

Leistungsbewertung und Notenbildung¹

Einige Betrachtungen
zur schulischen Beurteilungspraxis

Klaus Korossy

Institut für Bildungswissenschaft der
Ruprecht-Karls-Universität Heidelberg

2017

¹ Für eine kritische Durchsicht des Papiers und zahlreiche hilfreiche Anmerkungen danke ich sehr herzlich Prof. Franz Amann.

Gliederung

	Seite
1. Probleme der Notenbildung	2
1.1 Intention des Beitrags	2
1.2 Beispiele zur Praxis der Notenberechnung	4
1.3 Fragestellung und Gliederung des Beitrags	6
1.4 Exkurs zur Problematik der Skalenqualität	7
2. Leistung und Notenbildung	8
2.1 Einige Begrifflichkeiten	8
2.2 Leistungskategorien	12
2.3 Die Zuordnung von „gleichwertigen“ und „Mischleistungen“	13
3. Das Modell der Notenintegration	15
3.1 Noten- und Notenpunkte-Skalen	16
3.2 Grundmodell der Notenintegration	18
3.3 Gewichtung bei der Notenintegration	19
3.4 Mittelwert und Einzelleistungsgewichtung	20
3.5 Kategoriengewichtung als Bedeutsamkeitsgewichtung	24
3.6 Kategoriengewichtung versus Einzelleistungsgewichtung	26
3.7 Zuordnungsproblematik	31
3.8 Zusammenfassung und Empfehlungen	32
4. Logikmodelle der Notenbildung	33
4.1 Praktische Beispiele zur Notenberechnung	33
4.2 Adjunktives Modell (<i>ODER</i> -Modell) der Notenbildung	36
4.3 Konjunktives Modell (<i>UND</i> -Modell) der Notenbildung	37
4.4 Allgemeine Logikmodelle der Notenbildung	38
4.5 Zusammenfassung und Schlussfolgerung	40
5. Dimensionen und Abhängigkeiten von Leistungen	41
5.1 Das Dimensionen-Modell	41
5.2 Taxonomie-Modelle	42
5.3 Modelle mit komplexen Leistungsabhängigkeiten	44
5.4 Notenbildung für taxonomisch geordnete Leistungen	44
5.5 Zusammenfassung und Schlussfolgerung	47
6. Kleines Fazit	48
Literatur	

1 Probleme der Notenbildung

1.1 Intention des Beitrags

Im deutschen Bildungssystem setzt sich zunehmend eine „neue Lern- und Unterrichtskultur“ auf der Grundlage eines erweiterten Lern-, Wissens- und Bildungsverständnisses durch (siehe z.B. Jürgens & Lissmann, 2015, S. 35ff.). Der alt-hergebrachte Leistungsbegriff wird im Zuge der Umstellung auf eine Kompetenzorientierung modifiziert, durch zusätzliche Aspekte angereichert und zu einem „pädagogischen Leistungsbegriff“ erweitert (vgl. etwa Bohl, 2004; Bohl & Kucharz, 2010; Jürgens, 2000; Jürgens & Lissmann, 2015). Neben die traditionelle Leistungsbeurteilung tritt immer gewichtiger die Aufgabe der Lernstands- und Lernprozessdiagnostik im Dienste der Lernförderung; in bestimmten Schultypen (z.B. Sonderschulen, vielen Gemeinschaftsschulen), Schulstufen oder Versuchsschulen werden die Zensuren durch verbale Beurteilungen oder ziffernnoten-begleitende Wortgutachten ersetzt; Bewertungsinstrumente, die sich eher auf die Diagnose des Lernstandes oder des Lernpotenzials richten (Lernentwicklungsberichte, Einschätzungsbögen, Bewertung auf der Grundlage von Portfolios oder Lerntagebüchern, Beurteilungsraster, Selbstbewertung etc.), treten zunehmend auch als Alternativen zu der traditionellen Leistungsbeurteilung auf mit dem Anspruch diese zu ersetzen.

Dennoch behauptet sich die traditionelle Leistungsbeurteilung durch Zensuren in Deutschland² als eine wesentliche und vermutlich auch unverzichtbare Grundlage unseres schulischen Beurteilungswesens in der Erfüllung gesellschaftlich zugeschriebener Aufgaben. Dabei erfolgt im deutschsprachigen Raum die Beurteilung von Schülerleistungen in einem Schulfach in der Regel numerisch durch Zensuren in Form von Ziffernnoten oder auch Notenpunkten für Einzelleistungen, die unter Beachtung bestimmter Vorgaben zu einer Gesamtnote im Schuljahreszeugnis verrechnet werden. Neben der in den Vordergrund rückenden Aufgabe der Förderdiagnostik und individuellen Lernunterstützung bleibt diese Art der Leistungsbeurteilung sicher bis in absehbare Zeit einer der zentralen Aufgabenbereiche der Lehrperson.

In Deutschland fallen die Regelungen zur Leistungsbeurteilung in das Schulrecht und damit in die Kompetenz der Bundesländer. Die Richtlinien und Grundlagen der Notenbildung werden jeweils länderspezifisch durch Richtlinien, Verordnungen und Gesetze geregelt³. Der Gesetzgeber lässt im Rahmen der geltenden formalen Rechtsvorschriften dem Lehrer einen relativ weiten „pädagogischen Beurteilungsspielraum“ für die Notengebung⁴, verlangt aber

² Nicht so in den skandinavischen Ländern, die bis in höhere Jahrgangsstufen auf Noten und Ziffernzeugnisse verzichten

³ z.B. in Baden-Württemberg die *Verordnung des Kultusministeriums über die Notenbildung* (Notenbildungsverordnung, NVO) vom 5. Mai 1983 mit der letzten berücksichtigten Änderung der §§ 3 und 9 durch Artikel 9 Verordnung vom 11. November 2009 (GBl. S. 693, 709), in Bayern durch schulartbezogene Verordnungen auf der Grundlage des Artikel 52 des bayerisches Gesetzes über das Erziehungs- und Unterrichtswesen (BayEUG), in Hessen durch vom Kultusministerium erlassene Rechtsverordnung zur Gestaltung des Schulverhältnisses

⁴ „Die Bildung der Note in einem Unterrichtsfach ist eine pädagogisch-fachliche Gesamtwertung der vom Schüler im Beurteilungszeitraum erbrachten Leistungen.“ (NVO, § 7 (2))

von ihm, dass er die formalen Verfahrensregeln der Benotung einhält, dass er den gegebenen Beurteilungsspielraum verantwortlich im Sinne seiner erzieherischen Aufgaben nutzt, dass er die Chancengerechtigkeit für jede Schülerin und jeden Schüler wahrt und dass er die Regularien seiner Notenbildung im allgemeinen und die Entscheidungsgründe im Einzelfall gegenüber den Schülerinnen und Schülern und Erziehungsberechtigten transparent macht⁵. Nichtsdestoweniger ist die Leistungsbeurteilung in Form von Ziffernnoten oder Notenpunkten einer der häufigsten Konfliktanlässe zwischen Schülern, Erziehungsberechtigten und Lehrern und vielfach Gegenstand juristischer Auseinandersetzungen auf dem Verwaltungsrechtsweg, wobei überprüft wird, ob die Beurteilung formal vorschriftsgemäß zustande gekommen ist.

Der vorliegende Beitrag befasst sich nicht mit der juristischen Ebene der Leistungsbeurteilung, sondern mit einigen alltäglichen praktischen, aber auch grundsätzlichen Fragen der Notenbildung und Notenberechnung. Er bezieht sich schwerpunktmäßig auf die in Form von Ziffern- oder Punktenoten bewerteten Leistungen entsprechend der im deutschen Schulwesen üblichen Metriken und unterzieht einige Aspekte der Notenbildung und Notenberechnung einer detaillierten Betrachtung. Er geht von der Beobachtung aus, dass Lehrpersonen im Rahmen geltender Vorgaben der Notenbildungsverordnungen und der Beschlüsse und Empfehlungen aus Fachbereichskonferenzen verschiedene Verfahren der Notenbildung und –berechnung einsetzen, die sich oftmals der kritischen Analyse und Hinterfragung, letztlich auch der überzeugenden Legitimierbarkeit entziehen. Zwar stellt § 7, Abs. 2, der baden-württembergischen NVO fest, dass die Note "eine pädagogisch fachliche Gesamtwertung der vom Schüler im Beurteilungszeitraum erbrachten Leistungen" ist, und nicht etwa „das Ergebnis einer rein arithmetischen Rechnung“ (Schreiben des Kultusministeriums Baden Württemberg vom 05.07.1996), womit dem Lehrer die Möglichkeit gegeben sein soll, im Einzelfall auch einmal eine herausragende Leistung oder eben auch eine Fehlleistung eines Schüler aus pädagogischen Gründen normabweichend zu gewichten oder besondere Umstände bei der Leistungserbringung zu berücksichtigen; dennoch ist der nach klaren Verfahrensregeln berechnete Notenstand die unabdingbare Grundlage der Leistungsbeurteilung.

Auch wenn also die ganzheitliche Würdigung der von einem Schüler erbrachten Leistungen als eine Leitperspektive der Leistungsbeurteilung gesehen wird, so darf eine solche „holistische“ Sichtweise nicht die Problematik intransparenter, inkonsistenter und damit mit Blick auf die Chancengleichheit auch ungerechter Verfahren der Notenbildung verschleiern. Sehr wohl nämlich lassen sich die Verfahren der Integration von Einzelleistungsbewertungen in Form von Ziffernnoten oder Notenpunkten rechnerisch erfassen und rekonstruieren. Bei dieser rechnerischen Integration konsistent zu verfahren, ibs. bei allen Schülerinnen und Schülern einer Gruppe dasselbe Verfahren anzuwenden, ist ein Gebot der Gerechtigkeit und Transparenz; ausgehend von dem Resultat einer solchen Berechnung kann dann bei Vorliegen gewichtiger Gründe im Einzelfall, dem Prinzip der Billigkeit folgend, auch einmal der pädagogischen Freiraum für eine von der errechneten Note abweichende Beurteilung genutzt werden. Für die Noten in Abschlusszeugnissen müssen allerdings die Maximen der Verfahrensgerechtigkeit und Chancengleichheit Vorrang haben.

⁵ Als „Transparenzerlass“ bezeichnet man in Baden-Württemberg die in § 7 der NVO geregelte Informationspflicht des Lehrers in Verbindung mit dem Auskunftsrecht des Schülers.

Im vorliegenden Papier sollen einige Verfahrensweisen der Notenbildung im Hinblick auf ihr jeweiliges Begründungsprinzip genauer analysiert und ggf. in ihrer Problematik diskutiert werden. Besonderes Augenmerk kommt im vorliegenden Papier den Verfahren der Verrechnung von Einzelnoten zu Gesamtnoten zu. Ich fasse diese Verfahren unter dem Begriff „*Modelle der Notenbildung*“ zusammen. Sobald es um die Zusammenfassung einzelner bewerteter Leistungsaspekte eines Schülers zu einer Ziffernote geht oder um die Zusammenfassung einzelner Ziffernoten zu einer Gesamtnote, müssen bestimmte, festgelegte Verfahren der Integration der Teilleistungen zu einer Gesamtleistung eingesetzt werden; es muss dazu ihs. geklärt werden, in welchem „Verhältnis“ einzelne Leistungsaspekte oder Einzelnoten zueinander stehen, und es müssen in jedem Falle bestimmte Festlegungen getroffen werden, in welchen Gewichtungen Einzelaspekte in eine Gesamtbewertung einfließen. Diese Verfahren sollen hier konzentriert betrachtet werden.

Ausgeklammert bleiben in der vorliegenden Betrachtung einerseits grundlegende Fragen und Probleme der schulischen Leistungsbewertung wie etwa die fragwürdige Güte der Leistungsbeurteilung hinsichtlich der klassischen Testgütekriterien, das Problem subjektiver Fehlerquellen der Notengebung, die Problematik der Funktionsüberfrachtung von Noten und die Problematik angemessener Kombinationen von Bezugsnormen bei der Leistungsbeurteilung, andererseits messtheoretische und fachmethodische Fragen der Leistungsmessung.

1.2 Beispiele zur Praxis der Notenberechnung

Um die Problematik der Notenbildung, wie sie in diesem Beitrag fokussiert wird, zu umreißen, seien vier Beispiele vorangestellt.

1.2.1 Beispiel

- Schüler Markus hat als Hausaufgabe einen schriftlichen Aufsatz geschrieben. Die Lehrerin nimmt den Aufsatz mit nachhause, korrigiert und benotet ihn und fügt diese Note den schriftlichen Noten hinzu. Ihr Kollege argumentiert dagegen, Hausaufgaben seien grundsätzlich als mündliche Leistungen zu werten.
- Die Schülerinnen und Schüler einer Klasse sollten zu einem kontroversen Thema als Hausaufgabe jeweils eine Stellungnahme verfassen. Schülerin Ute trägt ihre Stellungnahme vor der Klasse vor und erhält dafür eine mündliche Note. Der Lehrer sammelt dann einige weitere Hefte mit den schriftlich verfassten Stellungnahmen ein, korrigiert und wertet diese Leistungen als schriftliche Noten.

Das Beispiel verweist auf Unklarheiten und Unsicherheiten hinsichtlich der Zuordnung von Leistungen zu bestimmten Leistungskategorien, die im Zusammenhang mit unterschiedlichen Gewichtungen der einzelnen Leistungskategorien folgenreich sein können.

1.2.2 Beispiel

Ein Schüler hat im Fach Englisch zwei schriftliche Klassenarbeiten mit den Noten 4 und 5 geschrieben; in drei Vokabeltests (schriftlichen Wiederholungsarbeiten) hatte er die Noten 1, 2 und 1,5 erhalten. Lehrer A und Lehrer B folgen den Empfehlungen der Fachkonferenz, Klassenarbeiten doppelt, schriftliche Wiederholungsarbeiten einfach zu gewichten.

- Lehrer A berechnet die schriftliche Note als doppelt gewichteten Durchschnitt der Klassenarbeiten und einfach gewichteten Durchschnitt der Vokabeltests folgendermaßen:

Schriftliche Durchschnittsnote =

$$(2 \cdot 4,5 + 1 \cdot 1,5) / 3 = 10,5 / 3 = 3,5$$

- Lehrer B berechnet die Summe der schriftlichen Noten, wobei er die schriftlichen Klassenarbeiten jeweils doppelt, die Vokabeltests einfach gewichtet, und dividiert durch die fiktive Anzahl $2 + 2 + 1 + 1 + 1 = 7$ schriftlicher Arbeiten:

Schriftliche Durchschnittsnote =

$$(2 \cdot 4 + 2 \cdot 5 + 1 \cdot 1 + 1 \cdot 2 + 1 \cdot 1,5) / 7 = 22,5 / 7 = 3,2^6$$

Der Unterschied ist eklatant und wird umso drastischer, je stärker sich die Noten der schriftlichen Klassenarbeiten von denen der Vokabeltests unterscheiden. Ich greife das Problem auf unter dem Stichwort *Kategorien-Gewichtung versus Anzahl-Gewichtung*.

