

Jutta Mägdefrau
(Hrsg.)

Schulisches Lehren und Lernen

Pädagogische Theorie
an Praxisbeispielen



KLINKHARDT

MÄGDEFRAU
SCHULISCHES LEHREN UND LERNEN

SCHULISCHES LEHREN UND LERNEN
Pädagogische Theorie an Praxisbeispielen

herausgegeben von
Jutta Mägdefrau

VERLAG
JULIUS KLINKHARDT
BAD HEILBRUNN • 2010



Bildnachweis Umschlagfoto: privat

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet abrufbar über <http://dnb.d-nb.de>.

2010.10.d. © by Julius Klinkhardt.

Das Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt.

Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Druck und Bindung: AZ Druck und Datentechnik, Kempten.

Printed in Germany 2010.

Gedruckt auf chlorfrei gebleichtem alterungsbeständigem Papier.

ISBN 978-3-7815-1765-3

Inhaltsverzeichnis

Vorwort der Herausgeberin	7
1 <i>Gerhard Tulodziecki, Bardo Herzig & Silke Grafe</i> Leistungsmessung und Leistungsbeurteilung	10
2 <i>Birgit Brouër</i> Selbstbeurteilungsfähigkeit der Schülerinnen und Schüler fördern	33
3 <i>Jutta Mägdefrau</i> Klassenführung	49
4 <i>Hanna Ferdinand</i> Interessenförderung	68
5 <i>Michaela Gläser-Zikuda</i> „Heute hat es im Unterricht wieder Spaß gemacht!“ Zur Bedeutung von Lern- und Leistungsemotionen in Schule und Unterricht	85
6 <i>Claudia Kastens & Frank Lipowsky</i> Hausaufgaben	109
7 <i>Ludwig Haag</i> Individualisierender Unterricht	127
8 <i>Maria Hallitzky</i> (Was) Lernen Schüler/innen in Phasen offenen Unterrichts?	141
9 <i>Torsten Nicolaisen & Waldemar Pallasch</i> Lerncoaching – Überlegungen und Erfahrungen	156
10 <i>Florian Söll</i> Sozialformen des Unterrichts	173
11 <i>Marita Kampshoff & Martina Walther</i> Mädchen und Jungen in der Schule	190

12	<i>Katharina Schwindt</i> Verknüpfung von Theorie und Praxis am Beispiel der Unterrichtswahrnehmung	212
13	<i>Johannes Mayr</i> Lehrerpersönlichkeit	232
14	Rudolf Kammerl Medien in der Schule	250
15	<i>Guido Pollak</i> Pädagogische Semantik und sozialwissenschaftliche Statistik – Fallarbeit mit empirischem Datenmaterial in der Lehrerbildung	268
	Autorenspiegel	294

Vorwort

Was haben erziehungswissenschaftliche Theorie und schulische Praxis miteinander zu tun? Diese Frage begegnet Lehrenden an Hochschulen in universitären Seminaren ebenso wie Lehrpersonen in Seminaren der zweiten Phase und schließlich wieder Lehrenden in der Lehrerfortbildung. Offenbar erschließt sich die Bedeutung theoretischen Wissens selbst dann nicht, wenn man den alltäglichen Handlungsproblemen der pädagogischen Praxis ausgesetzt ist.

Zentrales Anliegen dieses Buches ist es, zu zeigen, dass erziehungswissenschaftliche Theorien und empirisch gewonnene Erkenntnisse für die Praxis bedeutsam sein können. Das gelingt vielleicht besser, wenn vorher einige Missverständnisse angesprochen werden, die für das Theorie-Praxis-Verhältnis in der Pädagogik kennzeichnend sind und die oft bereits in der Lehrerbildung an den Hochschulen ihren Ausgang nehmen. Das erste dieser Missverständnisses ist das *Transfer-Missverständnis*:

Eine frisch aus dem ersten Schulpraktikum zurückgekehrte Studentin sagt in der Besprechung ihres Praktikumsberichts: „Aus den Theorien, die ich an der Uni lerne, möchte ich Handlungsanweisungen für den Unterricht ableiten und so theoretisches Wissen in die Praxis transferieren. Ich muss z.B. dringend Theorien dazu kennenlernen, was man machen soll, wenn ein Kind dauernd den Unterricht stört!“

Mit dieser Erwartung an die universitäre Lehrerbildung wird die Studentin wahrscheinlich enttäuscht werden, selbst dann, wenn sie fleißig „Theorien lernt“. Empirisch gewonnene Theorien sind Systeme (vorläufig) wahrer Aussagen, die in nomologischer (auf Gesetzmäßigkeiten beruhender) Form für definierte Bedingungen Gültigkeit beanspruchen. Handlungsprobleme der Praxis jedoch sind „Beispiele für etwas“. Für was aber sie ein Beispiel sind, muss der Praktiker oft blitzschnell in einer Handlungsdrucksituation erkennen. Ihm bleibt keine Zeit, lange nachzudenken, welche Theorie in dieser Situation ‚passt‘ und kann daher keine Handlungsanweisung deduzieren. Eine und dieselbe pädagogische Situation kann durch eine Vielzahl von Theorien erklärbar oder verständlich werden. Das Störverhalten eines Kindes im Unterricht könnte z.B. entwicklungstheoretisch erklärt werden, motivationstheoretisch, bedürfnistheoretisch, lerntheoretisch, lehrtheoretisch und so fort.

Verfügt jemand über keinerlei theoretische Wissensbestände, erklärt er sich seine Praxis vor dem Hintergrund seiner eigenen Erfahrungen und derjenigen seiner Berufsgruppe, in die er ‚einsozialisiert‘ wurde. Theoretisches Wissen hingegen ermöglicht eine kritische Reflexion seiner Praxis und der diese Praxis fundierenden subjektiven Theorien. Die Anleitung der Praxis ist daher ein Ergebnis der Reflexion von Theorie vor der Folie praktischer Erfahrung von Einzelsituationen, die als „Beispiel für etwas“ erfahrbar werden, Anleitung von Praxis entsteht nicht durch direkten „Transfer“ oder „Anwendung“ einer Theorie.

Ein mit der Transfervorstellung zusammenhängendes weiteres Missverständnis ist das *Generalisierungsmissverständnis*.

Eine Studentin, die im Rahmen eines Seminars zu Unterrichtsmethoden eine Studie zur Lernwirkung des Gruppenpuzzles vorgestellt hat, formuliert: „In meiner Studie kam raus, dass Schüler, die mit der Methode des Gruppenpuzzles unterrichtet wurden, zwar hinterher nicht unbedingt mehr wussten als Schüler, die normalen Klassenunterricht mit Lehrerinstruktion hatten, aber immerhin mit mehr Interesse und Lernfreude bei der Sache waren.“ Und engagiert fügt sie hinzu: „Also kann ich das auf meinen Unterricht anwenden.“

Dieses Buch zeigt unter anderem, wie durch empirische Unterrichtsforschung Wissen generiert wird und wozu dieses Wissen „taugt“ und wozu nicht. Eine bestimmte Unterrichtsmethode, von der ich wissen möchte, ob sie hinsichtlich der Schulleistungsentwicklung einer anderen überlegen ist, erweist sich unter definierten Bedingungen in einer bestimmten Altersstufe einer Schulform in einem Unterrichtsfach, nehmen wir Englisch, einer anderen Methode überlegen. Aber ist sie auch in einer anderen Klassenstufe, einer anderen Schulform und im Fach Deutsch überlegen? Und womit wurde die Methode in der Studie eigentlich konkret verglichen? Was also ist „normaler Klassenunterricht mit Lehrerinstruktion“? Gerade weil empirische Befunde oft so feinkörnig sind und Ergebnisse daher nicht mit grobem Pinselstrich verallgemeinert werden können, sind pädagogische Theorien bis heute so unvollständig. Wir müssen uns mit vorläufigem Wissen bescheiden und an der Vervollständigung der Theorien weiterarbeiten. Das heißt: Selbst wenn jemand Forschungsbefunde kennt, heißt das nicht, dass er ihre Ergebnisse eins zu eins auf seinen Unterricht übertragen kann, sie sind nicht per se generalisierbar, führen also nicht zu Gewissheiten, wie zu handeln ist, um ganz sicher erfolgreich zu sein.

Welche Rolle spielen vor diesem Hintergrund die Fälle, die Praxisbeispiele in diesem Buch? Es liegen – das zeigen die Beiträge in diesem Buch eindrucksvoll – bereits eine ganze Menge einigermaßen belastbarer Ergebnisse darüber vor, unter welchen Umständen Kinder in der Schule besser oder schlechter lernen, welches Lehrerverhalten lernförderlich oder -hemmend ist. Vor dem

Hintergrund der oben ausgeführten Überlegungen zum Theorie-Praxis-Verhältnis, die den Anstoß zu diesem Buch gaben, wird vermutlich bereits andeutungsweise klar, warum in diesem Buch „Fälle“ aus der Praxis und empirische Forschung bzw. empirisch gewonnene Theorien miteinander verwoben werden.

Die Autorinnen und Autoren in diesem Buch wurden gebeten, das von ihnen bearbeitete Thema (schul)pädagogischen Handelns vor dem Hintergrund eines Falles aus der Praxis oder eines konkreten Handlungsproblems darzustellen. Sie werden merken, dass die verschiedenen Autoren diese Verschränkung von Theorie und Praxisfall ganz unterschiedlich gelöst haben, was – so meine Hoffnung – zum Lesevergnügen beiträgt.

