer zu machen, gibt es übrigens noch weitere Methoden.
• Wenn solche Auswertungen von Prüfungsergebnissen nicht nur das Leistungsniveau einer Schule widerspiegeln sollen, sondern die Qualität des Unterrichts (das ist zu unterscheiden), dann ist noch eine weitere Voraussetzung zu erfüllen: Bekanntlich hängen die Schulleistungen nicht nur von der Qualität des Unterrichts ab, sondern auch von anderen Faktoren, z. B. von dem Leistungsniveau, das die Kinder mitbringen. Auch unter diesem Aspekt müssen Prüfungsdaten modifiziert und von den Auswirkungen anderer Faktoren bereinigt werden. Dieser zweite Schritt ist übrigens auch erforderlich, wenn bei Vergleichsuntersuchungen die Schulleistung mit extern entwickelten Instrumenten gemessen wird und Rückschlüsse auf die Unterrichtsqualität gezogen werden sollen. Eine solche Datenkorrektur wird auch in Deutschland bei einigen Projekten geplant oder praktiziert (s. Helmke/Schrader 2002, S. 249).

Auf diesem Hintergrund wird deutlich, wie unsinnig es wäre, Prüfungsergebnisse als Nachweis für die *Unterrichtsqualität einer Schule* anzusehen. Das entbindet allerdings nicht davon, den Ursachen auffallend negativer Prüfungsergebnisse bei einer Schule nachzugehen – als legitimes Verfahren der Qualitätssicherung. Dass aber auch mit Kompetenzzertifikaten weiterführende Zwecke im Bereich der allgemeinen Qualitätsverbesserung von Bildung angestrebt werden, ist in Abschn. 4.2.1 beschrieben worden.

5. Wirkungen und Wirkungsforschung

Projekte zur Leistungsmessung und einzelne Formen von Leistungsbeurteilungen können weitergehende oder auch völlig andere Wirkungen haben als ursprünglich beabsichtigt. Ein naheliegendes Beispiel ist die *PISA-Untersuchung*:
• Zunächst wurde der Informationsstand erweitert. Dass Deutschland höchstens im Mittelfeld der beteiligten Länder liegt, ist als Forschungsergebnis besonders interessant, weil es Erwartungen nicht bestätigt, sondern widerlegt. Wenn man nach den Bedingungen für dieses Leistungsniveau fragt, ist der Informationsgewinn durch PISA jedoch nicht so hoch einzuschätzen.
• PISA hat eine kaum vorstellbare öffentliche Diskussion ausgelöst – Bildung ist wieder in den Mittelpunkt öffentlichen Interesses gerückt.
• PISA hat darüber hinaus auch Innovationen bewirkt, wenn man die KMK-Beschlüsse und die entsprechenden Maßnahmen der Länder (KMK 2003, Teil D) als Hinweis darauf betrachtet. Für ein Ergebnis einer Schulleistungsmessung ist das keineswegs selbstverständlich.
• Nahezu unbemerkt ist aber geblieben, dass mit den KMK-Beschlüssen zur Durchführung von Internationalen Leistungsmessungen und der Akzeptanz der Ergebnisse in Deutschland eine bemerkenswerte bildungspolitische Entscheidung getroffen wurde: Deutschland hat zumindest indirekt die den Vergleichsuntersuchungen zugrunde liegenden internationalen Bildungsstandards anerkannt – nach Meinung des Autors ein längst fälliger Schritt.