1.2.3 Beispiel

- Schüler Markus bringt immer wieder in den Unterricht sehr anspruchsvolle Beiträge ein, die das erworbene Wissen reflektiert und ideenreich auf neue Sachverhalte anwenden und oft eigene Lösungsansätze für gestellte Probleme beinhalten.
- Schülerin Lisa hält sich bei der Erarbeitung neuer Sachverhalte sehr zurück; allerdings lässt sie keine Lücke erkennen, wenn es um die Wiedergabe gelernten Wissens oder die Anwendung von Kenntnissen auf bekannte Aufgabenstellungen geht.

Sind die mündlichen Beiträge dieser beiden Schüler zum Unterricht gleich zu bewerten? Das Beispiel verweist auf die Problematik, dass Leistungen mit Berücksichtigung unterschiedlicher Anforderungsniveaus zu bewerten sind.

1.2.4 Beispiel

- Für die Schülerin Anna hat der Lehrer die Einzelnoten 2,5; 1,0; 1,0; 1,5 für mündliche Fachleistungen notiert; da die Schülerin kein Referat gehalten hat, errechnet der Lehrer die mündliche Leistung nur aufgrund der vorliegenden mündlichen Einzelnoten zu 1,5.
- Für den Schüler Ben hat die Lehrerin die Einzelnoten 2,5; 1,0; 1,0; 1,5 für mündliche Fachleistungen notiert; der Schüler hat sich jedoch der Einbringung eines Referats verweigert. Die Lehrerin erachtet die Übernahme eines Referates für eine gute mündliche Note als notwendig und notiert für die nicht erbrachte Leistung eine 6,0. Da sie Referate grundsätzlich doppelt so stark wie eine sonstige mündliche Fachnote gewichtet, führt ihre Berechnung zu der mündlichen Gesamtleistung 3,0.
- Für die Schülerin Clara hat die Lehrerin nur eine mündliche Einzelnote 5,0 notiert, für zwei gut gelungene Referate allerdings die Noten 2,5 und 1,5. Die Lehrerin gewichtet im

⁶ Bei der praktischen Notenberechnung bricht man nach der 1., spätestens nach der 2. Dezimalen ab; ein Gleichheitszeichen wird dennoch gesetzt.

mündlichen Leistungsbereich die Referatenote doppelt und kommt in ihrer Berechnung auf die mündliche Gesamtnote 3,0.

- Der Schüler David steht gegen Schuljahresende mündlich auf der Note 2,0; ein Referat ist ihm allerdings misslungen und liefert nur die Note 4,5. Soll der Lehrer diese Note als Pflichtleistung werten oder als Wahlleistung deklarieren und in diesem Fall aus der Berechnung der mündlichen Gesamtleistung ausschließen?
- Aus Zeitgründen kann der Schüler Elias sein Referat nicht mehr im Unterricht vortragen. Die Lehrerin bittet ihn um eine schriftliche Ausarbeitung. Die Bewertung dieser Ausarbeitung ordnet sie den schriftlichen Leistungen zu und gewichtet sie dort halb so stark wie eine schriftliche Klassenarbeit.

In den jeweiligen Einzelfällen lässt sich die Notengebung durchaus legitimieren. Ist die Notengebung aber über diese Einzelfälle hinweg konsistent und gerecht? Problematisch ist sicherlich die inkonsistente Zuordnung gleichartiger Leistungen zu verschiedenen Leistungskategorien. Wir werden aber sehen, dass auch hier Gewichtungen eine Rolle spielen und dass unterschiedliche Notenbildungsprinzipien angewandt wurden.

Die Beispiele verweisen auf Fragen der Zuordnung von Leistungen zu bestimmten Kategorien, auf die Gewichtung von Einzelnoten bei der Berechnung einer Gesamtnote sowie auf unterschiedliche Strategien der Einrechnung von Einzelnoten in die Gesamtnote. Diese drei Problemfelder sind Gegenstand der Betrachtungen in dem vorliegenden Papier.

1.3 Fragestellung und Gliederung des Beitrags

Die Beurteilung schulischer Leistungen ist ein komplexes, induktives Verfahren in mehreren Schritten. Leistungsverhalten muss in arrangierten Anforderungssituationen auf eine geeignete Weise festgestellt, gemessen und bewertet werden. I.d.R. werden dann die resultierenden Leistungsdaten, je nach Art der Leistung, definierten Kategorien zugeordnet, die kategorienspezifischen Leistungsdaten sodann zu einer Kategorienbewertung, schließlich die einzelnen Kategorienbewertungen ihrerseits zu einer Bewertung der Gesamtleistung zusammengefasst und (wenn nicht durch eine Verbalbeurteilung) durch eine Ziffernnote oder Notenpunktzahl dargestellt. Die Auseinandersetzung mit dem Problembereich schulischer Leistungsbeurteilung durch Zensuren beinhaltet somit (mindestens) die Klärung der folgenden Fragen:

- (1) Welches Leistungsverhalten wird unter welchen Anforderungen in welcher Weise festgestellt und gemessen? In welcher Weise werden die festgestellten bzw. gemessenen Leistungen durch Punktzahlen oder Ziffernnoten bewertet?
- (2) Welche Einzelleistungen werden welchen Leistungskategorien zugeordnet? Oder: In welcher Weise werden „gleichartige“ Leistungen zu Kategorien zusammengefasst?
- (3) In welcher Weise werden die durch Ziffernnoten oder Punktzahlen bewerteten Leistungen einer Kategorie zu einer Kategorienbewertung zusammengefasst? Oder: Welche Modelle der Notenbildung sind zu unterscheiden?

Die Frage (1) ist eine fachliche bzw. methodische Frage und berührt wesentlich die messtheoretische Frage nach der Abbildung eines empirischen (beobachtetes Leistungsverhalten)

in ein numerisches Relativ (Punkte- oder Ziffernnotenskala) und der Skalenqualität. Die Beschäftigung mit dieser Frage bleibt im vorliegenden Papier, mit Ausnahme eines kurzen Exkurses in 1.4, ausgeklammert. Wir gehen stets von bewerteten Einzelleistungen aus, d.h. die bewertete Einzelleistung ist unsere unterste Analyseebene.

In Abschnitt 2 versuche ich zunächst, einige begriffliche Klärungen herbeizuführen sowie der Frage (2) unter der Perspektive nachzugehen, wie, nach welchen Leitprinzipien die Bildung von Leistungskategorien erfolgt bzw. erfolgen sollte. Der Abschnitt 3 ist Modellen der Notenintegration unter Verwendung von Gewichtungen gewidmet und schneidet damit einige grundlegende Fragen der Notenbildung im Sinne der Frage (3) an; mitlaufend ist dabei nochmals die Frage (2) der Kategorienzuordnung von Leistungen. Die Frage (3) wird weitergeführt in Abschnitt 4 mit einer grundsätzlichen Betrachtung zu Strategien der Notenbildung im Zusammenhang mit Wahl-, Wahlpflicht- und Pflichtleistungen. Abschnitt 5 wird verdeutlichen, dass die bis dahin angestellten Betrachtungen noch einmal komplexer werden, wenn wir die Modelle der Notenintegration nicht nur auf – unterstellter Weise – voneinander unabhängige, „dimensionierte“ Leistungen beziehen, sondern auf voneinander abhängige Leistungen und Leistungsbereiche anwenden, im einfachsten Fall auf taxonomisch geordnete Leistungsbereiche.

1.4 Exkurs zur Problematik der Skalenqualität

Nicht Gegenstand dieses Beitrages sind die hinreichend bekannten und häufig (allerdings weitgehend folgenlos) diskutierten Probleme um die sog. „Skalendignität“ von Schulnoten, denen ich hier nur einen kurzen Exkurs widme.

In der Praxis der Notenbildung werden von Zahlenfolgen, die Leistungsbewertungen eines bestimmten Schülers in Form von Ziffern- oder Punktenoten repräsentieren, routinemäßig arithmetische Mittelwerte im Sinne der „durchschnittlichen Leistung“ ausgerechnet; ebenso werden nach Leistungsüberprüfungen ganzer Schülergruppen, z.B. einer Schulklasse, Durchschnitte ausgerechnet. Spätestens hier macht uns der Messtheoretiker auf eine zentrale Problematik aufmerksam: Unsere üblichen Schulnoten haben im Allgemeinen gar nicht die „Messqualität“, um sie in dieser Weise miteinander verrechnen zu können. Um messtheoretisch zulässige Durchschnitte zu berechnen, müssten Schulnoten die gemessenen Leistungen auf einer Skala repräsentieren, die zumindest *Intervallskalenniveau* aufweist, einem Messniveau wie dies etwa bei Temperaturmessungen mit einer Celsius- oder Fahrenheit-Skala erreicht wird.

Die üblichen Notenskalen – etwa die von der Note 1 bis zur Note 6 oder auch die 15 Punkte-Skala – weisen jedoch in aller Regel nicht die Qualität einer „Intervallskala“ auf, sondern erreichen zumeist lediglich die Qualität einer Messung auf *Ordinalskalenniveau*, ein Messniveau wie es etwa bei der Beliebtheitseinschätzung von Politikern oder einer Wetterbeurteilung nach „heiß“, „warm“, „lau“, „kühl“ etc. erreicht wird. Ein solches Messniveau erlaubt höchstens qualitative Größenvergleiche: Die Note „4“ ist besser als die Note „5“; die Note „1“ ist besser als die Note „2“; aber die jeweiligen Differenzen sind quantitativ nicht vergleichbar; die für Intervallskalen geforderte Äquidistanz der Skaleneinheiten ist in den meisten Fällen nicht über die gesamte Skala hinweg gewährleistet. Plausibel wird diese Feststellung schon

dann, wenn man Differenzen zwischen Notenpaaren auf der Grundlage der verbalen Notenbeschreibungen zu vergleichen trachtet – beispielsweise den Unterschied zwischen „ausreichend“ und „mangelhaft“ mit dem Unterschied zwischen „sehr gut“ und „gut“. Das messtheoretisch für Schulnoten zu verwendende Maß der zentralen Tendenz wäre demnach im Falle einer zusammenfassenden Charakterisierung z.B. der Median, nicht aber das arithmetische Mittel (vgl. z.B. Eid et al., 2015, 139 f.).

In Ausnahmefällen mögen Noten sich durchaus der Messqualität einer Intervallskala annähern. Beispiele hierfür wären etwa eine Note, die proportional einer bestimmten Anzahl gelöster elementarer und voneinander unabhängiger Aufgaben in einem Rechentest, einer bestimmten Anzahl richtig angekreuzter Lösungen in einem Multiple-Choice-Test oder einer bestimmten Fehlerzahl in einem Rechtschreibe-Test zugeordnet ist⁷. Wenn hier die Anzahlen gelöster Aufgaben oder die Fehleranzahlen die Leistungen auf Intervallskalenniveau erfassen, dann bleibt diese Skalenqualität durch eine lineare Zuordnung von Noten auf einer Notenskala erhalten. Zudem macht z.B. Bortz (1979, 30) im Zusammenhang mit sozialwissenschaftlichen Messskalen darauf aufmerksam, dass die Einstufung als Ordinalskalen auch zu Unterbewertungen führen kann, weil zwar in den Extrembereichen von Skalen die Äquidistanz häufig nicht gewährleistet sei, während jedoch in den mittleren Bereichen durchaus Intervallskalengüte vorliegen könne, so dass solche Skalen zwar nicht durchgängig Intervallskalengüte aufweisen, aber dennoch ein höheres Messniveau als reine Ordinalskalen. Dieser Hinweis mag durchaus auch für bestimmte Notenskalen zutreffen; freilich geht eine näherungsweise Intervallskalengüte bereits dann wieder verloren, wenn der Punkte-Noten-Schlüssel, welcher einer leistungsnahen Bewertung mit Verrechnungspunkten Noten zuordnet, nicht-linear gestaltet ist, wobei diese spezifischen Umrechnungen dann häufig auch pädagogisch begründet werden (siehe z.B. Lehrerfreund, 2017). Dass dann Notendurchschnitte, manchmal aus unterschiedlichsten Leistungsbereichen, berechnet und häufig zu folgenreichen Vergleichszwecken herangezogen werden, wird unter Verweis auf pragmatische Argumente gerechtfertigt.

Die soeben beschriebene messtheoretische Problematik liegt auf der untersten Ebene der Abbildung beobachteter Leistungen und Leistungsaspekte auf eine Bewertungsskala. Diese Problematik wird für die Betrachtungen im vorliegenden Papier nur am Rande bedeutsam.

2 Leistung und Notenbildung

2.1 Einige Begrifflichkeiten

In der vorliegenden Arbeit werden (ähnlich Jürgens & Lissmann, 2015, 69) unterschieden:

- die *Leistungsfeststellung* im Sinne der Beobachtung und Beschreibung eines Leistungsverhaltens in einer gegebenen Anforderungssituation; z.B. die Beobachtung kommunikativen Verhaltens in einer Gruppensituation mit Hilfe eines Beobachtungsbogens;

⁷ Voraussetzung ist hier natürlich, dass der Test kriterien- und nicht normenorientiert angelegt ist.

- die *Leistungsmessung*, in welcher der festgestellten (beobachteten) Leistung unter Beachtung von Gütekriterien eine Zahl zugeordnet wird; z.B. die Anzahl richtiger Ankreuzungen in einem Multiple-Choice-Test, die Anzahl von Rechtschreibfehlern bei einem Rechtschreibediktat, die Anzahl richtig gelöster Aufgaben in einem Rechentest;
- Die *Leistungsbewertung*, die einer festgestellten oder gemessenen Leistung unter Zugrundelegung eines Bewertungsmaßstabes eine Bewertung auf einer Notenskala zuweist; z.B. die Bewertung eines Rechtschreibediktats mit 6 Fehlern auf der Notenskala von 1 bis 6 mit der Note 2,5;
- die *Leistungsbeurteilung*, die häufig synonym mit der Leistungsbewertung verwendet wird (so bei Jürgens & Lissmann, 2015), bevorzugt aber den Prozess der Zusammenfassung mehrerer bewerteter Einzelleistungen zu einer Gesamtbewertung bezeichnet, wobei dieser Prozess mehrere Stufen umfassen kann.