Gemeinsam ist allen Mitwirkenden die Überzeugung, dass Theorie und Empirie durchaus Antworten haben für die Bewältigung pädagogischer Probleme in der Schule, gemeinsam ist ihnen aber auch, dass wir noch immer vieles nicht wissen darüber, wie Menschen beim Lernen „funktionieren“, dass Fragen offen bleiben oder eine Ungewissheit bleibt, ob die gefundenen Antworten richtig sind.

Ich hoffe, dass dieses Buch einen Beitrag dazu leistet, den Leser oder die Leserin dafür zu gewinnen oder darin zu bestärken, dass die Vorläufigkeit des Wissens über Lehren und Lernen herausfordernd, ja vielleicht sogar vergnüglich ist und uns bestärkt, an den spannenden Fragen rund um die optimale Ermöglichung menschlichen Lernens weiterzuarbeiten.

Ich danke allen an diesem Buch beteiligten Autorinnen und Autoren für die hervorragende Zusammenarbeit und ihre Bereitschaft, sich auf die „Fälle“ einzulassen und ganz besonders Christine Herrmann für ihre zuverlässige und kompetente Vorbereitung der Druckvorlage.

Jutta Mägdefrau
Passau im Juli 2010

1 Leistungsmessung und Leistungsbeurteilung

1 Ausgangslage und Problemstellung

Die Erfassung, die Feststellung und die Bewertung von Lernleistungen können als zentrale Aufgaben von Lehrpersonen bezeichnet werden. Wie immer die Überprüfung des Lernerfolgs und seine Bewertung erfolgen – für Kinder und Jugendliche haben sie eine große Bedeutung, entscheiden sie in unserem Bildungssystem doch z.B. darüber, ob angestrebte Abschlüsse erreicht und weitere Bildungschancen eröffnet oder verwehrt werden.

Vor diesem Hintergrund ist es nicht verwunderlich, dass Lern- und Leistungskontrollen als schulische Aufgabe von Lehrpersonen in einer spannungsreichen Diskussion stehen und in unserem Bildungssystem zum Teil mit Widersprüchlichkeiten belastet sind. Ein Widerspruch besteht z.B. darin, dass Kinder und Jugendliche zwar als Subjekte ihrer Lernprozesse gelten sollen, bei Lernerfolgskontrollen unter Umständen jedoch zu Objekten gesellschaftlicher Ansprüche bzw. externer Kontrollen werden. Des Weiteren entsteht ein Spannungsfeld zwischen der Forderung nach einer individuellen Förderung der Kinder und Jugendlichen einerseits und dem – mit der Beurteilung von Leistungen verbundenen – Beitrag zu Selektionsentscheidungen andererseits. Nicht zuletzt führen die ideologischen Prämissen des Leistungsbegriffs bei einer Übertragung des gesellschaftlichen Leistungsprinzips auf die Schule zu einem Spannungsverhältnis.

Die hohe Bedeutsamkeit von Lernerfolgskontrollen und die damit verbundenen Ambivalenzen erfordern im Schulalltag ein sorgsames und reflektiertes Handeln von Lehrpersonen.

Bitte versetzen Sie sich vor diesem Hintergrund in die Situation eines Referendars, der im Fach Politik eine schriftliche Lernerfolgskontrolle durchführen möchte und den in Abbildung 1 dargestellten Entwurf vorbereitet hat.

Versuchen Sie bitte zunächst, die Lernziele zu beschreiben, die der Referendar mit diesem Entwurf vermutlich überprüfen wollte.

Wie viele Punkte würden Sie für die richtige Lösung bei den einzelnen Aufgaben vergeben?

Wie viele Punkte sollten für die Noten „mangelhaft“ bis „sehr gut“ in der Lernerfolgskontrolle jeweils verlangt werden?

Wie beurteilen Sie den Entwurf des Referendars insgesamt?

1. AUFGABE

Ordnen Sie bitte die vier Begriffe „relative Mehrheitswahl“/ „reine Verhältniswahl“/ „Partielle Mehrheitswahl“/ „personalisierte Verhältniswahl“ den folgenden Kurzbeschreibungen zu:

a) Die Mandate werden nach dem relativen Stimmanteil der einzelnen Parteien vergeben.

b) Ein bestimmter Prozentsatz der Sitze wird nach Mehrheitswahlgrundsätzen, der Rest nach Verhältniswahlgrundsätzen vergeben.

c) Mit der ersten Stimme wird in jedem Wahlkreis ein Kandidat gewählt. Mit der zweiten Stimme wird über den Anteil der Mandate jeder Partei entschieden.

d) In jedem Wahlkreis ist der Kandidat gewählt, der die meisten Stimmen erhält. Die Zahl der Mandate entspricht der Zahl der so gewählten Kandidaten.

2. AUFGABE

Für die Bundesrepublik gilt die „personalisierte Verhältniswahl mit Sperrklausel“. Stellen Sie sich bitte folgende Situation vor: Für die Sitze im Bundestag bewerben sich die fünf Parteien A, B, C, D, E. Für A wird mit einem Stimmenanteil von ca. 43%, für B mit einem Stimmenanteil von ca. 38% gerechnet, für C wird ein Stimmenanteil von ca. 10% erwartet. D und E sind in der Gefahr, die 5%-Grenze knapp zu verfehlen. A will mit D, B mit C koalieren. Die Frage einer möglichen Koalition mit E ist unklar. Welche Vergabe der Erst- und Zweitstimme ist für jemanden am günstigsten, der verhindern möchte, dass B und C die Regierung bilden?

Erststimme für die Partei _____, Zweitstimme für die Partei _____

Begründen Sie bitte Ihre Entscheidung.

3. AUFGABE

Kreuzen Sie bitte die richtigen Ergänzungen an:

a) Der Volkswillen ist im Parlament am ehesten repräsentiert bei dem Wahlverfahren der Mehrheitswahl der Verhältniswahl

b) Die Bildung regierungsfähiger Mehrheiten wird am besten gesichert durch das Wahlverfahren der Mehrheitswahl der Verhältniswahl

4. AUFGABE

Bevor das Bundeswahlgesetz von 1956 beschlossen wurde, waren verschiedene Wahlverfahren in der Diskussion: vor allem die „relative Mehrheitswahl“, die „reine Verhältniswahl“ und die „personalisierte Verhältniswahl mit Sperrklausel“. Die Entscheidung fiel zugunsten der personalisierten Verhältniswahl mit Sperrklausel. Wie bewerten Sie die damalige Entscheidung des Bundestages aus heutiger Sicht?

Abb. 1: Entwurf einer Leistungsüberprüfung im Fach Politik

Um begründete Beurteilungen zu Lernerfolgskontrollen abgeben und selbst geeignete Lernerfolgskontrollen entwickeln zu können, ist es sinnvoll, sich mit folgenden Fragen auseinander zu setzen:

- Welche Möglichkeiten bestehen, wünschenswerte Lernleistungen zu beschreiben?
- Wie können Lernleistungen erbracht und erfasst werden?
- Wie lassen sich Lernleistungen feststellen bzw. messen?
- Was sollte bei der Beurteilung von Lernleistungen beachtet werden?
- Welche Qualitätsmaßstäbe sind an die Feststellung und Beurteilung von Lernleistungen zu stellen?

2 Zur Beschreibung wünschenswerter Lernleistungen

Angestrebte Lernleistungen werden in der Regel in Form von Lernzielen beschrieben. Für die Lernzielformulierung lassen sich verschiedene Abstraktionsgrade, Lernzieldimensionen und Lernzielstufen unterscheiden. Beim Abstraktionsgrad kann man differenzieren zwischen

Richtzielen, z.B.: Eine positive Einstellung zur Demokratie gewinnen.

Grobzielen, z.B.: Verständnis für die Bedeutung verschiedener Wahlverfahren im Hinblick auf Regierungsbildungen und Repräsentanz des Volkswillens erzielen.

Feinzielen, z.B.: Vier verschiedene Wahlverfahren nennen können.

Nach Bloom u.a. (1972) lassen sich affektive, kognitive und psychomotorische Lernzieldimensionen trennen. Das obige Ziel „Eine positive Einstellung zur Demokratie gewinnen“ ist ein affektives Lernziel, das Ziel „Wahlverfahren nennen können“ ein kognitives Lernziel, und das Ziel „Videokamera und Fotoapparat im Rahmen von Erkundungen zum Demokratieverständnis in der BRD handhaben können“ ein psychomotorisches Lernziel.

Innerhalb der einzelnen Lernzieldimensionen können verschiedene Lernziel-niveaus oder -stufen unterschieden werden, die für die Lernzieltaxonomie von Bloom u.a. (1972) (vgl. zur Weiterentwicklung Anderson & Krathwohl 2001) für den kognitiven Bereich im Folgenden mit Beispielen dargestellt sind:

Wissen: Das Jahr der Konstituierung des ersten Deutschen Bundestages nennen können.

Verstehen: Ein Wahlverfahren in seiner Grundidee und den damit verbundenen Intentionen richtig interpretieren können.

Anwenden: Für eine Situation bei einer Wahl – bezogen auf eine bestimmte politische Absicht – die günstigste Stimmvergabe nennen können.

Analyse: Argumentationsstränge und -ebenen in Wahlaussagen erfassen und rekonstruieren können.