Auch die wenigen *Evaluationsstudien* zu speziellen Formen der Leistungsbeurteilung bestätigen, dass die Wirkungen sich beträchtlich von den ursprünglichen Zwecken entfernen können. Ein Beispiel ist die Evaluation der Zwischen-„Prüfung" in der Berufsausbildung (Lennartz/Klähn 1967). Wie schon erwähnt, ist hier die Bezeichnung „Prüfung" irreführend, es geht vielmehr um eine Feststellung des Ausbildungsstandes ohne Anrechnungen oder sonstige Belohnungen oder Sanktionen. Dieser Prüfungsform wird nach Einschätzung der Autorinnen wegen der unpräzisen Zweckbestimmung nur eine eingeschränkte Funktion als pädagogisch-diagnostisches Instrument beigemessen. Sie hat aber nach den Befragungsergebnissen dieser Studie indirekte pädagogische Wirkungen: Lerneffekte, die durch die notwendig erscheinende Prüfungsvorbereitung ausgelöst werden, Lerneffekte durch das Kennenlernen der Prüfungssituation (die weitgehend der Abschlussprüfung ähnelt) sowie Feedback-Wirkungen, die im betrieblichen Ausbildungsalltag in dieser Form häufig fehlen (S. 131).

Wie Zwecke und Wirkungen auseinander liegen können, zeigen besonders die unerwünschten Nebenwirkungen von Leistungsbeurteilungen (s.u.).

Die zitierte Evaluationsstudie zur „Zwischenprüfung" zeigt, wie notwendig eine *pädagogische Wirkungsforschung* wäre, spezialisiert auf die Wirkungen, die unterschiedliche Formen von Leistungsbeurteilungen haben. Solche Wirkungsanalysen für Leistungsuntersuchungen auf Länderebene (Rheinland-Pfalz, Brandenburg) sind in Vorbereitung oder werden bereits durchgeführt (Bos/Postlethwaite 2002, S. 255). Ein Beispiel ist das Projekt WALZER = Wirkungsanalyse Leistungsevaluation: Zielerreichung, Ertrag für die Bildungsqualität der Schule und die Rückmeldung von Evaluationsergebnissen (www.uni-koblenz-landau.de). Diese Untersuchungen beziehen sich zu Recht auch auf eine Zweckkombination (d: Qualitätssicherung/Bildungsplanung + a: Lernförderung), die als problematisch anzusehen ist.

6. „Vor Risiken und Nebenwirkungen wird gewarnt"

„Die vielen Fehler und Ungereimtheiten in den Handreichungen für den Lehrer haben bei mir Wut und Verzweiflung erzeugt, weil mein Teil der Arbeit derartig erschwert wurde, dass der gesamte Aufwand sinnlos erscheint ... dass pädagogisches Herzblut und Arbeitszeit von Tausenden von Deutschlehrern skrupellos vergeudet werden". So schreibt eine Lehrerin im Anschluss an Vergleicharbeiten, die im Fach Deutsch in Gymnasien durchgeführt wurden – nachdem sie ihr Interesse an Vergleichsarbeiten bekundet hatte. Ein Beispiel dafür, dass negative Nebeneffekte die positiven Wirkungen übertreffen können.

Nebenwirkungen (damit sind alle Wirkungen negativer Art gemeint) können *in vielen Varianten* auftreten. Zwei besonders kuriose *Beispiele* sind die folgenden Nebenwirkungen von Prüfungen:

- Immobilienbesitzer können über Nacht ärmer oder reicher werden – je nachdem, wie die Prüfungsergebnisse der benachbarten Schulen ausfallen. Dass es ein Problem für Testzentren darstellt, ihre schulspezifischen Ergebnisse vor Immobilienmaklern geheim zu halten, habe ich schon in den 80er Jahren in den USA gehört, kürzlich wurde es mir für England bestätigt (s. auch Jungclaussen 2003). Je nach Schul-Testergebnis werden die Immobilienpreise in der Nachbarschaft modifiziert.

• Die Dialer, die mit Hilfe der berüchtigten 0190-Telefonnummern betrügen, haben eine ergiebige Zielgruppe entdeckt: Die Schüler und Studenten, die gegen Bezahlung im Internet nach Hausaufgaben, Klausuren und Referaten surfen (Die Zeit vom 6. Nov. 20003, Nr. 48, S. 24).