Während auf der Grundlage einzelner Leistungsfeststellungen auch ein Wortgutachten aufsetzen kann, wird mit dem Begriff der *Leistungsmessung* die Weiche für eine Bewertung oder Beurteilung in Form einer Ziffernnote oder Punktezahl auf einer Notenskala gestellt. Wenn in den weiteren Betrachtungen von „Leistung“ die Rede ist, soll stets die *bewertete Leistung* gemeint sein.

Um den Begriff der *Notenbildung* im deutschen Beurteilungswesen hinreichend scharf in den Blick zu bekommen, müssen wir beachten, dass die Leistungsbeurteilung eines Schülers oder einer Schülerin in einem Schulfach von der Bewertung der einzelnen Leistungen bis hin zu einer Gesamtnote i.d.R. mehrere Stufen durchläuft. Hilfreich sind hier die folgenden begrifflichen Unterscheidungen:

- Die *Gesamtleistung*, bewertet durch eine Endnote in einem Schulfach, integriert alle Leistungen eines Schülers oder einer Schülerin in diesem Schulfach über ein Halbjahr oder ein gesamtes Schuljahr hinweg zu einer Gesamtbeurteilung.
- Die Gesamtleistung setzt sich in aller Regel aus *Teilleistungen* in einzelnen definierten *Leistungsbereichen zusammen* – wie etwa der schriftlichen Leistung, der mündlichen Leistung, der praktischen Leistung⁸ etc., die in einem Schulfach über eine festgelegte Zeitspanne hinweg jeweils als Gesamtnote für den jeweiligen Leistungsbereich gebildet wird. Werden Leistungsbereiche in sich noch weiter unterteilt, dann entstehen Unterbereiche, die in Ermangelung einer gängigen Nomenklatur als *Leistungssparten* bezeichnet seien – so etwa innerhalb des Leistungsbereichs der schriftlichen Leistungen die Sparte für Klassenarbeiten, die Sparte für schriftliche Wiederholungsarbeiten, die Sparte für den als Hausarbeit geschriebenen Aufsatz oder im mündlichen Leistungsbereich die Sparte für die Fachnoten, die Sparte für Referate oder vielleicht eine Sparte für die „Förderung des Unterrichtsprozesses“. Leistungsbereiche und Leistungssparten seien im Folgenden generalisierend als *Kategorien (Leistungskategorien)* bezeichnet.
- Der Begriff *Einzelleistung* soll sich in dieser Arbeit auf die unterste, elementare Stufe des Beurteilungsprozesses beziehen, also auf die Ebene der Leistungsmessung, auf der

⁸ wobei hier in verschiedenen Bundesländern unterschiedliche Einteilungen vorgenommen werden

festgestellte Leistungen erstmals durch eine Note oder Punktzahl bewertet und dokumentiert werden. Dies kann die einzelne Klassenarbeit in der Sparte der schriftlichen Klassenarbeiten sein, die einzelne mündliche Fachleistung, das Referat, der einzelne benotete Gesamteindruck über eine gewisse Zeitspanne hinweg (die „Eindrucksnote“) etc; es kann die einzelne Bewertung in einer separaten Leistungssparte für das Lesen, Schreiben, Hören, die Konversation oder Präsentation sein, die zusammen den Bereich der mündlichen Leistungen bilden; es könnte aber auch die Einzelleistung sein, die sich aus diesen Leistungsaspekten oder –dimensionen konstituiert und in eine einzelne Bewertung als mündliche Note einfließt. Grundsätzlich sind Ausdifferenzierungen hin zu sehr spezifischen Aspekten immer möglich.⁹

Der Begriff der *Notenbildung* steht in der vorliegenden Arbeit für den gesamten Prozess der Bewertung einzelner Leistungen und der stufenweisen Zusammenführung von Teilleistungen bis hin zu einer Gesamtbewertung in Form einer Ziffernote oder Punktezahl. Abbildung 1 veranschaulicht beispielhaft eine mögliche Grundstruktur eines Notenbildungsprozesses.

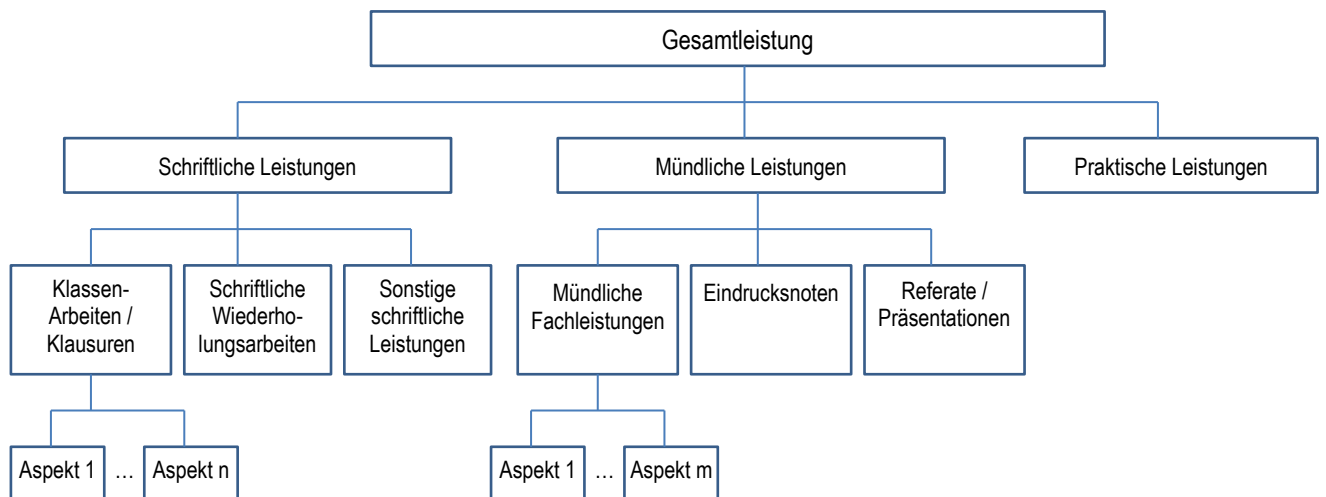


Abbildung 1 Beispiel für die Grundstruktur des Notenbildungsprozesses. Erläuterung im Text.

Tabelle 1 zeigt als Beispiel die Dokumentation einzelner Leistungen in verschiedenen Leistungskategorien in der Notenübersicht einer Lehrperson. In den Zeilen sind jeweils die bewerteten Einzelleistungen und die Bewertungen der Teilleistungen in den Leistungskategorien aufgeführt; der Notenbildungsprozess endet mit der Gesamtbewertung und der Endnote am Ende einer Zeile.

⁹ Wir treffen hier auf die grundsätzlichen Probleme, die mit jeder Operationalisierung von Konstrukten verbunden sind.

Tabelle 1 Beispiel für die Notentabelle einer Lehrperson. In den Zeilen sind pro Schüler/in die (möglicherweise unterschiedlichen Anzahlen von) Noten für die Einzelleistungen und die Teilleistungen in verschiedenen Kategorien (jeweils symbolisiert durch das Symbol „ \emptyset “) dokumentiert; der Notenbildungsprozess endet mit der Gesamtbewertung und der Endnote.

Schüler	Schriftliche Leistungen				Praktische Leistungen		Mündliche Leistungen				Gesamtbewertung	Endnote
	Klassenarbeiten	Tests	GFS	\emptyset		\emptyset	Interview (Abfragen)	Eindrücke	Referate	\emptyset		
1	A_1, \dots, A_n	T_1, \dots, T_m	G	\bar{S}	P_1, \dots, P_p	\bar{P}	F_1, \dots, F_r	E_1, \dots, E_s	R_1, \dots, R_t	\bar{M}	\bar{N}	
...												
...												
...												
n												

Abbildung 1 und Tabelle 1 spiegeln die im deutschen Beurteilungssystem häufig zugrunde gelegten Leistungskategorien wieder. Die Notenbildung kann aber freilich auch auf einer anderen Struktur von Leistungskategorien beruhen. Abbildung 2 zeigt ein Beispiel einer Grundstruktur, die auf die Differenzierung unterschiedlicher Anforderungsniveaus setzt und jedes Anforderungsniveau durch verschiedene schriftliche, mündliche und praktische Leistungen belegt. In Abschnitt 5 werden Probleme der Notenbildung im Zusammenhang mit solchen „Taxonomien“ aufgegriffen.

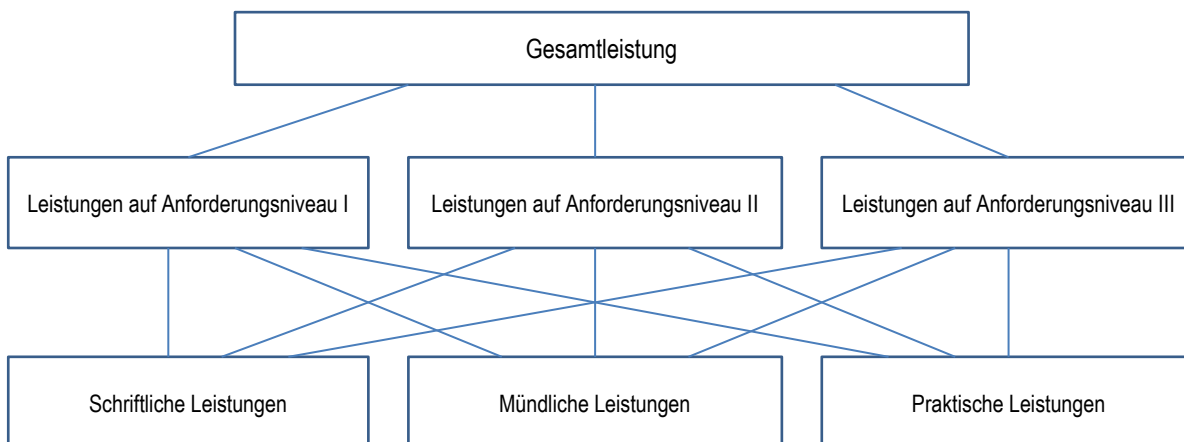


Abbildung 2 Beispiel für die Grundstruktur eines Notenbildungsprozesses. Erläuterung im Text.

Notenbildung kann sich auch in der Hinsicht unterschiedlich gestalten, dass bestimmte Einzel- oder Teilleistungen als „Wahl-“, „Wahlpflicht-“ oder als „Pflichtleistungen“ deklariert und mit unterschiedlicher Verbindlichkeit in die Gesamtnote einfließen. Eine systematisierende Übersicht verschiedener „Logikmodelle“ der Notenbildung entwickle ich in Abschnitt 4.

Eine genauere Beschreibung des Notenbildungsprozesses erfolgt in Abschnitt 3. Im Folgenden sollen zunächst einige Betrachtungen den üblichen Leistungskategorien und der Zuordnungsproblematik von Einzelleistungen gewidmet werden.

2.2 Leistungskategorien

Der Prozess der Notenbildung beruht auf einem gestuften Prozess der zusammenfassenden Bewertung einzelner Leistungskategorien. Welche Leistungen werden in einer Kategorie zusammengefasst? Man könnte z.B. folgende Prinzipien der Kategorienbildung unterscheiden:

- (1) Nach *Art der Prüfungssituation* kann man z.B. unterscheiden in mündliche Leistungen, schriftliche Leistungen, praktische Leistungen o.ä., die jeweils in unterschiedlichen Prüfungssituationen erbracht werden;
- (2) nach dem *Gegenstandsbezug der Leistung* in verschiedenen (fachspezifischen) Inhaltsbereichen könnte man z.B. in Mathematik Leistungen im Bereich der Bruchrechnung, Leistungen in der elementaren Geometrie o.ä. definieren, in denen dann jeweils sowohl schriftlich wie mündlich erbrachte Leistungen zusammengefasst wären;
- (3) nach dem *Anforderungsniveau* könnte man Leistungen taxonomisch zuordnen, wobei dann vielleicht bei den schriftlichen Klassenarbeiten oder in der mündlichen Prüfungssituation jeweils ein Teil einem bestimmten Anforderungsniveau zugeordnet wäre.

Denkbar sind viele verschiedene Kategorisierungen, und häufig oder gar in der Regel werden verschiedene Kategorisierungsprinzipien kombiniert. Dennoch erscheint folgendes Leitprinzip der Kategorienbildung als unabdingbar

Leitprinzip der Interpretierbarkeit einer Leistungskategorie:

Leistungen, die in einer Kategorie zusammengefasst und schließlich auch zusammenfassend bewertet werden, sollen unter Bezug auf das Kategorienbildungsprinzip interpretierbar sein.

Wenn eine Lehrperson ihre Beurteilung gegenüber der Schülerin oder Eltern transparent und verständlich machen soll, dann muss sie auf Gleichartigkeiten der in einer Kategorie zusammengefassten Leistungen rekurrieren können.

Im deutschen Beurteilungsweisen werden die Leistungen eines Schülers nach dem „Typ“ der Leistung oder nach der Art der Leistungserbringung verschiedenen Kategorien zugeordnet. In aller Regel wird hier unterschieden zwischen *schriftlichen Leistungen*, *mündlichen Leistungen* und *sonstigen Leistungen*; in der Notenverordnung von Baden-Württemberg werden neben den schriftlichen und mündlichen Leistungen die *praktischen Leistungen* als eigene Kategorie genannt; in Nordrhein-Westfalen werden die mündlichen Leistungen zusammen mit anderen Leistungen der Kategorie „sonstige Leistungen“ zugordnet. Vom Kerngedanken her soll die Kategorisierung unterschiedlicher Leistungsarten eine bessere Interpretierbarkeit und Legitimierbarkeit der aus heterogenen Leistungsarten zusammengefügt Gesamtleistung ermöglichen, in dem man diese in einzelne Kategorien unterteilt, die jeweils aufgrund größerer Homogenität eine bessere Interpretierbarkeit gewährleisten.

- Mündliche Leistungen* sind die bewerteten Beiträge, die ein Schüler im Zusammenhang mit Unterricht mündlich zeigt. Sie können im Rahmen verschiedenster Prüfsituationen (Abfragen, Beobachtung während eines Wiederholungsgesprächs in der Klasse, Referat

oder Präsentation etc.) oder auch als hochintegrierte Zusammenfassung von Beobachtungen über einen längeren Zeitraum (Eindrucksnote) erfasst werden.