Synthese: Einen Fragebogen zur Überprüfung einer Hypothese in einer Erhebung zum Demokratieverständnis in der BRD selbstständig entwickeln können.

Bewertung: Eine eigenständige Stellungnahme zu einem politischen Kommentar mit Bezug auf mehrere Kriterien abgeben können.

Insgesamt gibt es zur Frage der Lernziele und ihrer Formulierung unterschiedliche Einschätzungen. Einerseits wird betont, dass das Nachdenken über Lernziele wichtig für ein angemessenes Anspruchsniveau und eine sinnvolle Zielorientierung des Unterrichts ist; andererseits wird mit Blick auf die Unterrichtsplanung die Gefahr einer zu großen Vorab-Festlegung des Unterrichtsprozesses gesehen – insbesondere wenn die Ziele als Feinziele formuliert werden. Im Hinblick auf eine Lern- und Leistungskontrolle stellt sich dieses Problem allerdings nicht, weil der Unterrichtsablauf im Rückblick reflektiert werden kann und die zu formulierenden Aufgaben ohnehin einen noch höheren Konkretisierungsgrad aufweisen als ein Feinziel. Es ist wichtig, die Schülerinnen und Schüler bei der Unterrichtsgestaltung in den Prozess der Zielbeschreibung im Sinne von Zielvereinbarungen einzubeziehen. In Projektphasen kann ihnen die Zielformulierung auch ganz überlassen bleiben (vgl. zur Umsetzung z.B. Tulodziecki, Herzig & Blömeke, 2004, 101ff; Bohl, 2004, 89ff). Mit der Formulierung von Lernzielen ist die Auswahl von Inhalten verbunden. Diese sollten für eine Leistungsmessung und -bewertung den vorausgegangenen Unterricht in möglichst repräsentativer Weise abbilden. Weiterhin sollten Aufgabenformate bevorzugt werden, die Schülerinnen und Schüler bereits kennen gelernt haben und gewohnt sind.

In Bezug auf das eingangs vorgestellte Beispiel gehen wir davon aus, dass der Referendar in einer vorhergehenden Unterrichtseinheit kognitive Lernziele auf der Wissens-, der Verstehens-, der Anwendungs- und der Beurteilungsebene verfolgt hat. Dabei lässt sich zwischen den folgenden Mindestanforderungen, mittleren und höheren Anforderungen unterscheiden: Die Lernenden sollen mindestens verschiedene Wahlverfahren gegebenen Kurzbeschreibungen richtig zuordnen können, die Konsequenzen von Mehrheitswahl und Verhältniswahl für die Repräsentation des Volkswillens und für die Bildung regierungsfähiger Mehrheiten angeben können. Die Lernenden sollen möglichst für eine Situation bei einer Wahl – bezogen auf eine bestimmte politische Absicht – die günstigste Stimmvergabe nennen und begründen können. Darüber hinaus ist es wünschenswert, dass die Lernenden Wahlverfahren nach politisch relevanten Kategorien bewerten können.

3 Zur Erbringung und Erfassung von Lernleistungen

Für eine Feststellung und Bewertung von Lernleistungen müssen diese zunächst geäußert und erfasst werden. Die Lernleistungen können in drei Varianten erbracht werden: in schriftlicher, mündlicher oder praktischer Form.

Bei der schriftlichen Form können die Aufgaben gebundener, ungebundener oder freier Art sein. Bei gebundenen Aufgaben werden verschiedene Antwortmöglichkeiten vorgegeben, zwischen denen die Lernenden auswählen oder die sie in bestimmter Weise anordnen sollen. Dabei ist z.B. darauf zu achten, dass bei Auswahlantwort-Aufgaben die Auswahlmöglichkeiten – die so genannten Distraktoren – für diejenigen Lernenden, die raten, hinreichend wahrscheinlich klingen und nicht versteckte Lösungshilfen enthalten. Bei ungebundenen Aufgaben muss die Antwort von den Lernenden selbst formuliert werden. (Vgl. z.B. Beiner, 1982, 165ff, Sacher, 2009, 61ff; dort sind auch Beispiele für die verschiedenen Aufgabenformen zu finden.)

Mündliche Lern- und Leistungskontrollen bestehen darin, dass die oder der Lernende etwas in sprechender Weise darstellt beziehungsweise auf Fragen in mündlicher Weise antwortet. In Prüfungssituationen werden die mündlichen Leistungen in der Regel durch ein Protokoll festgehalten. Aufgrund der direkten Kommunikation zwischen Prüfenden und Prüflingen können die Fragen jeweils an das Leistungsniveau angepasst werden.

Als dritte Form der Lern- und Leistungskontrolle bietet sich die praktische Lernerfolgskontrolle an. Hierbei kann es darum gehen, ein Produkt zu erstellen, beispielsweise eine Collage, oder eine Situation zu gestalten, z.B. ein Rollenspiel, oder einen Bewegungsablauf auszuführen, z.B. Hochsprung. Als Leistungsindikatoren können dabei Merkmale des Prozesses, der zu einem bestimmten Produkt oder Ergebnis führt, und/oder das Produkt oder Ergebnis selbst und/oder die Präsentation eines Ergebnisses dienen.

Jede dieser Formen weist hinsichtlich ihrer Eignung zur Überprüfung bestimmter Ziele und Inhalte Vorzüge und Probleme auf. Beispielsweise lässt sich die Frage, ob jemand eine bestimmte Hochsprungtechnik beherrscht, letztlich nur durch eine praktische Prüfung und nicht durch eine schriftliche oder mündliche Prüfung feststellen. Letztere würden die Beschreibung und Einschätzung von Bewegungsabläufen, jedoch nicht ihre Ausführung kontrollierbar machen.

Bei der Darstellung der verschiedenen Erbringungsformen von Lernleistungen ist deutlich geworden, dass Lernleistungen in unterschiedlicher Weise erfasst werden können. So lassen sich verschiedene Formen unterscheiden, in denen die erfasste Leistung vorliegen kann, unter anderem als schriftliche Prüfungsleistung, z.B. als Klausur (vgl. Fallbeispiel), als Protokoll, z.B. von einer mündlichen Prüfung, als Beobachtungsbogen, z.B. zur Erfassung bestimmter Merkmale oder Kategorien bei praktischen Prüfungen, als Ton- oder Videoaufzeichnung, z.B. von mündlichen oder praktischen Prüfungen, als erstelltes Produkt, z.B. als Ergebnis einer praktischen Prüfung, als schriftliche Dokumentation, z.B. im Zusammenhang eines Projektes, oder in einer Mischform, z.B. im Rahmen eines Portfolios.

Mit der jeweiligen Form der Leistungserfassung sind – wie mit der jeweils zugrunde liegenden Prüfungsform – unterschiedliche Möglichkeiten und Probleme verbunden. Auch dabei sind Vorteile und Nachteile abzuwägen. So hätten Tonaufzeichnung und Transkription einer mündlichen Prüfung den Vorteil einer präzisen Dokumentation, wären jedoch gegenüber einem einfachen Protokoll aufwändiger.

Letztlich verlangt die Entscheidung für bestimmte Formen der Leistungserbringung beziehungsweise Leistungserfassung eine Abwägung unter verschiedenen Kriterien, z.B. Gütekriterien (s.u.).

Um auf vorliegende Schwächen in den Aufgabenstellungen aufmerksam zu werden, die Anforderungen – einschließlich der erforderlichen Zeiten – für die Schülerinnen und Schüler besser abschätzen zu können und eine Voraussetzung für eine reflektierte Leistungsfeststellung – gemäß dem nächsten Abschnitt – zu schaffen, empfiehlt sich die Erstellung einer Lösungsskizze.

Im oben vorgestellten Beispiel hat sich der Referendar entschieden, eine schriftliche Prüfungsleistung mit gebundenen und freien Aufgabenformen in der Lernerfolgskontrolle zu verwenden. Die gebundenen Aufgaben bieten z.B. den Vorteil, dass die Auswertungsobjektivität (s.u.) eher gewährleistet ist und die Messung der Lernleistung ökonomisch ist. Jedoch bergen sie beispielsweise eine höhere Wahrscheinlichkeit des Raten der richtigen Antwort. Bei Zuordnungsaufgaben (vgl. Aufgabe 1) zieht eine falsche Zuordnung zudem eine weitere nach sich. Die freie Aufgabenform ermöglicht die Überprüfung einer komplexen Leistung, jedoch hängen beispielsweise Präzision und Tempo der Darstellung stark von der sprachlichen Kompetenz der Lernenden ab.

4 Zur Feststellung bzw. Messung von Lernleistungen

Wenn die Leistung in einer der oben genannten Formen erfasst worden ist, stellt sich die Frage, wie sie im Sinne der Leistungsfeststellung charakterisiert werden kann. Hier bieten sich verschiedene Formen an. Eine – in vielen Fächern verwendete – Form ist die Zuteilung von Punkten. Dabei wird eine erfasste Leistung durch Punkte ausgedrückt. Eine andere Form der Leistungsfeststellung ist die Charakterisierung der Leistung in verbaler Form. Dabei wird die Leistung nach bestimmten Gesichtspunkten, z.B. sachlicher Gehalt eines Aufsatzes oder Aufbau und Gliederung, in Worten beschrieben. Eine dritte Form der Leistungsfeststellung ist mit der Verwendung von Schätzskalen gegeben. Dabei wird eine Leistung hinsichtlich einzelner Kategorien eingeschätzt. Beispielsweise kann der Grad der selbstständigen Leistung bei einer Klausur als „sehr hoch“, „hoch“, „mittel“, „gering“ oder „sehr gering“ eingeschätzt werden. Entsprechende Leistungsfeststellungen können von der Lehrperson oder von den Lernenden selbst oder gemeinsam vorgenommen werden. Die jeweils gewählte Art und Form der Leistungsfeststellung ist von verschiedenen Gesichtspunkten abhängig. Diese reichen von inhaltlichen und

fachlichen Akzenten über ziel- und voraussetzungsbezogene Fragen bis zu institutionellen Rahmenvorgaben (vgl. z.B. Baumann & Dehn, 2004).