Andere Nebenwirkungen, die Leistungsbeurteilungen nachgesagt werden, sind zwar nicht so absurd, aber trotzdem *von erheblicher praktischer Bedeutung*:
• Leistungstests können die Zielsetzungen der Schule verengen und zu einem verkürzten Verständnis von Bildung beitragen (u. a. Weiland 1999, Boenicke 2000).
• Umfassende und aufwendige Beurteilungen behindern die übrige pädagogische Arbeit (Jürgens 2001).
• Schlecht vorbereitete Leistungsmessungen, die zudem wenig Hilfe für die Unterrichtspraxis bringen, führen zu Frustrationen bei Lehrerinnen und Lehrern.
• Prüfungen fördern die Neigung zu Betrügereien bei Schülern und Studenten.
• Beurteilungen können zu einer „Etikettierung" des (schwachen) Schülers führen.

Die häufig geäußerte Befürchtung, dass besonders flächendeckende Testuntersuchungen zu einer *Verengung der pädagogischer Zielsetzungen und der Unterrichtspraxis* führen, war Gegenstand zahlreicher Untersuchungen und Veröffentlichungen. Auf den umfangreichen und nicht eindeutigen Forschungsstand zu diesen oder anderen Nebenwirkungen kann aber hier nicht eingegangen werden.

Wenn Nebenwirkungen erörtert werden, bezieht sich die Diskussion meist auf solche beobachtete oder vermutete Nebenwirkungen von externen schriftlichen Prüfungen oder sonstigen Leistungsmessungen, bei denen Aufgaben in gebundener Form wie Multiple-Choice-Aufgaben verwendet werden. Damit wird aber das Problem der Nebenwirkungen unzweckmäßig eingeschränkt: Nebenwirkungen sind sehr vielfältig, wie schon die wenigen hier aufgeführten Beispiele zeigen. Die Bedingungskonstellationen, die sie verursachen, sind komplex und nicht nur auf die Eigenheiten einer Beurteilungsmethode bezogen. Nach meinen Erfahrungen sind Nebenwirkungen nicht selten durch *handwerkliche Mängel* bedingt: Die Beurteilungs- oder Testinstrumente werden nicht mit dem nötigen Aufwand entsprechend der jeweils verwandten Methode entwickelt und eingesetzt. Das eingangs in diesem Abschnitt zitierte Beispiel fällt in diese Bedingungskategorie. Auch Faktoren außerhalb der Schule können Nebenwirkungen mit verursachen, wie die Beispiele „Immobilienpreise" und „0190-Dialer" zeigen.

Nebenwirkungen können so schwerwiegend sein, dass sie *die positiven Effekte von Leistungsbeurteilungen zunichte machen* können. So wird aus anderen Ländern gelegentlich von einem ausgesprochenen Desaster bei schulischen Prüfungen berichtet: Viele Beschwerden über schlechte Aufgaben, hohe Zahl falsch ausgewerteter Prüfungsarbeiten, extrem hohe oder niedrige Durchfallquoten, Unregelmäßigkeiten bei Durchführung und Auswertung der Prüfungen und die entsprechende Reaktion in der Öffentlichkeit. Ursachen waren häufig die zu kurzen Vorbereitungs- und Entwicklungszeiten, bedingt durch kurzfristige politisch gesetzte Fristen, zu einschneidende Umstellungen, fehlende Infrastruk-

tur usw. – weniger prinzipielle Mängel der Beurteilungsverfahren (s. für England McCaig 2003).

Wenn man Nebenwirkungen kennt, kann man auch versuchen, sie zu reduzieren. Beispielsweise besteht eine Nebenwirkung darin, dass mit Hilfe von Aufgabensammlungen Prüfungsaufgaben „gepaukt" werden. Daher sind bei einer qualifizierten Testanwendung Aufgabenbänke nicht zugänglich, an deren Stelle werden Aufgabenmuster mit Erläuterungen und didaktisch aufbereitete Übungsmaterialien angeboten, die das Auswendiglernen von Aufgaben wenig attraktiv erscheinen lassen.