- *Schriftliche Leistungen* werden in Form von schriftlichen Klassenarbeiten, Klausuren, schriftlichen Wiederholungsarbeiten etc. erhoben.
- *Praktische Leistungen* (z.B. NVO Baden-Württemberg) werden gemäß Lehrplan z.B. erhoben in Sport sowie in den musischen Fächern, ggf. in den Fächern Hauswirtschaft/Textiles Werken, in den Profulfächern Mensch und Umwelt, Natur und Technik der Realschule sowie in Bildungsgängen des beruflichen Schulwesens; sie sind ähnlich wie mündliche Leistungen zu bewerten.

Die Unterscheidung solcher Kategorien wie *schriftliche*, *mündliche* und *praktische Leistungen* und die Zuordnung von Einzelleistungen zu solchen Kategorien ist eine weichenstellende Grundlage der Notenbildung, denn diese Leistungskategorien gehen in aller Regel mit einer bestimmten Gewichtung in die Gesamtnote ein – in mathematisch-naturwissenschaftlichen Fächern häufig im Verhältnis schriftlich zu mündlich wie 2:1, in neusprachlichen Fächern meistens im Verhältnis 1:1 etc. Dabei beruhen solche Kategorisierungen vornehmlich auf der Tradition des Beurteilungswesens im öffentlichen deutschen Schulsystem und sind pädagogisch-psychologisch häufig nicht sachlogisch, sondern nur durch die unterschiedlichen Arten der Prüfungssituation zu rechtfertigen.

Welchen Informationsgehalt hat die Aussage, eine Schülerin sei im mündlichen Leistungsbereich „gut“, dagegen in den schriftlichen Arbeiten „eher befriedigend“? Dass ein Schüler eine Fremdsprache kommunikativ zu nutzen weiß, erweist sich sicherlich am validesten in einer mündlichen Gesprächssituation. Dass aber ein Schüler in Mathematik eine bestimmte Aufgabe der Differenzialrechnung lösen kann, sollte er in einer schriftlichen Klassenarbeit ebenso gut beweisen können wie in einer mündlichen Prüfungssituation, in welcher an der Tafel oder einem äquivalenten Medium eine Aufgabe (dialogisch) bearbeitet wird. Freilich ist hier die Art der Prüfungssituation durchaus einflussreich: In einer schriftlichen Klassenarbeit, wo im Regelfall die zu lösenden Aufgaben feststehen, ist ein Schüler weitestgehend auf sich allein gestellt, während die typische mündliche Leistungsprüfung in einer Gesprächssituation stattfindet, wo der Lehrer vor dem Hintergrund jeweils vorgängiger Informationen evtl. eine leichtere oder schwierigere weitere Frage stellt oder Aufgabe auswählt und wo er durch Nachfragen, Hilfestellungen, frühzeitige Korrekturen etc. flexibel auf die dargebotene Leistung des Schülers eingehen kann.

Den vorangehenden Überlegungen entnehmen wir die folgenden Fragen, die uns in diesem Papier vornehmlich beschäftigen:

- (a) Wie werden Leistungen jeweils bestimmten Kategorien zugeordnet? (Zuordnungsproblematik)
- (b) Wie wirken sich verschiedene Gewichtungen und Gewichtungstypen von Leistungen innerhalb der Kategorien und der Kategorien selbst auf die Notenbildung aus?

Der folgende Unterabschnitt 2.3 befasst sich in Form einiger besonderer Beispielfälle zunächst mit der Zuordnungsproblematik in Frage (a).

2.3 Die Zuordnung von „gleichwertigen“ und „Mischleistungen“

Gehen wir im Folgenden nun von der „Realität“ im deutschen Beurteilungswesen aus und nehmen die Einteilung in schriftliche, mündliche, praktische Leistungen als gegeben hin.

Aber auch unter dieser scheinbar klaren Leistungskategorisierung ist die Zuordnung von Leistungen zu den Leistungskategorien ein häufiger Gegenstand von Fragen und Verwirrungen. So wird z.B. immer wieder die Frage diskutiert, ob Leistungen in Verbindung mit Hausaufgaben den mündlichen oder den schriftlichen Leistungen zuzurechnen ist, oder ob die Beurteilung eines mündlich vorgetragenen, aber schriftlich ausgearbeiteten Referats als eine schriftliche oder mündliche Note erfolgt. In solchen Unsicherheiten schwingt oft unterschwellig die Frage mit, in welcher Gewichtung diese Einzelnoten in die Fachnote eingehen soll, was z.B. dann bedeutsam ist, wenn – wie in etlichen Schulfächern festgelegt – schriftliche Leistungen doppelt so stark gewichtet werden wie mündliche Leistungen.

Die Vermischung von Leistungsart, Prüfungssituation und Gewichtung kommt sehr deutlich in Festlegungen der baden-württembergischen NVO über die „Ersetzbarkeit“ von Klassenarbeiten durch andere Leistungen zum Ausdruck. Z.B. heißt es in § 9 (2) der NVO:

In den Fächern Technik sowie Mensch und Umwelt der Realschule können jeweils bis zu zwei Klassenarbeiten durch fachpraktische Arbeiten ... ersetzt werden; das gleiche gilt im Fach Naturwissenschaft und Technik des Gymnasiums mit der Maßgabe, dass eine Klassenarbeit ersetzt werden kann.

Oder in § 9 (6), der sich auf die beruflichen Schulen bezieht:

Von den nach Absatz 3 vorgesehenen Klassenarbeiten kann ... eine der Klassenarbeiten durch eine gleichwertige Feststellung von Leistungen der Schüler der Klasse ersetzt werden. Diese Leistungsfeststellung bezieht sich insbesondere auf schriftliche Hausarbeiten, Jahresarbeiten, Projekt, darunter auch experimentelle Arbeiten im naturwissenschaftlichen Bereich, Freiarbeit, Referate, mündliche, gegebenenfalls auch außerhalb der stundenplanmäßigen Unterrichtszeit terminierte Prüfungen oder andere Präsentationen. ... In den Hauptschulen, Realschulen und Gymnasien gelten die Sätze 1 bis 3 entsprechend mit der Maßgabe, dass die ... gleichwertigen Leistungen die Zahl der vorgeschriebenen Klassenarbeiten unberührt lässt.

Mit der „Ersetzbarkeit“ bzw. „Gleichwertigkeit“ von Leistungen wird die Gewichtung, die Bedeutsamkeit der Einzelleistung im Vergleich mit den schriftlichen Klassenarbeiten, festgelegt, während die Art der Leistung eine völlig andere als eine schriftliche Leistung sein kann. Für die Notenbildung wird mit diesen Festlegungen allerdings eine gewisse Verwirrung gestiftet.

- Soll eine mit einer schriftlichen Klassenarbeit „gleichwertige Leistung“ auch dann, wenn sie eine völlig andere als eine schriftliche Leistung beinhaltet, in die Kategorie der schriftlichen Klassenarbeiten eingeordnet und zusammen mit diesen zu einer schriftlichen Gesamtnote verrechnet werden? Rechnerisch ist der Gleichwertigkeit damit entsprochen; aber die Interpretation einer Leistungsnote für den Leistungsbereich „schriftliche Leistungen“ wird damit höchst fragwürdig.
- Wird eine solche Leistung aber in eine andere Kategorie eingeordnet, dann ergibt sich einerseits die Frage, welche anderen Leistungen in eine solche „Restkategorie“ einsortiert werden und wie angesichts der zu erwartenden Inhomogenität der darin vertretenen Leistungsarten eine sinnvolle Interpretation einer Gesamtnote für diese Kategorie lauten kann (siehe das o.g. *Leitprinzip der Interpretierbarkeit*), und die zweite Frage, wie die vorgeschriebene Gleichwertigkeit dieser Leistung mit einer schriftlichen Klassenarbeit rechnerisch sichergestellt werden kann.

Damit geraten also zwei Gesichtspunkte der Notenbildung tendenziell in einen Widerspruch, nämlich die Vorgabe bezüglich Gewichtung bzw. Bedeutsamkeit und die Anforderung, eine gebildete Gesamtnote für einen Leistungsbereich auf der Grundlage annähernd homogener Einzelleistungen interpretieren zu können.

Ich vertrete die folgende Position:

- Bei der Zuordnung von Leistungen zu Leistungskategorien sollte zugunsten der Homogenität der Einzelleistungen in den Kategorien und damit der Interpretierbarkeit soweit wie möglich die *Art der Leistung bzw. Leistungserbringung* maßgeblich sein.

Beispiel: Hausaufgaben

Die Leistung für die Hausaufgabe kann den schriftlichen oder den mündlichen Leistungen oder auch einem dritten Leistungsbereich (etwa den praktischen Leistungen) zugeordnet werden, je nachdem welche Art der Leistung erbracht wird. Dominiert in der Hausaufgabe eine schriftliche Leistung, wie etwa die Anfertigung eines Aufsatzes, so gehört die hierfür zu erteilende Note zum Leistungsbereich der schriftlichen Leistungen; steht in der Hausaufgabe eine mündliche Leistung im Vordergrund, z.B. das Erlernen und Vortragen eines Gedichtes, dann ist diese Leistung als mündliche Leistung zu werten; wird als Hausaufgabe eine praktische Leistung erbracht – z.B. die Erstellung eines Posters oder die Vorbereitung eines Rollenspiels –, dann gehört die zu erteilende Note in den Bereich der praktischen Leistungen. In welcher Gewichtung diese Noten in die jeweilige Teilleistungsnote eingehen, ist eine ganz andere Frage.

- Liegt eine rechtliche Vorgabe zur Gleichwertigkeit einer Leistung – und damit zu einer gleichen Gewichtung – im Vergleich mit anderen Leistungen vor, dann und nur dann und nur gerechtfertigt aus Gründen der rechnerischen Stimmigkeit und gewissermaßen als „Notlösung“ sollte die Leistung der entsprechenden Kategorie zugerechnet werden.

Beispiel: GFS (= gleichwertige Feststellung von Schülerleistungen)

Schreibt die NVO die Gleichwertigkeit einer Leistung mit einer schriftlichen Klassenarbeit vor, dann sollte die entsprechende Leistungsnote in die Kategorie der schriftlichen Leistungen eingeordnet werden. Es wäre für die Notenberechnung in der Praxis zu schwierig, eine rechnerische Gleichgewichtung einer GFS mit einer schriftlichen Klassenarbeit sicherzustellen, wenn die Note für die GFS in eine andere Kategorie, z.B. in eine Kategorie „Sonstige Leistung“ eingeordnet würde. In der Interpretation der Gesamtnote für die schriftlichen Leistungen muss dann ggf. die Heterogenität der Leistungen berücksichtigt werden. Allerdings könnten innerhalb dieser Kategorie zugunsten einer besseren Interpretierbarkeit durchaus wieder Unterkategorien gebildet werden können.

Diese Position der Vorrangigkeit rechtlich vorgegebener Gewichtungen einzelner Leistungen rechtfertigt letztlich auch die folgenden ausführlichen Betrachtungen zum Problem der Gewichtungen bei der Notenbildung im gegenwärtigen Schulsystem.

3 Das Modell der Notenintegration

Der in 2.1 beschriebene Prozess der Notenbildung geht von benoteten, nach Kategorien sortierten Einzelleistungen aus und berechnet daraus jeweils eine Benotung der Kategorie, fasst dann wiederum Kategoriennoten bewertend zusammen bis hin zu einer die Gesamtleistung bewertenden Gesamtnote. Beispielsweise wird in dem in Abbildung 1 dargestellten Notenbildungsprozess der Kategorie der Klassenarbeitsnoten eine Durchschnittsnote der Klassenarbeitsnoten zugeordnet, gleichfalls den Kategorien der schriftlichen Wiederholungsarbeiten und den sonstigen schriftlichen Leistungen. Schließlich fasst man alle schriftlichen Leistungsnoten in eine Kategorie und berechnet die Note für die schriftlichen Leistungen. Sinngemäß verfährt man mit den mündlichen Leistungsnoten. Schließlich werden schriftliche, mündliche, ggf. Leistungsnoten aus anderen Kategorien zu einer Gesamtnote zusammengefasst.

Der Prozess der Notenbildung muss nun noch durch Spezifizierung der Operationen ergänzt werden, mit denen die Leistungen und Teilleistungen in den einzelnen Kategorien jeweils zusammenfassend bewertet werden. Dazu bedarf es zum einen einer Skala zur Leistungsbewertung, zum anderen eines definierten Verfahrens der zusammenfassenden Bewertung einer Menge von Leistungsdaten. Die folgenden beiden Unterabschnitte befassen sich zunächst mit dem Konzept der „Noten-“ bzw. „Notenpunkte-Skala“ und dem Grundmodell der Notenintegration; in den nachfolgenden Unterabschnitten werden dann Verfahren der Mittelwertbildung und damit zusammenhängender Probleme näher betrachtet.

3.1 Noten- und Notenpunkte-Skalen

Wird eine Leistung durch Ziffernnoten oder Notenpunkte bewertet, dann liegt eine *Noten-* oder eine *Notenpunkte-Skala* zugrunde. Die Notenpunkte-Skala ist nicht zu verwechseln mit der Verteilung von Verrechnungspunkten aus einem Punktepool für einzelne Leistungsaspekte.

Was unter eine Noten- oder Notenpunkte-Skala zu verstehen ist, charakterisiert der folgende Kasten.

3.1.1 Noten- und Punkte-Skala:

Eine *Notenskala* bzw. eine *Notenpunkte-Skala* ist ein abgeschlossenes Intervall $Z := [z_1, z_2] = \{x \in Q \mid z_1 \leq x \leq z_2\}$ als zusammenhängende Teilmenge der (linear geordneten) Menge Q der rationalen Zahlen¹⁰, wobei die Intervallgrenzen z_1, z_2 i.d.R. ganze Zahlen sind und eineindeutig der besten bzw. der schlechtesten Leistung zugeordnet sind.

In der praktischen Anwendung für die Bewertung von Einzelleistungen werden auf der Notenskala häufig bestimmte Werte hervorgehoben, die eine Schrittfolge von 1, 0,5, 0,25 oder auch andere Unterteilungen bewirken¹¹; die Notenskala kann dann auch als endliche Teilmenge $\{z_1, \dots, z_2\}$ der rationalen Zahlen geschrieben werden.