Im weitesten Sinne des Wortes handelt es sich bei allen Formen der Leistungsfeststellung um eine Messung. Eine Messung ist dadurch gekennzeichnet, dass ein Sachverhalt in einen Bezug zu einer Bezugseinheit gesetzt wird. Bei der Messung der Länge einer Strecke, der Masse eines Körpers oder der Temperatur eines Gegenstandes ist dies offensichtlich. Mit entsprechenden Maßeinheiten für Längen, Massen oder Temperaturen liegen genau definierte Bezugseinheiten vor. Bei der Messung von Lernleistungen ist dies ungleich schwieriger, weil es keine generell anwendbaren Bezugseinheiten gibt und diese für die jeweilige Leistungsmessung erst definiert werden müssen. Zudem gilt es zu beachten, dass die Messung von Lernleistungen in der Regel nur eine Messung auf dem Ordinal- oder Rangskalenniveau ist (vgl. Sacher, 2009, 34f). Dies bedeutet, dass Formulierungen folgender Art für Lernleistungen möglich sind:

Die Schülerin A verfügt über ein umfangreicheres Wissen zu demokratischen Wahlverfahren als Schüler B.

Der Schüler C verfügt über geringere Fähigkeiten, Verfahren der Berechnung von Sitzverteilungen auf konkrete Wahlergebnisse anzuwenden, als die Schülerin D.

Das Rangskalenniveau einer Messung wird direkt offenbar, wenn man mit einer Schätzsкала arbeitet, also z.B. eine Leistung auf einer Fünferskala von „umfangreiche Kenntnisse“ bis „keine Kenntnisse“ einschätzt.

Auch wenn die verbale Beschreibung einer Leistung oder die Einschätzung auf einer Schätzsкала durch Punktzuordnungen ersetzt wird, ändert dies nichts an der Tatsache, dass damit letztlich nur das Rangskalenniveau erreicht wird. Allerdings können Punktzuordnungen helfen, die Rangskalennmessung zu präzisieren.

Deshalb soll im Folgenden kurz auf Möglichkeiten der Punktzuordnung eingegangen werden. Solche Punktzuordnungen werden in der gängigen Praxis von den Lehrpersonen vorgenommen, grundsätzlich können Schülerinnen und Schüler jedoch an Prozessen der Punktzuordnung beteiligt werden und sie unter Umständen auch selbst übernehmen.

Eine erste Möglichkeit der Punktzuordnung besteht darin, einer richtigen Lösung einen Punkt und einer falschen Lösung keinen Punkt zuzuordnen. Dieses Verfahren ist – wenn es auf alle Aufgaben einer Lern- und Leistungskontrolle angewendet wird – nur dann vertretbar, wenn alle Aufgaben den gleichen Komplexitätsgrad haben und Abstufungen in Lösungsvarianten nicht sinnvoll sind.

Bei unterschiedlichem Komplexitätsgrad von Aufgaben besteht die Möglichkeit, Lösungen oder Teillösungen gemäß dem Komplexitätsgrad zu gewichten.

Dazu kann beispielsweise die Taxonomie der kognitiven Lernziele herangezogen werden. Diese lässt sich so benutzen, dass eine richtige Lösung auf der Ebene des Erinnerns einen Punkt, auf der Ebene des Verstehens zwei Punkte, auf der Ebene der Anwendung drei Punkte und so weiter erhält.

Eine nicht akzeptable, manchmal jedoch (implizit) verwendete Form der Punktzuzuweisung besteht darin, die Punkte nach dem empirischen Schwierigkeitsgrad zu vergeben. Damit ist die Annahme verbunden, dass eine Aufgabe umso schwieriger ist, je weniger Schülerinnen und Schüler sie lösen. So könnte man auf die Idee kommen, für Aufgaben, die beispielsweise nur von 20% der Schüler gelöst werden, 8 Punkte zu vergeben, und für eine Aufgabe, die z.B. von 70% gelöst wird, nur 3 Punkte. Eine solche Art der Punktezuordnung würde zum einen dem pädagogischen Prinzip widersprechen, dass möglichst alle Schülerinnen und Schüler ein gesetztes Lernziel erreichen sollen. Zum anderen würde damit der Vorgang der Leistungsmessung mit dem Vorgang der Leistungsbewertung im Sinne einer sozialen Bezugsnorm vermischt (vgl. dazu den nächsten Abschnitt).

Die am ehesten zu empfehlende Art der Punktzuzuweisung ist nach unserer Auffassung eine Analyse der erwünschten Lösung hinsichtlich der erforderlichen gedanklichen Schritte oder hinsichtlich der zu berücksichtigenden Gesichtspunkte. Wenn z.B. zur Lösung einer Aufgabe sieben gedankliche Schritte vollzogen werden müssen, können bei einer richtigen Lösung sieben Punkte vergeben werden.

Es sei noch einmal ausdrücklich betont, dass die Zuordnung von Punkten zwar ein häufiges Verfahren der Leistungsfeststellung in der Praxis darstellt, dabei jedoch nur als eine von verschiedenen Möglichkeiten gelten kann. Vor allem für sehr komplexe Schülerleistungen mit besonderen individuellen oder kreativen Anteilen – etwa für freie Aufsätze, komplexe Bewegungsabläufe oder künstlerische Produkte und Leistungen – ist es unter Umständen angemessener, die Leistungsfeststellung mit Hilfe von Schätzskaleten oder freien Berichten beziehungsweise Kommentaren vorzunehmen (vgl. dazu z.B. Bohl, 2004, 41ff).

Im oben geschilderten Beispiel ergäbe das Verfahren der Zuordnung nach richtigen oder falschen Lösungen pro Aufgabe einen Punkt. Dieses Verfahren ist aber ungeeignet, weil die Aufgaben einen unterschiedlichen Komplexitätsgrad aufweisen. Das Verfahren der Punktzuzuweisung nach Lernzieltaxonomien ergäbe für Aufgabe 1 (Wissensebene) 4 x 1 Punkt, für Aufgabe 2 (Anwendungsebene) 2 x 3 Punkte, für Aufgabe 3 (Verstehensebene) 2 x 2 Punkte und für Aufgabe 4 sechs Punkte. Dieses Verfahren berücksichtigt die unterschiedlichen Komplexitätsgrade der Aufgaben, eine Erfassung von Teilleistungen bei Aufgabe 4 ist jedoch nicht unmittelbar ableitbar. Das Verfahren nach empirischem Schwierigkeitsgrad würde den Vorgang der Leistungsmessung mit dem Vorgang der Leistungsbewertung vermischen und ist daher ungeeignet. Empfehlenswerter für dieses Beispiel wäre das Verfahren der Punktvorgabe nach Analyse der erwünschten Lösung hinsichtlich der erforderlichen

gedanklichen Schritte bzw. der zu berücksichtigenden Gesichtspunkte, da mit diesem Verfahren auch Teillösungen gut begründet festgestellt werden können.

Bei Aufgabe 1 muss jeweils überlegt werden, ob ein Definitionsmerkmal gegeben ist oder nicht. Daher ergeben sich insgesamt 4 Punkte für diese Aufgabe.

Bei Aufgabe 2 könnte man für die Entscheidung zwischen den vier möglichen Gruppierungen 4 Punkte vergeben und für die Abwägung zwischen Erst- und Zweitstimme für die Parteien A und D nochmals 4 Punkte.

Bei Aufgabe 3 müssen die Informationsfelder mit jeweils zwei Aspekten in Beziehung gesetzt werden, daher ergeben sich insgesamt 4 Punkte.

Bei Aufgabe 4 kann z.B. in Bezug auf die drei Gesichtspunkte „Abbildung des Volkswillens“, „regierungsfähige Mehrheiten“ und „Entwicklung der Parteienlandschaft/Innovationen“ diskutiert werden, inwiefern diese auf gute, mittlere oder schlechte Weise erfüllt werden. Folgt man dieser Sichtweise ergeben sich $3 \times 3 = 9$ Punkte.

Diese Lösung stellt eine von weiteren Möglichkeiten dar. Insbesondere in Bezug auf Aufgabe 4 wären nach diesem Verfahren – je nach behandelten Unterrichtsinhalten – auch andere Punktzuordnungen denkbar.

5 Zur Bewertung von Lernleistungen

Neben der Leistungsfeststellung ist es im schulischen Alltag üblich und in vielen Fällen bei den gegebenen Rahmenbedingungen notwendig, die Leistung zu beurteilen beziehungsweise zu bewerten. Bei der Leistungsbewertung kann man grundsätzlich eine Bewertung durch die Lehrperson von einer Bewertung durch den Lernenden selbst unterscheiden.