7. Folgerungen für die Planung von Leistungsbeurteilungen

Wenn man Vorhaben zur Leistungsbeurteilung plant, neue Beurteilungsformen entwickelt, bei bereits laufenden Projekten die Zweckbestimmung kontrollieren möchte sowie die derzeitige Beurteilungspraxis an kritischen Punkten verbessern will, kann man die hier beschriebene *Zweck- und Wirkungsanalyse* einsetzen. Das gilt – entsprechend modifiziert – für alle Varianten von schulischen Leistungsbeurteilungen wie Vergleichsarbeiten und Tests, um die Bildungsvoraussetzungen zu bestimmen, Vergleichsuntersuchungen auf nationaler oder internationaler Ebene oder Varianten der Abiturprüfung in zentraler Form.

Dabei kommen je nach Vorhaben auf die Schuladministration und auch auf das Lehrpersonal die folgenden Aufgaben zu:

7.1. Den Prozess der Zweckbestimmung initiieren und moderieren

Um die Diskussion und Konsensfindung unter den Beteiligten voranzutreiben, können u. a. folgende Gesichtspunkte für die Einschätzung von Zwecken verwendet werden, die man auch auf die zitierten Beispiele anwenden kann:

- Sind die Zwecke genügend präzis formuliert, um eine Umsetzung und Evaluation zu ermöglichen?
- Werden zu wenig oder zu viele Zwecke vorgeschlagen, die das Vorhaben überladen (was häufiger vorkommt)?
- Welchen Zweckkategorien sind die Zwecke zuzuordnen, werden dabei eventuell Zwecke miteinander kombiniert, die schwierig miteinander zu verbinden sind und sind dann die unterschiedlichen Zwecke gewichtet?
- Sind im pädagogischen Umfeld die Voraussetzungen gegeben oder herstellbar, um die Ergebnisse von Beurteilungen auch in der pädagogischen Praxis umzusetzen (z. B. Fördermaßnahmen bei lerndiagnostischen Tests)?

Es trägt zu Klärung bei, wenn man Zwecke anhand der Beurteilungsergebnisse diskutiert, die erforderlich sind, um diese Zwecke zu erreichen. Wenn beispielsweise die Lehr-/Lernförderung im Vordergrund steht, ist zu fragen, welche Testergebnisse und welche Ergebnisdarstellungen dann notwendig sind. Diese wird für die einzelnen Zielgruppen – Schüler, Lehrerinnen und Lehrer, Schulverwaltung auf allen Ebenen – auch unterschiedlich sein müssen.

7.2 Klären, ob die Ziele mit dem geplanten oder praktizierten Beurteilungskonzept erreichbar sind

Dies bedeutet, dass mit Experten für Leistungsmessung und Testentwicklung bereits im Planungsstadium kooperiert werden muss, um zu erkunden, ob die

vorgesehene Beurteilungsform oder -methode den Zwecken entspricht und welche Konsequenzen sich daraus für das gesamte Vorhaben ergeben. Beispielsweise wird ein Testexperte darauf hinweisen, dass sich ein Problem ergibt, wenn eine Vergleichsarbeit zur individuellen Lernförderung und gleichzeitig als Grundlage für ein Übergangsgutachten mit prognostischen Aussagen verwendet werden soll. Hierfür müssten unterschiedliche Vergleichsmaßstäbe herangezogen werden, was auf Schwierigkeiten stößt. Diese lassen sich lösen, wenn man einem der beiden Zwecke Priorität zumisst.