Die Standard-Notenskala in Deutschland ist für die Unter- und Mittelstufe i.d.R. die Notenskala $[1, 6]$, die für die Bewertung von Einzelleistungen häufig mit feineren Unterteilungen in der Form $\{1; 1,5; 2; 2,5; 3; 3,5; 4; 4,5; 5; 5,5; 6\}$, sogar mit Viertelsnoten-Unterteilung durch ein „+“ oder „-“ an einer ganzzahligen Note verwendet wird. In der gymnasialen Oberstufe wird i.d.R. die 15-Punkte-Skala als Menge $\{0, \dots, 15\}$ mit 0 als der schlechtesten, 15 als der besten Note verwendet. In anderen Ländern findet man die unterschiedlichsten Skalen.

¹⁰ Diese Festlegung einer Noten- bzw. Notenpunkte-Skala als ein Intervall in der Menge der rationalen Zahlen ermöglicht, die Ergebnisse der Notenintegration auf der gleichen Skala abzubilden, so lange wir im Bereich der Grundrechenarten bleiben. Würde man bei der Notenintegration z.B. auf das geometrische Mittel zurückgreifen, dann müsste man die Menge der reellen Zahlen zugrunde legen.

¹¹ Bei der Leistungsbewertung an Universitäten findet man häufig die Skala $\{1,0; 1,3; 1,7; 2,0; 2,3; 2,7; \dots; 5,0\}$.

So lange es nur um unterschiedliche Skalen der Leistungsbewertung und nicht um eine Umrechnung von Verrechnungspunkten in Notenpunkte oder Noten geht, folgen wir der üblichen Unterstellung, dass die Skalen Intervallskalenniveau aufweisen, somit durch lineare Transformationen (und ggf. Rundungen) ineinander überführt werden können; wenn nämlich den auf den Skalen abgebildeten Leistungsdaten Intervallskalenniveau zugesprochen werden soll, dann sollte diese Präsomtion zumindest auf der Ebene von Skalentransformationen nicht unterlaufen werden. Wir unterstellen also für die folgenden Betrachtungen:

3.1.2 Unterstellung der Intervallskalennqualität der Noten- bzw. Notenpunkte-Skala:

Die auf der Noten- bzw. Notenpunkte-Skala abgebildeten Leistungen sind mit Intervallskalennqualität gemessen; oder einfach: Die der Notenbildung zugrunde gelegte Noten- oder Notenpunkte-Skala besitzt Intervallskalennqualität.

Sind Z und Z' zwei Notenskalen, so sind die Skalen durch lineare Transformationen ineinander überführbar.

Die Intervallskalennqualität einer Skala bleibt erhalten, wenn sie durch eine lineare Transformation auf eine andere Intervallskala abgebildet wird.

Beispiel:

Eine lineare Transformation T einer 15-Punkte-Skala (mit der besten Note 15 Punkte und der schlechtesten Note 0 Punkte) in die übliche Notenskala von 1 bis 6 ist gegeben durch

$$T : \begin{cases} \{0, 1, 2, \dots, 15\} \rightarrow \{1, 2, \dots, 6\} \\ P \rightarrow T(P) := -\frac{1}{3}P + 6 \end{cases} ,$$

wobei durch Rundung oder durch Abbruchregeln nach einer Dezimalen die Ganzzahligkeit des Bildes erreicht wird; praktisch werden dann allerdings meistens auch feinere Unterteilungen der 15- Punkte- bzw. der Notenskala von 1 bis 6 berücksichtigt¹².

Alle in diesem Papier dargestellten Überlegungen beziehen sich auf Noten, die auf einer solchen Notenskala mit unterstellter Intervallskalennqualität verortet sind; sie können auf alle Notenskalen übertragen werden, die sich durch lineare Transformationen aufeinander abbilden lassen. In unseren Beispielen wird jeweils die Notenskala von 1 bis 6 zugrunde gelegt.

¹² Die in Baden-Württemberg verwendete Transformation der 15-Punkte-Skala in die Notenskala von 1 bis 6 mit + und – ist allerdings im Bereich zwischen 5 und 6 nicht linear.

3.2 Grundmodell der Notenintegration

Vorneweg soll in diesem Unterabschnitt zunächst das *Modell der Notenintegration* näher charakterisiert werden. Es schärft das in 2.1 beschriebene Konzept der Notenbildung. Die nachfolgenden Unterabschnitte werden das der Notenintegration zugrunde liegende Verfahren der Mittelwertbildung genauer im Hinblick auf seine Anwendung bei der schulischen Leistungsbewertung analysieren.

3.2.1 Modell der Notenintegration:

Gegeben sei eine (intervallskalierte) Noten- oder Punkteskala $Z := [z_1, z_2]$.

- (1) Ist $X := \{x_1, x_2, \dots, x_m\}$ die Menge der bewerteten Einzelleistungen¹³ einer Person p aus einer bestimmten Kategorie (z.B. der schriftlichen Klassenarbeiten), wobei alle Einzelleistungen auf derselben Skala Z bewertet wurden, also $x_i \in Z$, für alle $i = 1, \dots, m$, dann wird der Menge X nach einem vorgegebenen Berechnungsverfahren eine Bewertung $\bar{x} \in Z$ (eine Bewertung der Kategorienleistung) zugewiesen.
- (2) Ist $Y := \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ eine Menge bewerteter Kategorienleistungen der Person p , wobei $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n \in Z$, dann wird der Menge Y nach einem vorgegebenen Berechnungsverfahren eine Kategorienbewertung $\bar{y} \in Z$ zugewiesen.
- (3) Auf der letzten Stufe steht die Gesamtbewertung als Note oder Punktzahl aus Z .

In der Notengebungspraxis erfolgt die Zuweisung der Bewertung für eine Kategorie in der Regel durch Berechnung des *arithmetischen Mittels*, evtl. eines gewichteten arithmetischen Mittels. Diese Berechnungen werden uns in den folgenden Abschnitten beschäftigen.

Beispiel:

Das in 3.2.1 induktiv charakterisierte Modell der Notenbildung beschreibt z.B. den folgenden Notenbildungsprozess in einem Schulfach auf einer Notenskala von 1 bis 6:

Es seien

$A := \{A_1, A_2, A_3, A_4\}$ die Leistungssparte der schriftlichen Klassenarbeitsnoten mit Bewertung \bar{A} ,

$T := \{T_1, T_2, T_3\}$ die Leistungssparte der schriftlichen Wiederholungsarbeiten mit Bewertung \bar{T} ;

¹³ Im Anschluss an die Unterscheidung der Notenbildung auf den beiden Ebenen der Einzelleistungen und der Kategorienleistungen von 2.1.1 wird auch hier unterschieden zwischen der Bewertung von Einzelleistungen und der Kategorienbewertung. Die damit verbundene Problematik wird in den folgenden Abschnitten diskutiert.

dann wird der Kategorie $S := \{\bar{A}, \bar{T}\}$ der schriftlichen Leistungen die Bewertung \bar{S} zugeordnet.

Es seien

$F := \{F_1, F_2, F_3, F_4, F_5\}$ die Menge mündlicher Fachleistungen einer bestimmten Leistungssparte mit der Bewertung \bar{F} ,

$R := \{R_1, R_2\}$ die Menge mündlicher Fachleistungen einer zweiten Leistungssparte (z.B. von Referaten) mit Bewertung \bar{R} ;

dann wird der Kategorie $M := \{\bar{F}, \bar{R}\}$ der mündlichen Leistungen die Bewertung \bar{M} zugeordnet.

Es sei weiterhin

$P := \{P_1, P_2, P_3\}$ die Kategorie der praktischen Leistungen mit Bewertung \bar{P} .

Dann wird der Kategorie $N = \{\bar{S}, \bar{M}, \bar{P}\}$ der bewerteten schriftlichen, der mündlichen und der praktischen Leistung im betrachteten Schulfach eine Bewertung der Gesamtleistung, d.h. eine Gesamtnote \bar{N} zugeordnet. Der Prozess der Notenbildung gemäß 3.1 ließe sich fortsetzen, indem man aus den Gesamtnoten der einzelnen Schulfächer eine Gesamtnote für das Zeugnis errechnet.

Ein Alternativmodell zu diesem Modell der Notenintegration ist das Modell eines „*Punktepools*“, bei dem eine maximal erreichbare Punktzahl für eine Gesamtleistung („*Punktepool*“) auf verschiedene Einzelleistungen¹⁴ oder Leistungsbereiche aufgeteilt wird, so dass sich die Punkte der Einzelleistungen oder der einzelnen Leistungsbereichen zur Gesamtpunktzahl aufaddieren, die dann ihrerseits auf eine Notenskala abgebildet wird; diese alternative Form der Leistungsbewertung wird in diesem Papier nicht näher betrachtet.

3.3 Gewichtung bei der Notenintegration

Ein wesentliches Merkmal bei Modellen der Notenintegration sind Gewichtungen. *Gewichtungen* werden im Zusammenhang mit der Notenbildung als *Bedeutsamkeitsgewichtungen* und als *Anzahlgewichtungen* verwendet.

- *Bedeutsamkeitsgewichtungen* im Bereich der Notenbildung werden vorgenommen, um die Wichtigkeit, Bedeutsamkeit oder Zuverlässigkeit von Einzelnoten in der Ermittlung einer Gesamtnote oder von einzelnen Aspekten für eine Einzelnote zu bewerten. Wichtige, bedeutsame oder auch zuverlässige Leistungsaspekte erhalten mit einer höheren Gewichtung einen größeren Einfluss auf die resultierende Note – so etwa die schriftliche Klassenarbeit im Verhältnis zu einem Vokabeltest. Bedeutsamkeitsgewichtungen treten

¹⁴ Auf unterster Ebene entspricht dieses Modell z.B. dem Bewertungsmodell aus der Abiturprüfung im Fach Mathematik, wo sich die auf einzelne Aufgaben verteilten Verrechnungspunkte zu maximal 60 Punkten aufaddieren (die dann ihrerseits durch einen genau definierten Schlüssel in Notenpunkte umgerechnet werden); auf höherer Ebene könnte jedoch auch eine Gesamtbewertung aus einer Punkteverteilung auf einzelne Leistungsbereiche resultieren.

bei der Notenbildung aber auch als Kategoriengewichtung auf, etwa dann, wenn die Durchschnittsnote der schriftlichen Leistungen und die Durchschnittsnote der mündlichen Leistungen mit unterschiedlichen Gewichtungen in die Gesamtnote eingehen. Unterschiedliche Gewichtungen können das Resultat einer Berechnung erheblich beeinflussen. Die Angemessenheit der Gewichtungsfaktoren ist stets von der Sache her zu legitimieren und nicht Gegenstand des Berechnungsverfahrens.

- *Anzahlgewichtungen* spielen vor allem für die rechnerische Vereinfachung eine Rolle, nämlich dann, wenn eine Anzahl gleicher Datenwerte zu einer Kategorie zusammengefasst werden; in diesem Fall wird die Kategorie mit der Anzahl der enthaltenen Daten gewichtet. Anzahlgewichtungen können im Bereich der Notenbildung z.B. genutzt werden, um widerzuspiegeln, dass dem schriftlichen Durchschnitt der Klassenarbeiten eine bestimmte Anzahl von Klassenarbeiten zugrunde liegt oder dass eine bestimmte Note (z.B. eine mündliche Note für einen bestimmten Leistungsaspekt) aufgrund mehrfacher Messungen oder Beobachtungen zustande gekommen ist. Tendenziell geht dann die Anzahlgewichtung in eine Bedeutsamkeitsgewichtung der Kategorie über, die sich durch die Genauigkeit bzw. Verlässlichkeit aufgrund mehrfacher Messungen begründen lässt.

Im Bereich der Notenbildung werden Gewichtungen vor allem im Zusammenhang mit der Bestimmung von Durchschnittsnoten auf der Ebene von Einzelleistungen sowie von Gesamtnoten auf der Grundlage verschiedener Kategorien vorgenommen. Dabei sind verschiedene Typen gewichteter Durchschnittsbildungen strikt zu unterscheiden – wie dies Beispiel 1.2.2 zeigt. Nach der kurzen Erinnerung an die Berechnung gewichteter Mittelwerte beschreibe ich diese beiden Gewichtungstypen bei der Notenberechnung als „Gewichtung *innerhalb* einer Kategorie“ bzw. als „Einzelleistungsgewichtung“ und als „Kategoriengewichtung“. Entscheidender Unterschied ist, ob Anzahlen von Einzelleistungen in die Gewichtung eingehen oder nicht. Anschließend werden die beiden Gewichtungstypen vergleichend diskutiert und Risiken von Verzerrungen dargestellt.

3.4 Mittelwert und Einzelleistungsgewichtung

Liegen in einem Leistungsbereich n gleichartige, auf derselben Notenskala $Z := [z_1, z_1]$ abgebildete Leistungsbewertungen eines Schülers in Form von Ziffernnoten vor, wie etwa vier gleichgewichtig zu wertende Noten für schriftliche Klassenarbeiten, dann wird deren Durchschnitt als arithmetisches Mittel der n Einzelnoten berechnet, wobei alle (in Zahlen ausgedrückte) Noten aufsummiert und durch die Anzahl der Noten dividiert werden.¹⁵

Allgemein erfolgt die Berechnung des *arithmetischen Mittelwertes* (kurz: *arithmetisches Mittel*) \bar{x} von n (intervall-skalierten) Daten x_1, x_2, \dots, x_n mit der Formel

¹⁵ Das arithmetische Mittel wird i.a. auch berechnet, um die durchschnittliche Leistung aller Schüler einer Gruppe in einer Leistungsüberprüfungssituation, also etwa den Durchschnitt einer schriftlichen Klassenarbeit, zu bestimmen. In unserer Betrachtung geht es allerdings nur um die „längsschnittliche“ Betrachtung der Leistungen eines einzelnen Schülers.

$$(1) \quad \bar{x} = \frac{\sum_1^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Die Formel besagt, dass die Summe der n Daten x_1, x_2, \dots, x_n (symbolisiert durch das Zeichen $\sum_1^n x_i$) durch die Anzahl n der Daten zu dividieren ist. Leicht einsichtig ist dabei, dass für Noten auf ein und derselben Skala Z gilt:

$$x_1, x_2, \dots, x_n \in Z = [z_1, z_2] \Leftrightarrow z_1 \leq \bar{x} = \frac{\sum_1^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \leq z_2;$$

d.h. das arithmetische Mittel von Noten auf einer Skala liegt wieder auf dieser Skala und kann also als Note auf dieser Skala verstanden werden, wobei ggf. Rundungen auf eine bestimmte Anzahl von Kommastellen vorgenommen werden. Die Anwendung der Formel (1) auf eine Durchschnittsnotenberechnung, beispielsweise die Berechnung des Durchschnittes von vier schriftlichen Klassenarbeitsnoten A_1, A_2, A_3, A_4 , ist klar.