Für die Bewertung durch eine Lehrperson ist in Deutschland – im Regelfall – die Vergabe einer Note vorgesehen. Dazu werden im schulischen Alltag die Noten von „sehr gut“ bis „ungenügend“ (1 bis 6) verwendet. Neben der Bewertung durch eine Note gibt es die Bewertung im Rahmen eines freien Wortgutachtens, die vor allem in den ersten Jahrgängen der Grundschule üblich ist. Darüber hinaus gibt es Lern- und Entwicklungsgutachten, bei denen unter anderem mit vorgegebenen Kategorien gearbeitet werden kann. Zum Teil werden Einschätzungen, die frei oder mit Hilfe vorgegebener Kategorien erstellt wurden, auch in Kombination mit einem Zifferzeugnis vergeben. (Vgl. Bohl, 2003, 552ff und 2004, 132f)

Bewertungen, bei denen Lernende ihre eigenen Leistungen beurteilen, können ebenfalls als Noten oder im Sinne einer verbalen Beurteilung erfolgen. Beispielsweise kann ein Portfolio neben der Dokumentation eigener Leistungen, z.B. im Bereich des Sprachenlernens, und ihrer Reflexion im Sinne der Leistungseinschätzung eine ausdrückliche Beurteilung des eigenen Leistungsstandes beziehungsweise der eigenen Lernfortschritte enthalten.

Die Bewertung einer Lernleistung stellt formal gesehen zunächst einen Zuordnungsvorgang zwischen z.B. einer Punktzahl und einer Ziffernote dar. Dies bedeutet, dass die Bedeutung der Note in zentraler Weise von dem

(Mess-)Ergebnis der Leistungsfeststellung abhängt. Mit anderen Worten: eine ungenaue Messung führt immer auch zu einer nicht aussagekräftigen Beurteilung. Eine transparente Leistungsbewertung setzt zudem eine bewusste Entscheidung für eine reflektierte Bezugsnorm voraus. Wie wichtig eine entsprechende Reflexion ist, zeigt z.B. folgende Überlegung:

In der Praxis wird in mehr oder weniger willkürlicher Form die Grenze für eine ausreichende Leistung häufig so gelegt, dass 50 Prozent der Gesamtpunktzahl erreicht werden müssen. Stellt man sich einmal vor, dass für eine Prüfung von Ärztinnen und Ärzten verlangt würde, dass (nur) 50 Prozent der Diagnosen richtig oder 50 Prozent der Operationen erfolgreich abgeschlossen sein müssten, oder dass für eine Prüfung von Piloten nur 50 Prozent der in einem Flugsimulator praktizierten Landungen erfolgreich verlaufen müssten, so wird die Problematik einer solchen 50 Prozent-Grenze unmittelbar einsichtig.

Somit stellt sich die Frage, wie eine pädagogisch sinnvolle Bezugsnorm für Leistungsbewertungen aussehen könnte. Grundsätzlich lassen sich zunächst drei mögliche Bezugsnormen unterscheiden: Lehrziele als Bezugsnorm (kriteriale Bezugsnorm), Leistungsverteilung in einer Gruppe als Bezugsnorm (soziale Bezugsnorm) und eigene frühere Leistungen als Bezugsnorm (individuelle Bezugsnorm).

Diese Bezugsnormen werden im Folgenden mit Bezug auf die zurzeit dominante Praxis der Notengebung diskutiert. Allerdings bleiben die Überlegungen zu den Bezugsnormen auch relevant, wenn statt Ziffernoten Bewertungen in Form von freien oder kategoriengebundenen Gutachten vorgenommen werden (s.u.).

Bei der *kriterialen Bezugsnorm* wird lehrzielbezogen eine Mindestleistung definiert. Diese ist für die Note „ausreichend“ maßgebend. Danach können – ebenfalls lehrzielbezogen – eine befriedigende, gute und sehr gute sowie eine mangelhafte und ungenügende Leistung definiert werden. Bei der kriterialen Bezugsnorm ergibt sich – neben der Transparenz der Anforderungen – im positiven Sinne die Notwendigkeit einer Reflexion der Anforderungen im Sinne ihrer Bedeutung und Angemessenheit. Zugleich schafft sie förderliche Bedingungen für die Kooperation, weil Noten nicht auf einem Vergleich zu anderen Schülerinnen und Schülern, sondern nur auf den erreichten Lernzielen beruhen. Darüber hinaus kann die intrinsische Motivation, d.h. eine positive Einstellung zu Lerninhalten, gefördert werden, wenn die Herausforderungen insgesamt bewältigbar erschienen. Sind Schülerinnen oder Schüler allerdings über- oder unterfordert, zeigt die kriteriale Norm eher negative Effekte auf die Motivation bzw. Anstrengungsbereitschaft (vgl. Schlag, 1995, 114). Rheinberg weist zudem darauf hin, dass eine durchgängige Orientierung an der kriterialen Bezugsnorm die Festlegung von Kompetenzanforderungen für

verschiedene Niveaustufen mit Bezug auf die im Schulgesetz verankerten Notenstufen in allen Schulstufen und Unterrichtsfächern voraussetze (vgl. 2002, 67). Insgesamt liegen jedoch nur wenige Forschungsergebnisse zu Auswirkungen der kriterialen Bezugsnorm vor (vgl. Sacher, 2009, 95).

Bei der *sozialen Bezugsnorm* werden die Noten mit Bezug auf die Leistungsverteilung in einer Zielgruppe festgelegt. Dabei kann beispielsweise eine Leistungsverteilung gemäß der so genannten Gaußschen Normalverteilung angenommen werden (vgl. dazu z.B. Sacher 2009, 87ff). Die für die angenommene Normalverteilung zugrunde liegende Stichprobe kann von einer Klasse bis zur repräsentativen Stichprobe für einen Schuljahrgang reichen.

Bei einer Notenverteilung nach der Gaußschen Normalverteilung geht man von der Annahme aus, dass sich Leistungen derart verteilen, dass die meisten Leistungen im Mittelfeld einzuordnen sind und gute beziehungsweise sehr gute sowie schlechte beziehungsweise sehr schlechte Leistungen in geringerer Häufigkeit auftreten. Eine solche Annahme wird in der Testtheorie z.B. bei Messung von Persönlichkeitsmerkmalen zugrunde gelegt, bei denen vermutet wird, dass sie durch sehr viele Faktoren bedingt sind, die insgesamt keiner unmittelbaren Steuerung unterliegen.

Wird die Gaußsche Normalverteilung der Notengebung unterlegt, kann man mit Hilfe statistischer Maße die Anteile für die einzelnen Noten festlegen. So könnte man für die Notenbereiche z.B. folgende Festlegung treffen (vgl. ebd., 123): In einer nach der Punktzahl erstellten Rangreihe der Schülerleistungen, sollen die ersten 2,5% der Schüler ein „sehr gut“ erhalten, die nächsten 13,5% ein „gut“, die nächsten 34% ein „befriedigend“, die nächsten 34% ein „ausreichend“, die nächsten 13,5% ein „mangelhaft“ und die letzten 2,5% ein „ungenügend“. Bei einer Rangreihe für beispielsweise 26 Schüler würde der Notenspiegel stets so aussehen, dass ein Schüler ein „sehr gut“ erhält, drei Schüler ein „gut“ bekommen, jeweils acht Schülern ein „befriedigend“ und ein „ausreichend“ zuerkannt wird, drei Schüler in ihren Leistungen mit „mangelhaft“ bewertet werden und einer ein „ungenügend“ erhält.

Eine solche Umsetzung einer sozialen Bezugsnorm macht die generelle Problematik dieser Bezugsnorm offensichtlich. Letztlich ist die Bewertung einer Leistung immer nur davon abhängig, wie gut oder schlecht die anderen Schülerinnen und Schüler der Lern- oder Zielgruppe bzw. der Stichprobe im Vergleich zur einzelnen Leistung sind. Die Note ist unabhängig vom Erreichen eines inhaltlichen bzw. lernzielbezogenen Gütemaßstabs. Dadurch kann zum einen Konkurrenzkampf unter Schülern begünstigt werden, zum anderen lassen sich Auswirkungen auf die Motivation und die Ursachenzuschreibung von Leistungsdefiziten befürchten. In ungünstigen Fällen können individuelle Lernzuwächse im sozialen Vergleich ausgeblendet werden, ebenso wie der

Lernzuwachs einer gesamten Gruppe in der Bewertung unberücksichtigt bleibt (vgl. z.B. Rheinberg, 2002, 64).

Außerdem kann man das oben demonstrierte Verfahren unter dem Gesichtspunkt kritisieren, dass die oben genannte testtheoretische Annahme zur Merkmalsverteilung (als Voraussetzung für die Verwendung der Gaußschen Normalverteilung) bei der schulischen Leistungsmessung nicht vorliegt. Schulische Leistungsmessung soll ja stets auf einen Unterricht bezogen sein, in dem bewusst auf eine bestimmte Leistung hin gelernt bzw. gesteuert wird. Aber auch wenn man von vornherein andere Annahmen über die Verteilung der Leistungen in einer Gruppe trifft oder wenn man nach Vorliegen aller Ergebnisse eine Mindestnorm so festlegt, dass ein bestimmter Anteil der Schülerinnen und Schüler mindestens „ausreichend“ oder höchstens ein „mangelhaft“ erhält, handelt es sich um eine soziale Bezugsnorm, sodass die oben aufgezeigte generelle Problematik dieser Bezugsnorm auch in diesen Fällen erhalten bleibt. Darüber hinaus ist aus mathematischer Sicht die Verteilung der Noten nach diesem Verfahren schon deshalb nicht zu rechtfertigen, weil stillschweigend vorausgesetzt wird, dass die Leistungen intervallskaliert erfasst wurden.