7.3 Die Verwendung der Ergebnisse und parallele pädagogische Maßnahmen planen.

Viele der hier beschriebenen Zwecke von Leistungsbeurteilungen sind nur zu erreichen, wenn entsprechende pädagogische Maßnahmen – abgestimmt mit der jeweiligen Leistungsbeurteilung – bereits praktiziert oder rechtzeitig entwickelt werden. Ein anschauliches Beispiel ist dafür das Projekt Bärenstark (2002), bei dem eine Sprachstandsermittlung mit einem informellen Test und entsprechende Fördermaßnahmen miteinander kombiniert werden. Wie notwendig eine Koppelung pädagogischer Maßnahmen mit entsprechenden Leistungsbeurteilungen ist, wird besonders deutlich, wenn diese zur Lernförderung gedacht sind. Aber auch bei Kompetenzzertifikaten ist zu fragen, wie pädagogisch mit denjenigen umzugehen ist, die den vorgegeben Standard nicht erreicht haben – im ersten Anlauf oder auch bei Prüfungswiederholungen. Auf die pädagogischen Aspekte von Beurteilungen für Laufbahnentscheidungen wurde schon eingegangen. Dass Leistungsvergleiche zur Qualitätsverbesserung wenig beitragen, wenn sie nicht zu pädagogischen Konsequenzen führen, ist offensichtlich. Kurz: Auch die Verwendung der Ergebnisse entscheidet im erheblichen Umfang darüber, ob die jeweilige Leistungsbeurteilung einen sinnvollen Zweck erfüllt.

Im vorliegendem Beitrag wurde eine Vielzahl von Gesichtspunkten und komplexen Zusammenhängen aufgezeigt, die bestimmen, ob schulische Leistungsbeurteilungen als notwendige Teile pädagogischen Handelns gesehen werden oder eher als Fehlentwicklungen einzuschätzen sind. Dabei kommt der Definition von Beurteilungszwecken eine *wichtige Steuerungsfunktion* zu. Vielleicht kann diese Darstellung der Problematik, aber auch der Gestaltungsgesichtspunkte bei schulischen Leistungsbeurteilungen dazu beitragen, dass diese in unser Bildungssystem integriert werden – und nicht durch unüberlegtes Vorgehen in einigen Jahren zu einem Roll-back führen.

Literatur

Arbeitsgruppe „Internationale Vergleichsstudie" 2003: Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten. Frankfurt/Berlin: DIPF (www.dipf.de)

Bärenstark. Berliner Sprachstandserhebung und Materialien zur Sprachförderung für Kinder in der Vorschul- und der Schuleingangsphase 2002. Berlin: Senatsverwaltung für Bildung, Jugend und Sport (www.senbjs.berlin.de)

Boenicke, Rose 2000: Leistung und Qualitätssicherung. In: Die Deutsche Schule, 92, 2000, 2, S. 139-151

Bos, Wilfried; T. Neville Postlethwaite 2002: Internationale Schulleistungsforschung: Ihre Entwicklung und Folgen für die deutsche Bildungslandschaft. In: F. E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim/Basel: Beltz, S. 251-268