Liegen aus k verschiedenen Kategorien X_1, X_2, \dots, X_k für jede Kategorie X_i das jeweilige Mittel \bar{x}_i aufgrund von s_i gleich gewichteten Daten (auf derselben Intervallskala) vor, so liefert eine leichte Variante zu Formel (1) die folgende Formel¹⁶

$$\bar{x} = \frac{\sum_{i=1}^k s_i \bar{x}_i}{\sum_1^k s_i} = \frac{s_1 \bar{x}_1 + \dots + s_k \bar{x}_k}{s_1 + \dots + s_k}$$

3.4.1 Bemerkung

Ein Spezialfall dieser Formel hilft in der in der Praxis öfter auftretenden Situation, dass nach bereits erfolgter Berechnung eines Durchschnittes \bar{x} von s Daten noch ein weiterer Wert (eine zusätzliche Note) x_{s+1} mit gleicher Gewichtung wie die anderen Werte hinzukommt. Der neue Durchschnitt \bar{x}_{neu} errechnet sich dann gemäß

$$\bar{x}_{neu} = \frac{s\bar{x} + x_{s+1}}{s + 1}$$

Nun werden im schriftlichen Leistungsbereich neben schriftlichen Klassenarbeiten häufig auch schriftliche Wiederholungsarbeiten (häufig und auch im Folgenden meist kurz als „Tests“ bezeichnet) geschrieben. Nach der NVO von Baden-Württemberg § 8 sollen Klassenarbeiten Aufschluss geben „über Unterrichtserfolg und Kenntnisstand einer Klasse und einzelner Schüler ...“ und „daher in der Regel nach Phasen der Erarbeitung, Vertiefung, Übung und Anwendung angesetzt“ werden; schriftliche Wiederholungsarbeiten sollen dagegen Aufschluss geben „über den erreichten Unterrichtserfolg der unmittelbar vorangegangenen Unterrichtsstunden“ oder auch „als Nachweis dafür dienen, mit welchem Erfolg die Hausaufgaben bewältigt wurden“. Deshalb sind für die Anfertigung einer schriftlichen Wie-

¹⁶ die als Spezialfall der späteren Formel (5) zu verstehen ist

derholungsarbeit „in der Regel bis zu 20 Minuten vorzusehen“. Diese Vorgaben rechtfertigen eine höhere Gewichtung der Bedeutsamkeit von Klassenarbeiten gegenüber schriftlichen Wiederholungsarbeiten aufgrund ihres größeren Aussagewertes bezüglich der schriftlichen Leistungsfähigkeit eines Schülers. Nichtsdestoweniger gehören beide Leistungsarten in den schriftlichen Leistungsbereich. Noch etwas unübersichtlicher wird die Berechnung, wenn eine Bestimmung lautet, dass eine „Gleichwertige Feststellung einer Schülerleistung“ (GFS) wie eine schriftliche Klassenarbeit zu werten ist.

Wie werden die Klassenarbeitsleistungen nun mit den Leistungen in den schriftlichen Wiederholungsarbeiten und der GFS unter Berücksichtigung der Bedeutsamkeitsgewichtung zur schriftlichen Leistung verrechnet?

Die Berechnung des *gewichteten Mittelwertes* \bar{x}_g von n einzelnen (intervall-skalierten) Daten x_1, x_2, \dots, x_n mit den Gewichten $\gamma_1, \gamma_2, \dots, \gamma_n$ und der Gewichtssumme $\sum_1^n \gamma_i = \gamma_1 + \gamma_2 + \dots + \gamma_n = G$ erfolgt mit der Formel

$$(2) \quad \bar{x}_g = \frac{\sum_1^n \gamma_i x_i}{\sum_1^n \gamma_i} = \frac{\gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_n x_n}{\gamma_1 + \gamma_2 + \dots + \gamma_n} = \frac{\gamma_1}{G} x_1 + \frac{\gamma_2}{G} x_2 + \dots + \frac{\gamma_n}{G} x_n$$

Leicht ersichtlich ergibt sich Formel (1) als Spezialfall aus Formel (2), wenn alle $\gamma_i = 1$ sind und damit $\sum_1^n \gamma_i = G = n$.

Die Faktoren $\frac{\gamma_1}{G}, \frac{\gamma_2}{G}, \dots, \frac{\gamma_n}{G}$ auf der rechten Seite der Formel können als *normierte Gewichtungsfaktoren* verstanden werden, deren Summe gleich 1 ist. Sind allgemein $\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_n$ Gewichtungsfaktoren, die auf 1 normiert sind, gilt also $\sum_1^n \hat{\gamma}_i = 1$, so vereinfacht sich die Formel (2) (aber nicht unbedingt die praktische Rechnung) zu

$$(3) \quad \bar{x}_g = \sum_1^n \hat{\gamma}_i x_i$$

Werden die Daten k Kategorien X_1, X_2, \dots, X_k so zugeordnet, dass sich jeweils in einer Kategorie X_i die Daten $x_{i1}, x_{i2}, \dots, x_{is_i}$ mit gleichen Gewichten γ_i befinden, dann schreibt sich die Formel (2) als

$$(4) \quad \bar{x}_g = \frac{\sum_{i=1}^k (\sum_{j=1}^{s_i} \gamma_i x_{ij})}{\sum_1^k \gamma_i s_i} = \frac{\sum_{i=1}^k \gamma_i (\sum_{j=1}^{s_i} x_{ij})}{\sum_1^k \gamma_i s_i} = \frac{\gamma_1 (x_{11} + x_{1s_1}) + \dots + \gamma_k (x_{k1} + \dots + x_{ks_k})}{\gamma_1 s_1 + \dots + \gamma_k s_k}$$

Wurde für jede der Kategorien X_1, X_2, \dots, X_k bereits der jeweilige Durchschnitt $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ ermittelt, so erhält Formel (4) die folgende gleichwertige Gestalt:

$$(5) \quad \bar{x}_g = \frac{\sum_{i=1}^k \gamma_i s_i \bar{x}_i}{\sum_1^k \gamma_i s_i} = \frac{\gamma_1 s_1 \bar{x}_1 + \dots + \gamma_k s_k \bar{x}_k}{\gamma_1 s_1 + \dots + \gamma_k s_k}$$

Hierin stellen die Gewichte γ_i die vorgegebene Bedeutsamkeitsgewichtung der Einzeldaten in der Kategorie X_i (und damit der Kategorie X_i), die Gewichte s_i Anzahlgewichtungen der Kategorie X_i entsprechend der Anzahl der in der Kategorie X_i vorhandenen Einzeldaten dar. Die Gewichtung der einzelnen Kategoriendurchschnittswerte für die berechnete Leistungsnote ist dann eine Kombination aus der Bedeutsamkeitsgewichtung der (Einzeldaten in der) Kategorie und der Anzahlgewichtung entsprechend der Anzahl der in der Kategorie enthaltenen Daten¹⁷.

Formel (5) ist relevant für die Frage, wie etwa die mit einem bestimmten Faktor gewichteten schriftlichen Klassenarbeiten, die mit einem bestimmten Faktor gewichteten Tests und zusätzliche Einzelleistungen mit einem bestimmten Gewicht in einer dritten Kategorie (etwa die GFS) zu einer schriftlichen Note zu verrechnen sind. Das Berechnungsverfahren ist folgendermaßen anzuwenden:

Seien

- A_1, A_2, \dots, A_r die Noten der r schriftlichen Arbeiten mit einem jeweiligen (gleichen) Gewicht γ_A ,
- T_1, T_2, \dots, T_s die Noten in s schriftlichen Wiederholungsarbeiten (Tests) mit jeweiligem (gleichen) Gewicht γ_T
- B_1, B_2, \dots, B_t die t Noten in einer dritten Sparte schriftlicher Arbeiten mit jeweiligem (gleichen) Gewicht γ_B ;

dann berechnet sich die schriftliche Durchschnittsnote \bar{S} entsprechend Formel (4) gemäß

$$(6) \quad \bar{S} = \frac{\sum_1^r \gamma_A A_i + \sum_1^s \gamma_T T_i + \sum_1^t \gamma_B B_i}{\sum_1^r \gamma_A + \sum_1^s \gamma_T + \sum_1^t \gamma_B} = \frac{\gamma_A \sum_1^r A_i + \gamma_T \sum_1^s T_i + \gamma_B \sum_1^t B_i}{\gamma_A r + \gamma_T s + \gamma_B t} =$$

$$= \frac{\gamma_A (A_1 + \dots + A_r) + \gamma_T (T_1 + \dots + T_s) + \gamma_B (B_1 + \dots + B_t)}{\gamma_A r + \gamma_T s + \gamma_B t}$$

oder unter Verwendung des schon berechneten Durchschnittes \bar{A} der Klassenarbeiten und des Durchschnittes \bar{T} der Tests gemäß

$$(7) \quad \bar{S} = \frac{\gamma_A r \bar{A} + \gamma_T s \bar{T} + \gamma_B t \bar{B}}{\gamma_A r + \gamma_T s + \gamma_B t}$$

¹⁷ Man beachte, dass in dieser Formel die Kategorienbildung nicht einer Sachlogik, sondern der arithmetisch bedingten Zusammenfassung gleichgewichteter Einzelleistungen folgt.

3.4.2 Beispiel

Angenommen, wir haben von einem Schüler folgende Notentabelle mit schriftlichen Klassenarbeiten, Tests und einer GFS:

Klassenarbeiten				Durchschnitt	Tests					Durchschnitt	GFS
A_1	A_2	A_3	A_4	\bar{A}	T_1	T_2	T_3	T_4	T_5	\bar{T}	B
2,0	4,5	3,0	2,5	3,0	2,0	3,5	2,5	4,0	1,5	2,7	1,5

Angenommen weiterhin, die schriftlichen Arbeiten A_1, A_2, A_3, A_4 werden mit einem Gewicht von 3, die schriftlichen Wiederholungsarbeiten T_1, T_2, T_3 mit einem Gewicht von 1, die gleichwertige Schülerleistung GFS wie die schriftlichen Arbeiten mit dem Gewicht 3 gewertet. Man errechnet dann gemäß Formel (5) bzw. (6) die Durchschnittsnote \bar{S} gemäß

$$\begin{aligned}\bar{S} &= \frac{3 \cdot (2,0 + 4,5 + 3,0 + 2,5) + 1 \cdot (2,0 + 3,5 + 2,5 + 4,0 + 1,5) + 3 \cdot 1,5}{3 \cdot 4 + 1 \cdot 5 + 3 \cdot 1} \\ &= \frac{3 \cdot 12,0 + 1 \cdot 13,5 + 3 \cdot 1,5}{20} = 2,7\end{aligned}$$

oder auch – unter Verwendung der Durchschnitte $\bar{A} = 3,0$ bzw. $\bar{T} = 2,7$ bzw. $\bar{B} = B = 1,5$

$$\bar{S} = \frac{3 \cdot 4 \cdot 3,0 + 1 \cdot 5 \cdot 2,7 + 3 \cdot 1,5}{20} = 2,7.$$

3.5 Kategoriengewichtung als Bedeutsamkeitsgewichtung

Wenn die Rede davon ist, dass die schriftlichen zu den mündlichen Leistungen im Verhältnis 2 : 1 gewichtet in die Gesamtnote \bar{N} eingehen, dann wird damit eine Bedeutsamkeitsgewichtung vorgenommen: Die schriftliche Leistung hat dann in der gesamten Leistungsbeurteilung ein doppelt so großes Gewicht wie die die mündliche Leistung. Auf diese Kategoriengewichtungen richten sich die nachfolgenden Betrachtungen in diesem Unterabschnitt.

3.5.1 Bemerkung

Um einer häufigen Verwirrung vorzubeugen, sei vermerkt, dass Gewichtungsvorgaben wie die genannten stets *relative Gewichtungen* der Leistungskategorien im Verhältnis zueinander betreffen, nicht die Anteile an der Gesamtleistung. Dies mögen die beiden folgenden Beispiele verdeutlichen.

- (1) Falls neben den schriftlichen und mündlichen Leistungen keine weiteren Leistungskategorien vorgesehen sind, entspricht der Vorgabe 2 : 1 in der Gewichtung der schriftlichen zu den mündlichen Leistungen folgendes Berechnungsverfahren: Berechne den Durchschnitt \bar{S} der schriftlichen Einzelleistungen; berechne den Durchschnitt \bar{M} der mündlichen Einzelleistungen; berechne dann die Gesamtnote \bar{N} als Summe des zweifach gewichte-

ten Durchschnittes der schriftlichen Leistungen und des einfach gewichteten Durchschnittes der mündlichen Noten und teile die Summe durch 3, also gemäß Formel (2):

$$\bar{N} = \frac{2 \cdot \bar{S} + 1 \cdot \bar{M}}{3} = \frac{2}{3} \bar{S} + \frac{1}{3} \bar{M}.$$

Die schriftliche Durchschnittsnote geht somit zu zwei Dritteln, die mündliche Durchschnittsnote zu einem Drittel in die Gesamtnote ein. Die Gewichtung der schriftlichen zur mündlichen Leistung im Verhältnis $\frac{2}{3} : \frac{1}{3} = 2 : 1$ ist gewahrt.

- (2) Falls außer den schriftlichen und mündlichen Leistungen weitere Leistungskategorien vorgesehen sind, müssen die relativen Gewichtungen der schriftlichen zu den mündlichen Leistungen erhalten bleiben, jedoch ändert sich der jeweilige Anteil an der Gesamtleistung. Angenommen etwa, neben den im Verhältnis 2 : 1 zu gewichtenden schriftlichen und mündlichen Leistungen seien in die Gesamtbewertung noch praktische Leistungen aus der Kategorie \bar{P} mit gleichem Gewicht wie die mündlichen Leistungen einzurechnen. Die Gesamtnote \bar{N} errechnet sich dann gemäß

$$\bar{N} = \frac{2 \cdot \bar{S} + 1 \cdot \bar{M} + 1 \cdot \bar{P}}{4} = \frac{2}{4} \bar{S} + \frac{1}{4} \bar{M} + \frac{1}{4} \bar{P}$$

Natürlich bleiben die relativen Gewichtungen von schriftlichen zu mündlichen zu praktischen Leistungen im Verhältnis $\frac{2}{4} : \frac{1}{4} : \frac{1}{4} = 2 : 1 : 1$ erhalten, während sich die jeweiligen Anteile an der Gesamtleistung vermindern.