Des Weiteren gilt die generelle Problematik der sozialen Bezugsnorm unabhängig davon, ob für die Notenvergabe nur die Leistungen einer Klasse oder mehrerer Klassen oder sogar repräsentativer Stichproben als Bezugspunkte dienen. Stets bleibt die Note – ohne Beachtung von Lernzielbezügen – davon abhängig, wie gut oder schlecht die Leistungen der anderen Schülerinnen und Schüler sind.

Diese Kritikpunkte gelten in erster Linie für die Vergabe von individuellen Noten nach der sozialen Bezugsnorm. Die Kritikpunkte sollen nicht besagen, dass es generell unzulässig sei, individuelle Ergebnisse in Bezug auf Ergebnisse einer größeren Stichprobe einzuordnen, wie dies beispielsweise bei großen nationalen oder internationalen Vergleichsstudien geschieht. Ein solcher Zusammenhang bedarf einer gesonderten Diskussion (vgl. dazu Baumert, Stanat & Demmrich, 2001, 34ff). Schließlich widersprechen die obigen Kritikpunkte nicht dem Anliegen, bei der Festlegung von Mindestnormen im Sinne der kriterialen Bezugsnorm die Leistungsmöglichkeiten der jeweiligen Zielgruppe zu berücksichtigen. Dies kann im Sinne angemessener Schwierigkeitsgrade durchaus sinnvoll sein (vgl. Sacher, 2009, 87ff).

Neben der kriterialen und der sozialen Bezugsnorm ist die *individuelle Bezugsnorm* zu nennen. Hier wird die aktuelle Leistung im Vergleich zur eigenen, früheren Leistung bewertet und danach beurteilt, ob die Leistungen besser oder schlechter geworden sind oder stagnieren. Eine individuelle Bezugsnorm würde auch schwächeren Schülerinnen und Schülern bei kleinen Leistungsverbesserungen eine positive Rückmeldung vermitteln und insofern

zu einer geringeren Misserfolgsängstlichkeit und einer besseren Leistungsentwicklung führen (vgl. z.B. Rheinberg & Krug, 2005, 68). Mit Blick auf die Anstrengungsbereitschaft, die Selbstverantwortlichkeit und die Motivation spricht Heckhausen auch vom motivationspsychologischen Primat der individuellen Bezugsnorm (vgl. Heckhausen & Brunstein, 2006, 182f). Allerdings könnte auf Dauer die Orientierung an einer sachlichen Anforderung beziehungsweise einem inhaltlichen oder lernzielbezogenen Gütemaßstab erschwert werden, weil die Konfrontation mit solchen Normen bei alleiniger individueller Orientierung ausbliebe. Darüber hinaus gäbe es – in Abhängigkeit vom Entwicklungsstand der Kinder- und Jugendlichen – Probleme mit der Vorstellung von einer „gerechten“ Bewertung, weil das Gerechtigkeitsverständnis in bestimmten Entwicklungsphasen auf dem Gedanken der Gleichheit basiert, sodass die Bewertung ungleicher Leistungsergebnisse mit gleicher Rückmeldung oder gar mit gleicher Note als ungerecht empfunden würde.

Insgesamt empfiehlt sich für den schulischen Zusammenhang letztlich die kriteriale Bezugsnorm, wobei diese nach unserer Auffassung – durch eine freie Kommentierung von Leistungen – mit Aspekten der individuellen Bezugsnorm verbunden werden sollte. Solche Verbindungen von kriterialer und individueller Bezugsnorm sind besonders für Länder wie Deutschland wichtig, in denen z.B. im Vergleich zu skandinavischen Ländern schon viel früher und häufiger Noten vergeben werden.

Dieser Argumentation folgend könnte man auf Basis der formulierten Lernziele (vgl. Abschnitt zur Beschreibung wünschenswerter Lernleistungen) mit Rückgriff auf die Lernzieltaxonomie für das Ausgangsbeispiel festlegen, dass für eine ausreichende Leistung mindestens die Wissens- und Verstehensaufgaben richtig gelöst sein müssen. Die Note „ungenügend“ würde bei unvollständigem Wissen (Aufgabe 1) und die Note „mangelhaft“ bei unvollständigem Verstehen (Aufgabe 3) vergeben. Für eine befriedigende Leistung könnte man – im Vergleich zur ausreichenden Leistung – zudem eine gewisse Anwendungsfähigkeit (Aufgabe 2) voraussetzen, für eine gute Leistung die vollständige Lösung von Aufgabe 2 und für eine sehr gute Leistung darüber hinaus Bewertungsfähigkeit (Lösung der Aufgabe 4). Die Benotung sollte durch einen freien Kommentar ergänzt werden, der z.B. die im Rahmen dieser Lernerfolgskontrolle gezeigte Leistung in Bezug zu vorherigen Leistungen setzt.

6 Zu Gütekriterien für Lern- und Leistungskontrollen

Lern- und Leistungskontrollen können an verschiedenen Stellen mit Fehlern behaftet sein. Fehler können bereits im Verhältnis von Aufgabenstellung und Leistungserbringung auftauchen, sie können sich des Weiteren bei der Leistungserfassung, bei der Leistungsfeststellung und der Leistungsbewertung einschleichen. Um solche Fehler kontrollieren zu können, sollte zum einen die Leistungsfeststellung von der Leistungsbewertung getrennt werden und

nacheinander in zwei Schritten erfolgen, zum anderen sollten Lern- und Leistungskontrollen unter Beachtung von Gütekriterien – Objektivität, Reliabilität und Validität – entwickelt beziehungsweise hinsichtlich ihrer Qualität reflektiert werden.

Unter *Objektivität* versteht man, dass das Ergebnis eines Vorgangs unabhängig von den Durchführenden ist. Dabei kann man die Durchführungsobjektivität von der Auswertungs- und der Interpretationsobjektivität unterscheiden (vgl. Sacher, 2009, 36). Die Durchführungsobjektivität bezieht sich darauf, dass eine Lernerfolgskontrolle, z.B. in zwei Parallelklassen, unter gleichen Bedingungen, beispielsweise mit gleichen Hilfsmitteln, durchgeführt wird. Auswertungsobjektivität ist mit der Leistungsfeststellung verbunden und bedeutet beispielsweise, dass zwei oder mehr Gutachterinnen oder Gutachter für eine bestimmte Schülerleistung dieselbe Punktzahl vergeben. Wenn zwei oder mehr Gutachterinnen oder Gutachter für eine vorliegende Arbeit darüber hinaus zu derselben Note als Bewertung kommen, spricht man von Interpretationsobjektivität. Objektivität ist zugleich eine wichtige Voraussetzung für das Gütekriterium der Reliabilität.

Reliabilität bezeichnet die Zuverlässigkeit einer Leistungsfeststellung oder Leistungsbewertung. Empirisch kann sie z.B. dadurch überprüft werden, dass man einen Test nach einer gewissen Zeit wiederholt oder zwei Tests mit gleichen Anforderungen parallel durchführt oder einen Test in zwei anforderungsgleiche Teile aufteilt und die Ergebnisse miteinander vergleicht. Je größer die Übereinstimmung zwischen den Ergebnissen ist, desto höher ist die Reliabilität. In schulischen Zusammenhängen kann eine Reliabilitätsprüfung z.B. auch darin bestehen, dass eine Lehrperson eine Aufsatzbeurteilung nach einem gewissen Zeitraum ohne Erinnerung an die zunächst abgegebene Note wiederholt (vgl. unter anderem Sacher, 2009, 36f).

Validität betrifft die Frage, ob durch eine Lern- und Leistungskontrolle auch wirklich das festgestellt und bewertet wird, was festgestellt und bewertet werden soll. Wenn es beispielsweise die erklärte Zielvorstellung für einen Aufsatz wäre, kreative Ideen zu einer Frage zu entwickeln, so wäre es ein Verstoß gegen die Validität, wenn die Note stärker von den Rechtschreibleistungen als von den geäußerten Ideen abhängt.

Neben diesen drei Hauptkriterien werden die Nebengütekriterien der Ökonomie (vertretbarer Aufwand), der Nützlichkeit und Zumutbarkeit (angemessene Beanspruchung im Verhältnis zur Bedeutung) und der Akzeptanz (Übereinstimmung mit vorhandenen Einstellungen) genannt (vgl. Sacher, 2009, 42). Verschiedene empirische Untersuchungen und zusammenfassende Darstellungen gehen der Frage nach, ob die Hauptgütekriterien für die in der Schule praktizierte Notengebung erfüllt sind oder nicht. Das durchgängige Ergebnis lautet, dass die Gütekriterien bei der schulischen Leistungsbewertung in der

Regel nicht beziehungsweise höchstens zum Teil erfüllt sind (vgl. z.B. Ingenkamp, 1976; Ziegenspeck, 1999; Sacher, 2009). Schon bei der Objektivität gibt es erhebliche Probleme. So stellt z.B. Sacher fest, dass die Koeffizienten für die Übereinstimmung zwischen den Auswerterinnen und Auswertern bei (psychologischen) Intelligenztests zwischen 0,95 und 0,99 (bei einem möglichen Minimum von 0,00 und einem möglichen Maximum von 1,00) liegen, während sie für schulische Leistungsbeurteilungen zwischen 0,35 und 0,85 schwanken (vgl. 2009, 43).