Brügelmann, Hans 2002: Kontroversen um die Schulleistungsmessung in Deutschland. Eine fiktive Diskussion über Positionen und Perspektiven in verteilten Rollen. In: F. E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim/Basel: Beltz, S. 33-44
Deutscher Zertifizierungsrat 1989: Allgemeine Kriterien für Stellen, die Personal zertifizieren. EN 45013. Berlin: Beuth
Deutsches PISA-Konsortium (Hg.) 2001: PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich
Ditton, Hartmut; Daniela Merz, Tanja Edelhäußer 2002: Einstellung von Lehrkräften und Schulleitern zu zentralen Testuntersuchungen an Schulen. In: Empirische Pädagogik, 16, 2002, 1, S. 17-33
Hanisch, Günter 2001: Systemvergleiche. In: D. H. Rost (Hg.): Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz, Psychologie-VerlagsUnion, S.615-622
Helmke, Andreas; Friedrich-Wilhelm Schrader 2002: Jenseits von TIMMS: Messungen sprachlicher Kompetenzen, komplexe Längsschnittstudien und kulturvergleichende Analysen. Ergebnisse und Perspektiven ausgewählter Leistungsstudien. In: F. E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim/Basel: Beltz, S. 237-250
Jungclaussen, John F. 2003: Testen, testen, testen. In: Die Zeit Nr. 32, S. 61
Jürgens, Eiko 2001: Lern- und Leistungsbericht. In: Die Deutsche Schule, 2001, 93, 4, S. 470-485
Kingdon, Michael 2001: Finite Human Resources – the approaching assessment crisis in England. Paper 27th Conference International Association for Educational Assessment. Rio de Janairo: IAEA
KMK 2003: Bildungsbericht Teil D: Maßnahmen der Länder. Bonn: KMK (www.kmk.org.de)
Lennartz, Dagmar; Margitta Klähn 1987: Die Zwischenprüfung in der Berufsausbildung. Berlin: Bundesinstitut für Berufsbildung
Lolwana, Peliwe 2001: If learner Assessments is the Tail that Wags the Dog: what is the dog in school education? Paper 27th Conference International Association for Educational Assessment. Rio de Janairo: IAEA
McCaig, Colin 2003: The Politics of Exam Crisis. Paper 29th Conference International Association for Educational Assessment. Manchester: IAEA (www.aqa.org.uk)
Reisse, Wilfried 1977: Prüfungen und lerndiagnostische Maßnahmen in der beruflichen Bildung – ein systemtechnischer Ansatz zu ihrer Verbesserung. In: Berufsbildung in Wissenschaft und Praxis 6, 1977, 4, S. 5-10
Reisse, Wilfried 2001: Development and Improvement of Assessment & Certification Systems in Vocational Education and Training. Paper 27th Conference International Association for Educational Assessment. Rio de Janairo: IAEA
Robinson, Colin 2001: In the Eye of the Beholder: Perception of Quality in Examinations. Paper 27th Conference International Association for Educational Assessment. Rio de Janairo: IAEA
Rolff, Hans-Günter 2001: Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit in Schulen? In: F. E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim/Basel: Beltz, S. 337-352
Rost, Detlef H. (Hg.) 2001: Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz, Psychologie-VerlagsUnion
Schlömerkemper, Jörg 2002: Leistungsmessung und die Professionalität des Lehrerberufs. In: F. E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim/Basel: Beltz, S. 311-322
Scott, Eva Poon 2001: Public Examinations – Servant or Master. Paper 26th Conference International Association for Educational Assessment. Jerusalem: IAEA
Shaw, Cecil T. 1996: Qualitätssicherung am Beispiel des Educational Testing Service als international arbeitende Prüfungs- und Testinstitution. In: Berufliche Bildung – Kontinuität und Innovation. Dokumentation des 3. BIBB Fachkongresses. S. 866-869. Berlin: BIBB

Weiland, Dieter 1999: Evaluation muss einem anspruchsvollen Leistungsbegriff entsprechen! In: Die Deutsche Schule, 91, 1999, 4, S. 409-410
Winter, Felix 2004: Leistungsbewertung. Schneider/Hohengehren
Wottawa, Heinrich 2001: Evaluation. In: A. Knapp; B. Weidemann (Hg.): Pädagogische Psychologie. Weinheim/Basel: Beltz, S. 647-574

Wilfried Reisse, geb. 1933, Dr. phil., langjähriger Leiter der Abteilung Prüfungen und Lernkontrollen im Bundesinstitut für Berufsbildung (bis 1997). Internationale Beratungstätigkeit, Mitglied der International Association for Educational Assessment (IAEA). Arbeitsschwerpunkt: Systementwicklung für Prüfungen und Lernkontrollen.

Anschrift: Deidesheimer Str. 11, 14197 Berlin;
Email: wm-reisse@t-online.de