Eine allgemeine Formel für die Errechnung eines gewichteten Gesamtwertes \bar{X}_g aus k Kategorien X_1, X_2, \dots, X_k mit den Durchschnitten $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ und den Gewichten Φ_1, \dots, Φ_k lautet entsprechend Formel (2):

$$(8) \quad \bar{X}_g = \frac{\sum_1^k \Phi_i \bar{x}_i}{\sum_1^k \Phi_i} = \frac{\Phi_1 \bar{x}_1 + \dots + \Phi_k \bar{x}_k}{\Phi_1 + \dots + \Phi_k}$$

oder, mit normierten Gewichten, $\hat{\Phi}_1, \dots, \hat{\Phi}_k$,

$$(9) \quad \bar{X}_g = \sum_1^k \hat{\Phi}_i \bar{x}_i$$

3.5.2 Bemerkung

Werden in den Gewichten Φ_1, \dots, Φ_k die Anzahlen s_1, \dots, s_k der Einzeldaten in den k Kategorien X_1, X_2, \dots, X_k und die Bedeutsamkeitsgewichtungen γ_i der Einzeldaten in jeder Kategorie X_i derart berücksichtigt, dass sich die Kategoriengewichtung als $\Phi_i = \gamma_i s_i$ für alle $i = 1, \dots, k$, berechnet, dann geht Formel (8) in Formel (5) über:

$$\bar{x}_g = \frac{\sum_1^k \Phi_i \bar{x}_i}{\sum_1^k \Phi_i} = \frac{\Phi_1 \bar{x}_1 + \dots + \Phi_k \bar{x}_k}{\Phi_1 + \dots + \Phi_k} = \frac{\gamma_1 s_1 \bar{x}_1 + \dots + \gamma_k s_k \bar{x}_k}{\gamma_1 s_1 + \dots + \gamma_k s_k} = \frac{\sum_{i=1}^k \gamma_i s_i \bar{x}_i}{\sum_1^k \gamma_i s_i} = \bar{x}_g.$$

In diesem Sinne kann Formel (8) als die allgemeinere Formel betrachtet werden, die Formel (5) als Spezialfall enthält. Praktisch dürfte dieser Fall der speziellen Gewichtung allerdings irrelevant sein, da Kategoriengewichtungen im Sinne von Bedeutsamkeitsgewichtungen gerade der Kategorie als solcher eine besondere Bedeutsamkeit beimessen, ohne dass die Anzahl der Einzeldaten eine Rolle spielt.

3.5.3 Beispiel

Werden in einem Schulfach die schriftlichen Leistungen eines Schülers mit dem Durchschnitt $\bar{S} = 3,5$ doppelt, die mündlichen Leistungen mit dem Durchschnitt $\bar{M} = 2,5$ einfach und die praktischen Leistungen mit dem Durchschnitt $\bar{P} = 2,5$ ebenfalls einfach gewichtet, so berechnet sich die Gesamtleistung \bar{N} gemäß Formel (8) zu

$$\bar{N} = \frac{2 \cdot \bar{S} + 1 \cdot \bar{M} + 1 \cdot \bar{P}}{2 + 1 + 1} = \frac{2 \cdot 3,5 + 1 \cdot 2,5 + 1 \cdot 2,5}{2 + 1 + 1} = 3,0$$

oder, mit normierten Gewichten, gemäß Formel (9) zu

$$\bar{N} = \frac{2}{4} \bar{S} + \frac{1}{4} \bar{M} + \frac{1}{4} \bar{P} = \frac{2}{4} \cdot 3,5 + \frac{1}{4} \cdot 2,5 + \frac{1}{4} \cdot 2,5 = 3,0^{18}$$

Die Formel (8) oder ersatzweise (9) ist immer anzuwenden, wenn eine Leistungskategorie – ohne Betrachtung der darin beinhalteten Einzelleistungen – nur mit deren Durchschnitt und einem fest vorgegebenen Gewicht in eine Gesamtnote eingehen soll.

3.6 **Kategoriengewichtung versus Einzelleistungsgewichtung**

Auch in Beispiel 3.4.2, wo es um die Verrechnung der schriftlichen Klassenarbeiten, Tests und der GFS ging, hätte man auf die Idee kommen können, die unterschiedlichen Typen schriftlicher Leistungen in drei verschiedene Sparten innerhalb der schriftlichen Leistungen einzuordnen und unter Anwendung der Formel (8) die errechneten Durchschnitte der drei Sparten jeweils mit einem Bedeutsamkeitsgewicht für die einzelne Sparte zu einer Durchschnittsnote zu verrechnen. Im Beispiel würde dann die Rechnung lauten

¹⁸ Das Rechenbeispiel zeigt, dass die Rechnung mit normierten Gewichten nicht unbedingt bequemer ist.

$$\bar{S} = \frac{3 \cdot \bar{A} + 1 \cdot \bar{T} + 3 \cdot \bar{B}}{3 + 1 + 3} = \frac{3 \cdot 3,0 + 1 \cdot 1,5 + 3 \cdot 1,5}{3 + 1 + 3} = \frac{15,0}{7} = 2,1$$

Der errechnete Durchschnitt der schriftlichen Leistungen „verbessert“ sich bei dieser Berechnung also von 2,5 auf 2,1 (gerundet auf die erste Dezimale). Die Abweichung ist beachtlich und nicht undramatisch. Dem Grund für den Unterschied kommt man schnell auf die Spur: In Beispiel 3.4.2 fallen die schlechteren Noten der schriftlichen Klassenarbeiten stärker ins Gewicht, weil in der verwendeten Formel (5) bzw. (6) die größere Anzahl der schriftlichen Arbeiten berücksichtigt wird. Im eben gerechneten Verfahren spielen dagegen die Anzahlen der Leistungen in den einzelnen Sparten keine Rolle. Diese unterschiedliche Berechnung liegt auch dem Eingangsbeispiel 1.2.2 sowie teilweise 1.2.3 zugrunde.

Gehen wir diesem Unterschied mit einer allgemeinen Betrachtung nach. Unter welchen Umständen spiegelt die Bedeutsamkeitsgewichtung einer Kategorie die Bedeutsamkeitsgewichtung der Einzelleistungen wider?

Nehmen wir aus den k Kategorien X_1, X_2, \dots, X_k zwei beliebige Kategorien X_u und X_v (z.B. schriftliche und mündliche Leistungen) mit den Durchschnittswerten \bar{x}_u, \bar{x}_v und den – der Einfachheit halber als normiert angenommenen – Kategoriengewichten $\hat{\Phi}_u, \hat{\Phi}_v$. Dann gilt für die jeweiligen Kategorien:

$$\hat{\Phi}_u \bar{x}_u = \hat{\Phi}_u \frac{1}{s_u} \sum_{j=1}^{s_u} x_{uj} = \sum_{j=1}^{s_u} \frac{\hat{\Phi}_u}{s_u} x_{uj}$$

$$\hat{\Phi}_v \bar{x}_v = \hat{\Phi}_v \frac{1}{s_v} \sum_{i=1}^{s_v} x_{vi} = \sum_{j=1}^{s_v} \frac{\hat{\Phi}_v}{s_v} x_{vj}$$

wobei jede Einzelleistung in Kategorie X_u das Gewicht $\frac{\hat{\Phi}_u}{s_u}$, jede Einzelleistung in Kategorie X_v das Gewicht $\frac{\hat{\Phi}_v}{s_v}$ erhält¹⁹. Sollen sich nun die Gewichtungen $\hat{\Phi}_u$ und $\hat{\Phi}_v$ der Kategorien X_u und X_v im gleichen Verhältnis auf die Gewichtungen der Einzelleistungen übertragen, so muss gelten

$$\frac{\hat{\Phi}_v}{\hat{\Phi}_u} = \frac{\frac{\hat{\Phi}_v}{s_v}}{\frac{\hat{\Phi}_u}{s_u}} = \frac{\hat{\Phi}_v \cdot s_u}{\hat{\Phi}_u \cdot s_v} = \frac{\hat{\Phi}_v}{\hat{\Phi}_u} \cdot \frac{s_u}{s_v}$$

und damit $s_u = s_v$.

Wir halten fest:

¹⁹ Nimmt man über alle k Kategorien die neuen Gewichte $\frac{\hat{\Phi}_i}{s_i}$ in ihrer jeweiligen Vielfachheit, dann gilt natürlich auch hier $\sum_{i=1}^k (\sum_1^{s_i} \frac{\hat{\Phi}_i}{s_i}) = \sum_{i=1}^k \hat{\Phi}_i = 1$. D.h. $\frac{\hat{\Phi}_i}{s_i}$ sind die neuen normierten Gewichte der Einzelleistungen.

Die relative Gewichtung zweier Kategorien X_u und X_v überträgt sich auf die Gewichtung der Einzelleistungen in diesen Kategorien dann und nur dann, wenn die Anzahlen der Einzelleistungen in den Kategorien X_u und X_v gleich sind.

Sind die Anzahlen $s_i =: s$ in allen Kategorien gleich, so erhalten wir unter Verwendung der Formeln (4) und (5):

$$\begin{aligned} \bar{x}_g &= \frac{\sum_{i=1}^k (\sum_{j=1}^s \gamma_i x_{ij})}{\sum_1^k \gamma_i s_i} = \frac{\gamma_1(x_{11} + x_{1s}) + \dots + \gamma_k(x_{k1} + \dots + x_{ks})}{\gamma_1 s + \dots + \gamma_k s} = \\ &= \frac{\gamma_1 s \bar{x}_1 + \dots + \gamma_k s \bar{x}_k}{\gamma_1 s + \dots + \gamma_k s} = \frac{\gamma_1 \bar{x}_1 + \dots + \gamma_k \bar{x}_k}{\gamma_1 + \dots + \gamma_k} = \frac{\sum_1^k \gamma_i \bar{x}_i}{\sum_1^k \gamma_i}. \end{aligned}$$

Die Einzelleistungsgewichtung ist hier also mit der Kategoriengewichtung entsprechend Formel (8) identisch.

Man kann diese Überlegung noch etwas weiterführen zu der Aussage

$$(10) \quad \frac{\hat{\Phi}_v}{\hat{\Phi}_u} > \frac{\hat{\Phi}_v \cdot s_u}{\hat{\Phi}_u \cdot s_v} \Leftrightarrow \frac{s_u}{s_v} < 1 \Leftrightarrow s_u < s_v,$$

die folgende Aussage beinhaltet:

Mit zunehmender Anzahl der Einzelleistungen in einer Kategorie X_v verringert sich die Bedeutsamkeitsgewichtung der Einzelleistungen in Kategorie X_v gegenüber der Bedeutsamkeitsgewichtung einer anderen Kategorie X_u .

Enthält z.B. die Kategorie X_v doppelt so viele Einzelleistungen wie Kategorie X_u , ist also $s_v = 2s_u$, so ist

$$\frac{\hat{\Phi}_2 \cdot s_1}{\hat{\Phi}_1 \cdot s_2} = \frac{\hat{\Phi}_2 \cdot s_1}{\hat{\Phi}_1 \cdot 2s_1} = \frac{1}{2} \frac{\hat{\Phi}_2}{\hat{\Phi}_1}$$

d.h. die Bedeutsamkeit einer Einzelleistung in Kategorie X_v gegenüber der Kategoriengewichtung selbst beträgt nur die Hälfte.

Im Zusammenhang mit der Notenbildung steckt in diesen Betrachtungen einige Brisanz, weil bei vorgegebener Kategoriengewichtung die Bedeutsamkeit einer Einzelleistung in einer Leistungskategorie von der Anzahl der Einzelleistungen abhängt. Verdeutlichen wir die Problematik an einem Beispiel.

3.6.1 Beispiel

Angenommen im Notenbuch eines Lehrers stehen folgende Einzelnoten:

Schriftliche Noten				Durchschnitt	Mündliche Noten		Durchschnitt	Gesamtnote
A_1	A_2	A_3	A_4	\bar{A}	M_1	M_2	\bar{M}	
4,0	4,5	3,5	4,0	4,0	1,5	2,5	2,0	

Wird entsprechend einer fachbezogenen Vorgabe die Kategorie der schriftlichen Leistungen zur Kategorie der mündlichen Leistungen im Verhältnis 2 : 1 gewichtet, dann errechnet sich die Gesamtnote \bar{N} gemäß Formel (8) zu

$$\bar{N} = \frac{2 \cdot 4,0 + 1 \cdot 2,0}{3} = 3,3.$$

Faktisch ist hier die Bedeutsamkeit einer schriftlichen Einzelnote die gleiche wie die einer mündlichen Einzelnote, da die Kategorie der schriftlichen Noten zwar doppelt gewichtet wird im Vergleich zur Kategorie der mündlichen Noten, dort aber doppelt so viele Noten eingehen wie hier, ohne dass die Anzahlen berücksichtigt würden. Man könnte in diesem Falle alle Leistungen gleich gewichten und mit gleichem Ergebnis wie folgt rechnen:

$$\bar{N} = \frac{4,0 + 4,5 + 3,5 + 4,0 + 1,5 + 2,5}{6} = 3,3.$$

Die in der höheren Gewichtung der schriftlichen Leistungen für die Gesamtnote sich wider-spiegelnde Vorstellung, dass die schriftlichen Leistungen für die Beurteilung der Gesamtleistung bedeutsamer sind, wird durch die faktische Gleichgewichtung der schriftlichen und mündlichen Einzelleistungen konterkariert. Würde man dagegen jede Einzelleistung bei den schriftlichen Noten zweifach gewichten und jede Einzelleistung bei den mündlichen Noten einfach, dann ergäbe die Rechnung laut Formel (5)

$$\bar{N} = \frac{2 \cdot (4,0 + 4,5 + 3,5 + 4,0) + 1 \cdot (1,5 + 2,5)}{2 \cdot 4 + 1 \cdot 2} = \frac{2 \cdot 4 \cdot 4,0 + 1 \cdot 2 \cdot 2,0}{10} = 3,6.$$

Natürlich fallen bei dieser Berechnung die schlechteren Noten der schriftlichen Arbeiten wegen Berücksichtigung ihrer größeren Anzahl stärker ins Gewicht. Faktisch hat man durch diese Berechnung allerdings die Bedeutsamkeit der schriftlichen Gesamtleistung gegenüber der mündlichen Gesamtleistung gesteigert – von 2 : 1 auf das Verhältnis 8 : 2 = 4 : 1.