Als Gründe für die relativ geringe Erfüllung der Gütekriterien bei schulischer Leistungsbeurteilung (im Vergleich zu psychologischen Messungen) sind unter anderem verschiedene sachfremde Einflüsse, z.B. Schulform oder Region, sowie Urteilsfehler, z.B. Strenge- oder Mildefehler, und Voreingenommenheiten, z.B. Erwartungseffekte nachgewiesen (vgl. dazu ausführlich z.B. Bohl, 2004, 66ff; Beiner, 1982, 103).

Vor dem Hintergrund dieser Forschungslage stellt sich die Frage, was man tun kann, um Beeinträchtigungen der Gütekriterien entgegenzuwirken. Dazu sind eine Reihe von Maßnahmen denkbar, die hier allerdings nicht alle diskutiert werden können (zumal es dazu in der Literatur hilfreiche Empfehlungen gibt, vgl. Sacher, 2001, 29ff; Bohl, 2004, 68ff).

Wir konzentrieren uns im Folgenden darauf, auf der Grundlage bisheriger Überlegungen wichtige Schritte zur Entwicklung einer Lern- und Leistungskontrolle zusammenzufassen. Ihre sorgfältige Durchführung dient zugleich der Objektivität, Reliabilität und Validität von Lern- und Leistungskontrollen. Die Schritte sind als idealtypisches Grundmuster des Vorgehens gedacht, das in der Praxis flexibel gehandhabt werden sollte.

(1) Auswahl der Inhalte und Formulierung der Lernziele

Hierbei sollen die Prüfungsinhalte und ihr Bearbeitungsniveau den Unterricht möglichst in repräsentativer Form abbilden. Bei Auswahlentscheidungen sollten Aufgaben, die für den weiteren Unterricht grundlegend sind, den Vorrang vor weniger grundlegenden Aufgaben erhalten. Außerdem sollten Aufgaben, die für die Lernenden in Bezug auf ihre Lebenswelt wichtig sind, Vorrang vor Aufgaben haben, die weniger bedeutsam sind. Mit der Bestimmung der Prüfungsinhalte ist es wichtig, sich die Lernziele bewusst zu machen. Falls sie aufgrund vorheriger Unterrichtsplanung noch nicht in ausdrücklicher Form vorliegen, müssten sie für die Lern- und Leistungskontrolle formuliert werden.

(2) Wahl der Form der Lern- und Leistungskontrolle

Die gewählte Prüfungsform – schriftlich, mündlich oder praktisch – sollte geeignet sein, die Inhalte beziehungsweise Lernziele zu überprüfen. Darüber hinaus soll sie der Art der Bearbeitung des Themas im Unterricht weitgehend entsprechen und eine angemessene Form der Leistungserfassung auch im

Hinblick auf die spätere Leistungsfeststellung und Leistungsbewertung ermöglichen.

(3) *Formulierung und Anordnung der Aufgaben*

Für die Formulierung der Aufgaben bei einer schriftlichen Prüfung kann man zwischen gebundenen und freien Antwortformen unterscheiden. Die gewählten Aufgabenformen sollten den zu prüfenden Zielen und Inhalten angemessen und den Schülerinnen und Schülern aus dem Unterricht bekannt sein. Sie sollten verständlich und eindeutig formuliert sein und bezogen auf die Voraussetzungen der Lernenden einen angemessenen Schwierigkeitsgrad aufweisen. Für den Beginn einer Lernerfolgskontrolle sollten Aufgaben mit einem geringeren Schwierigkeitsgrad gewählt werden.

(4) *Entwickeln einer Lösungsskizze*

Eine Lösungsskizze kann der Lehrperson helfen, vorliegende Schwächen in der Aufgabenstellung zu bemerken, die Anforderungen – einschließlich der erforderlichen Zeiten – für die Schülerinnen und Schüler besser abschätzen zu können und eine Voraussetzung für eine reflektierte Leistungsfeststellung zu schaffen.

(5) *Feststellung der Lernleistung*

Die Feststellung der Lernleistung ist ein Messvorgang auf Rangskalenniveau. Dieser „Messvorgang“ kann durch eine charakterisierende Beschreibung der Leistung, durch die Verwendung von Schätzskalen oder durch die Zuordnung von Punkten zu den Lösungen transparenter gemacht werden. Bei der Zuordnung von Punkten ist vor allem eine Zuordnung auf der Basis einer Analyse der richtigen Aufgabenlösung zu empfehlen.

(6) *Festlegen der Bewertung*

Die Bewertung sollte in der Regel nach der kriterialen Bezugsnorm erfolgen. Wünschenswert ist es, über geeignete Kommentare eine Kombination mit der individuellen Bezugsnorm herzustellen. Die Anwendung der kriterialen Bezugsnorm erfordert eine lernzielbezogene Festlegung der Noten.

(7) *Einschätzung nach Gütekriterien*

Für eine Einschätzung einer Lern- und Leistungskontrolle bieten sich die Gütekriterien Objektivität, Reliabilität, Validität und Ökonomie an.

(8) *Durchführung und Reflexion*

Sind die vorgenannten Schritte vollzogen, kann die Lern- und Leistungskontrolle durchgeführt werden. Auf der Basis der Ergebnisse sollte die Lehrperson beispielsweise Fragen nach der Angemessenheit der Annahmen zum Leistungsstand der Schülerinnen und Schüler, zur Angemessenheit der Aufgabenformulierungen, zur Einschätzung der Gütekriterien usw. reflektieren. Mit entsprechenden Reflexionen bleiben nicht nur Gestaltungs- und Durchführungsfragen für Lern- und Leistungskontrollen im Blick, es kann dadurch auch eine insgesamt reflexive Einstellung zum Aufgabenbereich der Lern- und

Leistungskontrolle im pädagogischen und gesellschaftlichen Zusammenhang erreicht werden.

7 Zur Rückmeldung von Leistungsmessungen und -bewertungen

Im Verlauf der bisherigen Überlegungen haben wir uns darauf konzentriert, wichtige Rahmenbedingungen der Feststellung und der Bewertung von Leistungen ins Bewusstsein zu heben. Die Bedeutung dieser Schritte wird deutlich, wenn man sich mögliche Funktionen von Lernerfolgskontrollen – Selektion und Stigmatisierung, Sozialisation, Legitimation, Kontrolle, Prognose, Disziplinierung, Rückmeldung, Diagnose und Leistungserziehung (vgl. Sacher, 2009, 22ff) – vergegenwärtigt. Auf eine dieser Funktionen soll hier noch gesondert eingegangen werden: die Rückmeldung und Information. Lernerfolgskontrollen sollen die Schülerinnen und Schüler, die Eltern oder auch weitere Personen, z.B. Arbeitgeber, darüber informieren, welche Lernstände und Kompetenzen erreicht sind, welche Fortschritte erzielt wurden und wo Entwicklungschancen liegen. Dies bedeutet, dass eine Rückmeldung Informationen verschiedener Art enthalten muss, damit Schülerinnen und Schüler im Verlaufe weiterer Lernprozesse konstruktiv und gezielt an ihrer individuellen Kompetenzentwicklung arbeiten können.

Eine Form der Rückmeldung besteht in der bereits angesprochenen Ziffernote. Darüber hinaus können Leistungen durch verbale Berichte oder durch implizite verbale und nonverbale Formen rückgemeldet werden. In der Diskussion um die Anwendung von Bezugsnormen bei der Bewertung von Leistungen ist deutlich geworden, dass sich neben einer Orientierung an Lernzielen oder Standards auch die Berücksichtigung individueller Leistungsfortschritte oder -rückschritte empfiehlt. Hiermit kann insbesondere die Möglichkeit genutzt werden, spezifische Informationen rückzumelden, die in einer abstrakten Ziffer nicht zum Ausdruck kommen (können).

Aus empirischer Sicht lassen sich mit der Rückmeldung in Form von Ziffernoten verschiedene Problemlagen benennen. So haben Arbeiten von Valtin (2002) gezeigt, dass durch Vergabe von Ziffernoten bei Grundschulkindern Leistungsangst und Misserfolgsorientierung bei leistungsschwächeren Schülern stärker steigen als bei denen, die eine verbale Bewertung als Rückmeldung erhielten. Auf der anderen Seite werden leistungsstarke Schüler durch Ziffernoten eher extrinsisch motiviert (vgl. 37). Auf die Praktikabilität von Zifferzensuren im Vergleich zu verbalen Kommentaren und den geringeren Arbeitsaufwand für die Lehrpersonen verweist Becker (vgl. 2005, 74).

Mit verbalen Rückmeldungen kann die Informationsarmut der Ziffernoten prinzipiell aufgefangen werden. Sie können Stärken und Schwächen klar benennen und förderdiagnostische Anregungen im Sinne konkreter Lernhilfen

bieten. Dies macht jedoch nur Sinn, wenn eine solche Rückmeldung differenziert und individuell erfolgt.

Entsprechende Untersuchungen zeigen, dass dies im Schulalltag „nur in Spurenelementen“ verwirklicht ist (Wagner & Valtin, 2003, 33). Zudem erweisen sich verbale Beurteilungen häufig als interpretationsbedürftig (vgl. z.B. Becker, 2005, 84), was durch Teilstandardisierungen von Formulierungen aufzufangen versucht wird. Die potenziellen Vorteile der verbalen Rückmeldung von Leistungen konnten bisher empirisch allerdings nur in Ausnahmefällen nachgewiesen werden (vgl. z.B. Lißmann & Paetzold, 1982; Lißmann, 1987; Valtin, 2002; Wagner & Valtin, 2003).

Die enorme Bedeutung, die Schülerinnen und Schüler den Ziffernoten beimessen, und ihre hohe sozialisatorische Funktion zeigen sich auch in Untersuchungen von Butler (2005), in denen Variationen von Leistungsrückmeldungen (Ziffernoten, verbale Rückmeldungen, Kombination aus beiden) auf ihre Wirkungen hin geprüft wurden. Die verbalen Rückmeldungen konnten dabei keine besonderen Effekte zeigen: „Sie [Schüler mit kombinierter Rückmeldung] maßen dem Kommentar also weder Bedeutung zu noch lernten sie aus ihm. Die meisten konnten sich nicht einmal an ihn erinnern, als sie gebeten wurden, sich die Inhalte des Feedbacks ins Gedächtnis zu rufen“ (132). Ob diese Ergebnisse die Bedeutung verbaler Rückmeldung generell in Zweifel ziehen, kann nicht sicher bestimmt werden. Möglicherweise handelt es sich eher um ein Akzeptanzproblem, das im Rahmen eines veränderten Umgangs mit Leistungsrückmeldungen einer konstruktiveren Aufnahme solcher Rückmeldungen für die eigene Kompetenzentwicklung weichen könnte.

Mit den bisherigen Überlegungen haben wir auf verschiedene Problemlagen der Leistungsbewertung mit Ziffernoten hingewiesen. Diese Aspekte nehmen wir abschließend noch einmal in grundsätzlicher Weise unter der Frage auf, ob für die Leistungskontrolle Zifferzensuren überhaupt notwendig und sinnvoll sind.

8 Zum Sinn oder Unsinn von Ziffernoten

Gegnerinnen und Gegner der Ziffernote führen vor allem die empirisch nachgewiesene Tatsache ins Feld, dass die Ziffernote eine nicht vorhandene beziehungsweise unangemessene Genauigkeit suggeriere und bei der gängigen Bewertungspraxis weder hinreichend objektiv, noch reliabel, noch valide sei. Außerdem wird auf eine Reihe möglicher unerwünschter Nebenwirkungen verwiesen, beispielsweise dass Ziffernoten zur Demotivierung schwächerer Schülerinnen und Schüler, zu Stress, Angst und Frustrationen, zu bloß extrinsischer (statt intrinsischer) Motivation, zu Konkurrenzverhalten, zu einer Schädigung des Lehrer-Schüler-Verhältnisses oder zu einer Behinderung

einer sinnvollen Leistungserziehung führen können und insgesamt nachteilige Folgen für die Entwicklung der Schülerpersönlichkeit zu befürchten seien. Gegen diese Argumente wird für die Ziffernote z.B. mit den Hinweisen geworben, dass es durchaus Möglichkeiten gäbe, bei der Ziffernote die Objektivität, die Reliabilität und die Validität zu erhöhen, dass die Ziffernote den Vorteil einer schnellen Orientierung für Schülerinnen und Schüler sowie für Lehrpersonen und potentielle „Abnehmer“, vor allem aber auch für die wichtige Gruppe der Eltern habe, dass sie eine wirksame Rückmeldung für Schülerinnen und Schüler darstelle, dass sie der Leistungsentwicklung dienen könne, dass es keine hinreichenden Alternativen zur Ziffernote gäbe und dass mit der Abschaffung der Ziffernote die Gefahr bestehe, dass der Leistungswillen untergraben werde und dass für die Vergabe von Berechtigungen und Positionen nicht mehr die Leistung zähle und damit unkontrollierbaren Einflüssen und irrationalen Kriterien Tür und Tor geöffnet werde (vgl. zur Diskussion u.a. Sacher, 2009, 31 und 165f; Barnitzky & Christiani, 1994, 22ff; Solzbacher & Freitag, 2001).

Ohne diese Diskussion hier bis ins Einzelne wiederholen zu wollen, möchten wir doch drei Punkte betonen, die uns wichtig sind:

Das Argument, dass es Möglichkeiten gibt, die Objektivität, Reliabilität und Validität von Zifferzensuren gegenüber der jetzigen Situation zu verbessern, ist grundsätzlich richtig. Die obigen Überlegungen verweisen darauf, was getan werden könnte, um Zifferzensuren objektiver, reliabler und valider zu machen. Allerdings würde dies einen erheblich höheren Aufwand – gegenüber der jetzigen Handhabung und Praxis – und eine deutlich gesteigerte Qualifizierung von Lehrpersonen für diese wichtige Aufgabe voraussetzen. Außerdem ist festzustellen, dass mit verbalen Rückmeldungen (Wortgutachten) die Probleme der Objektivität, Reliabilität und Validität keineswegs gelöst, sondern zum Teil noch verschärft würden.

Zugleich gilt aber auch, dass ein Wegfall von Zifferzensuren keineswegs den Leistungswillen untergraben müsste. Viele Beispiele aus dem Ausland zeigen, dass ohne Zifferzensuren deutlich bessere Leistungen als in Deutschland möglich sind. Insbesondere die PISA-Studie hat dies offenbar werden lassen (vgl. Deutsches PISA-Konsortium, 2001). Auch in alternativen Schulen oder Schulversuchen in Deutschland zeigt sich, dass schulische Leistungen und Leistungsentwicklung auch ohne Zifferzensuren keinen Schaden nehmen.

Betrachtet man die Funktionen der Lern- und Leistungskontrolle im Hinblick auf die Frage, ob sie durch differenzierte, beschreibende Rückmeldungen oder durch Zifferzensuren besser wahrzunehmen sind, gilt im Grundsatz für alle Funktionen – von der Motivierung bis zur Einordnung –, dass sie durch differenzierte, beschreibende Rückmeldungen beziehungsweise Gutachten besser zu erfüllen wären als durch eine bloße Ziffer.

Allerdings würde das einen höheren Aufwand sowohl beim Verfassen der Rückmeldungen durch die Lehrpersonen als auch bei der angemessenen Entschlüsselung durch Schülerinnen und Schüler, Eltern und sonstige „Abnehmer“ erfordern.

Trotz starker Argumente gegen die Zifferzensur ist nicht davon auszugehen, dass diese in Zukunft ihre Bedeutung verlieren wird. Dafür ist sie zu sehr im Denken und Handeln der im Bildungswesen agierenden Personen sowie in bildungspolitischer und gesellschaftlicher Hinsicht verankert. Vor diesem Hintergrund halten wir die folgenden Aspekte für besonders wichtig.

Grundsätzlich sollten Lernsituationen, in denen man in entspannter und stressfreier Situation unter anderem auch aus Fehlern lernen kann, und Leistungssituationen, in denen es um den Nachweis des Gelernten geht, deutlich unterschieden werden. Außerdem sollte es neben Bereichen, in denen Leistungskontrollen (gegebenenfalls mit Zifferzensuren) notwendig erscheinen, zensurenfreie Bereiche geben. Wenn zeugnisrelevante Bewertungen – als Ziffernoten oder Wortgutachten – vergeben werden, sollte auf eine klare Trennung von Leistungsfeststellung und Leistungsbewertung sowie auf eine größtmögliche Objektivität, Reliabilität und Validität geachtet werden.

Bei der Vergabe von Ziffernoten oder Wortgutachten sollten sich Lehrpersonen, Eltern und „Abnehmer“ sowie Schülerinnen und Schüler (mindestens von einem gewissen Alter an) der Probleme der Objektivität, Reliabilität und Validität sowie der vielfältigen individuellen und gesellschaftlichen Bedingungen schulischer Leistungen möglichst bewusst sein. Für Lehrpersonen und für Eltern ist darüber hinaus wichtig, dass sie in der Lage sind, pädagogische und gesellschaftliche Funktionen von Leistungsdiagnose und Leistungsbewertung zu durchschauen, sowie ein Bewusstsein für Widersprüche zwischen Förderung und möglicher Auslese, zwischen gesellschaftlichem und pädagogischem Leistungsbegriff sowie zwischen gesellschaftlichen Ansprüchen und dem Recht von Kindern und Jugendlichen auf eine erfüllte Gegenwart haben.

Im Übrigen wird (auch) bei der Diskussion um Zifferzensuren und Wortgutachten in der Regel ohne weitere Reflexion davon ausgegangen, dass die Bewertung der Schülerleistungen durch die Lehrperson – letztlich im Sinne einer Fremdbegutachtung – erfolgt. Wenn auch gilt, dass die Lernenden bis zu einem gewissen Grad eine Rückmeldung durch die Lehrperson für ihre Leistungsentwicklung benötigen und die gegenwärtigen institutionellen Rahmenbedingungen die Bewertung durch die Lehrperson erforderlich machen, sollte dafür Sorge getragen werden, dass Schülerinnen und Schüler zu einer selbstständigen Bewertung ihrer Leistungen kommen können und deshalb schon früh in angemessener Weise in Bewertungsprozesse einbezogen werden