Wie groß kann die Differenz unter realistischen Voraussetzungen maximal werden? Nehmen wir ein einfaches Extrembeispiel.

3.6.2 Beispiel

Ein Schüler habe vier schriftliche Klassenarbeiten jeweils mit der Note 6 geschrieben, und er habe eine mündliche Einzelnote von 1,0 erhalten. Dann sieht die Notentabelle folgendermaßen aus:

Schriftliche Noten				Durchschnitt	Mündliche Note	Durchschnitt	Gesamtnote
A_1	A_2	A_3	A_4	\bar{A}	M_1	\bar{M}	
6,0	6,0	6,0	6,0	6,0	1,0	1,0	

Nehmen wir nun an, der Lehrer gewichtet die Kategorie der schriftlichen Noten doppelt so stark wie die Kategorie der mündlichen Noten. Dann errechnet sich die Gesamtnote \bar{N} gemäß Formel (8)

$$\bar{N} = \frac{2 \cdot \bar{A} + 1 \cdot \bar{M}}{2 + 1} = \frac{2 \cdot 6,0 + 1 \cdot 1,0}{3} = 4,3$$

Würde man dagegen jede einzelne schriftliche Note zweifach gewichten und die einzelne mündliche Note einfach, dann ergäbe die Rechnung laut Formel (5)

$$\bar{N} = \frac{2 \cdot (6,0 + 6,0 + 6,0 + 6,0) + 1 \cdot 1,0}{2 \cdot 4 + 1 \cdot 1} = \frac{2 \cdot 4 \cdot 6,0 + 1 \cdot 1,0}{9} = \frac{48 + 1}{9} = 5,4.$$

Der Unterschied zwischen den beiden Berechnungsweisen ist dramatisch und könnte zwischen Versetzung und Nichtversetzung entscheiden. Die Bedeutsamkeit der schriftlichen Gesamtleistung gegenüber der mündlichen Gesamtleistung wäre im zweiten Berechnungsverfahren nun freilich auf das Verhältnis 8 : 1 gestiegen, was dem Grundgedanke der Kategoriengewichtung augenscheinlich nicht gerecht würde.

Das aufgezeigte Paradoxon lässt sich wohl nicht restlos auflösen, aber relativieren: Wird in einem Schulfach die schriftliche Teilleistung gegenüber der mündlichen Teilleistung doppelt gewichtet, dann lässt sich möglicherweise auf die höhere Zuverlässigkeit der schriftlichen gegenüber der mündlichen Leistung verweisen. Die schriftlichen Klassenarbeiten in Mathematik beinhalten vermutlich eine höhere Zuverlässigkeit bei Beurteilung des mathematischen Leistungsvermögens von Schülern als mündliche Prüfungen; die höhere Zuverlässigkeit wird aber testtheoretisch auch durch eine gewisse Anzahl von schriftlichen Arbeiten gewährleistet. Testtheoretisch sollte zudem die Zuverlässigkeit einer Teilleistungsbeurteilung umso höher sein, je mehr einzelne Noten in die Teilleistungsnote eingehen. D.h. also: Wenn sich in der angemessenen Kategoriengewichtung der schriftlichen Leistungen die höhere Zuverlässigkeit widerspiegelt, dann tritt dieser Bedeutsamkeitsgewichtung gegenüber die Bedeutsamkeit der einzelnen schriftlichen Klassenarbeit in den Hintergrund. Dennoch lässt sich das zutage getretene Paradoxon wirklich befriedigend nur auflösen mit der Empfehlung, die einzelnen, durch spezifische Vorgaben der Bedeutsamkeitsgewichtung akzentuierten Leistungskategorien möglichst durch die gleiche Zahl von Einzelleistungen zu belegen. Unter diesen Bedingungen führt nach den vorangegangenen Betrachtungen die Berechnung der Gesamtnote über die Einzelleistungsgewichtung zur gleichen Gesamtnote wie die Berechnung über die Kategoriengewichtung, wobei die letztere für diesen Fall das einfachere Berechnungsverfahren darstellt.

3.7 Zuordnungsproblematik

Im Zusammenhang mit der Diskussion um die Kategorien- vs. Einzelleistungsgewichtung muss nun auch noch einmal die Problematik der Zuordnung von Leistungen zu bestimmten Kategorien aufgenommen werden. Die Problematik sei an folgendem einfachen Beispiel dargestellt.

3.7.1 Beispiel

Angenommen, von einem Schüler liegen zwei schriftliche Klassenarbeiten mit den Noten 1,0 und 2,0 und drei mündliche Noten 4,0, 5,0, 3,0 vor; zudem hat er für eine Hausaufgabe – nehmen wir an, es sei ein in der Unterrichtsstunde vorgelesener schriftlicher Aufsatz – die Note 3,0 erhalten. Dann sieht die Notentabelle folgendermaßen aus:

Schriftliche Noten		Mündliche Note			Hausaufgabe	Gesamtnote
A_1	A_2	M_1	M_2	M_3	H	
1,0	2,0	4,0	5,0	3,0	3,0	

Nehmen wir nun an, die Hausaufgabe werde in Fall 1 den schriftlichen Noten zugeordnet und gegenüber den schriftlichen Klassenarbeiten halb so stark gewichtet, in Fall 2 den mündlichen Noten zugeordnet und dort ebenso wie die anderen mündlichen Noten gewichtet. Insgesamt werde die Kategorie der schriftlichen Leistungen doppelt im Vergleich zu den mündlichen Noten gewichtet. Dann errechnet man für die Gesamtnote \bar{N} in

Fall 1:

$$\bar{S} = \frac{2 \cdot 1,0 + 2 \cdot 2,0 + 1 \cdot 3,0}{2 + 2 + 1} = \frac{9,0}{5} = 1,8$$

$$\bar{M} = \frac{4,0 + 5,0 + 3,0}{3} = \frac{12,0}{3} = 4,0$$

$$\bar{N} = \frac{2 \cdot 1,8 + 1 \cdot 4,0}{2 + 1} = \frac{7,6}{3} = 2,5$$

Fall 2:

$$\bar{S} = \frac{1,0 + 2,0}{2} = \frac{3,0}{2} = 1,5$$

$$\bar{M} = \frac{4,0 + 5,0 + 3,0 + 3,0}{4} = \frac{15,0}{4} = 3,7$$

$$\bar{N} = \frac{2 \cdot 1,5 + 1 \cdot 3,7}{2 + 1} = \frac{6,7}{3} = 2,2$$

Das Beispiel verdeutlicht, welchen Einfluss die Zuordnung einer bestimmten Leistung zu einer der beiden Kategorien der schriftlichen bzw. mündlichen Leistungen haben kann. Im Zusammenhang mit den unterschiedlichen Gewichtungen für Einzelleistungen bzw. für Kate-

gorien kann die Kategorienzuordnung die Gesamtnote erheblich beeinflussen. Deshalb muss die Kategorienzuordnung von Einzelleistungen in jedem Fall sachlich gerechtfertigt werden.

3.8 Zusammenfassung und Empfehlungen

Ich fasse die vorangehenden Überlegungen in folgende Empfehlungen zusammen:

- Soweit möglich sollten gleichartige Leistungen in dieselben Kategorien eingeordnet werden – schriftliche Leistungen wie Klassenarbeiten und schriftliche Wiederholungsarbeiten, aber auch andere Leistungen, in denen der schriftliche Leistungsaspekt dominiert – gehören in die Kategorie der schriftlichen Leistungen, mündliche Fachleistungen und solche Leistungen, in der der mündliche Leistungsaspekt dominiert, gehören in die Kategorie der mündlichen Leistungen etc. Die fachliche und fachdidaktische Diskussion muss klären, inwieweit ein erweiterter schulischer Leistungsbegriff und neue Beurteilungsformen im Zusammenhang mit neuen Unterrichtsformen zu weiteren Kategorien der Leistungsbeurteilung führen. Dabei sollte ausgewogen entschieden werden einerseits zwischen dem zwangsläufig entstehenden Interpretationsproblem, wenn zu viele heterogene Leistungen in einer Restkategorie („Sonstige Leistungen“) untergebracht werden, andererseits der praktischen Handhabbarkeit, wenn zu viele Kategorien kreiert werden.
- Innerhalb eines Leistungsbereichs sollten unterschiedliche Einzelleistungen gemäß der Formeln (4), (5) bzw. (6), (7) nicht nur unter Berücksichtigung ihres Gewichtes (Bedeutsamkeitsgewichtung), sondern auch ihrer Anzahlen (Anzahlgewichtung) zur Berechnung der Gesamtnote für den Leistungsbereich verrechnet werden. So müssen z.B. innerhalb der Leistungskategorie der schriftlichen Noten die schriftlichen Klassenarbeiten ebenso wie andere schriftliche Leistungen jeweils nicht nur mit ihrem jeweiligen Bedeutsamkeitsgewicht, sondern auch mit ihrer Anzahl in die Durchschnittsbildung der schriftlichen Leistungsnote eingehen. Darüber hinaus ist zu beachten, dass jeder Teilleistungsbereich durch eine hinreichende Anzahl von Einzelleistungen belegt ist, um die Zuverlässigkeit der Teilleistungsbeurteilung sicherzustellen.
- Ist für eine bestimmte Art von Leistung eine Gleichwertigkeit (Gleichgewichtung) mit bestimmten Leistungen in einer Kategorie vorgegeben (Beispiel: Die einer Klassenarbeit gleichwertige Feststellung einer Schülerleistung), dann sollte die entsprechende Note in diese Kategorie eingeordnet werden und mit entsprechendem Gewicht in die Berechnung eingehen. In der Interpretation der Gesamtnote für diesen Leistungsbereich ist ggf. die Heterogenität der Einzelleistungen zu berücksichtigen.
- Bestimmt eine Vorgabe (etwa aus der Fachkonferenz für das jeweilige Fach) die Gewichtung von Leistungsbereichen (z.B. in der Mathematik häufig die doppelte Gewichtung der schriftlichen im Verhältnis zu den mündlichen Leistungen), dann ist unbedingt darauf zu achten, dass die Leistungsbereiche durch möglichst *gleich viele* Einzelnoten belegt sind, da die Bedeutsamkeitsgewichtung ansonsten konterkariert wird.

4 Logikmodelle der Notenbildung

In diesem Abschnitt greife ich die in 1.2.4 dargestellten Beispiele auf. In diesen Beispielen geht es um die Frage, welche Leistungen bzw. welche Leistungskategorien als „Pflichtleistungen“ deklariert und unbedingt in die Notenberechnung einbezogen werden, und welche Leistungen bzw. Leistungskategorien als „Wahl-“, oder „Wahlpflichtleistungen“ nur bedingt in die Notenberechnung eingehen können. In 4.1 werden zunächst einige Beispiele explizit gerechnet und miteinander verglichen, ehe in 4.2 und 4.3 die dahinterstehenden Notenbildungsmodelle unter Bezug auf ähnliche Konzepte in der pädagogisch-psychologischen Diagnostik vorgestellt werden.

4.1 Praktische Beispiele zur Notenberechnung

Die folgende Tabelle 2 greift die Notenbilder der ersten vier Schülerinnen und Schüler aus dem Eingangsbeispiel 1.2.4 auf und ersetzt den Schüler Elias durch eine Schülerin Emilie. Der Einfachheit halber beschränken wir uns auf den Fall, dass von diesen 5 Schülern jeweils einige einzelne mündliche Fachnoten vorliegen sowie im einen oder anderen Fall Noten für eingebrachte Referate. Ich zeige einige Strategien der Integration der mündlichen Fachnoten und ggf. der Noten für die Referate zu einer mündlichen Gesamtleistungsnote.

Tabelle 2: Beispiele der Notenbilder von 5 Schüler und 5 Strategien der Notenberechnung. Die Ziffern in der Strategiekennzeichnung beziehen sich auf die Strategien 1, 2, 3 innerhalb der Grundentscheidung (I); K und E beziehen sich auf die Strategien innerhalb der Grundentscheidung (II). Näheres im Text.

	Mündliche Fachnoten					Durchschnitt	Referate			Durchschnitt	Mündliche Gesamtnote gemäß Strategie				
	M_1	M_2	M_3	M_4	\bar{M}		R_1	R_2	\bar{R}		1K	1E	2	2K	2E
Anna	2,5	1,0	1,0	1,5	1,5	-	-	-			1,5				
Ben	2,5	1,0	1,0	1,5	1,5	-	-	6,0	4,5	3,0					
Clara	5,0	-	-	-	5,0	2,5	1,5	2,0	3,0	2,6		3,0	2,6		
David	2,5	1,0	1,0	1,5	1,5	4,5	-	-			1,5				
Emilie	2,5	1,0	1,0	1,5	1,5	4,5	-	4,5	3,5	2,5					

Die denkbaren Strategien der Notenbildung lassen sich in diesem Beispiel, wo es um den Status von Referaten bei der Bildung der mündlichen Note geht, entlang zweier Grundentscheidungen sortieren:

Grundentscheidung (I):

Sind die Referate als *Pflicht-*, als *Wahl-* oder als *Wahlpflichtleistungen* anzusehen?

Strategie 1: Die Referate sind als *Pflichtleistungen* in dem Sinne anzusehen, dass von jedem Schüler die Einbringung eines Referates eingefordert wird und ihm auch die Gele-

genheit hierzu gegeben wird. Deshalb ist auch ein nicht erbrachtes Referat mit der Note „ungenügend“ (Ziffernote 6) zu bewerten und diese in die Notenberechnung einzubeziehen.

Strategie 2: Die Referate haben den Status von *Wahlleistungen* im Sinne einer freiwilligen Zusatzleistung; es kann gewählt werden, ob ein Referat gehalten und wenn ja, ob es gewertet wird oder nicht; ibs. wird ein nicht erbrachtes Referat aus der Berechnung der mündlichen Gesamtnote ausgenommen.

Eine Variante der Strategie 2 ist die folgende

Strategie 3: Die Referate werden als *Wahlpflichtleistungen* gewertet. Wird ein Referat gehalten, dann wird es auf jeden Fall gewertet. Wird ein Referat nicht gehalten, dann geht es nicht in die Bewertung ein.

□ **Grundentscheidung (II):**

Wie werden die Einzelleistungen gewichtet?

Strategie K: Die Fachleistungen und Referate werden separaten Kategorien zugeordnet und mit einer Kategoriengewichtung ohne Anzahlberücksichtigung eingerechnet gemäß Formel (8).

Strategie E: Die Fachleistungen und Referate werden mit ihren jeweiligen Anzahlen berücksichtigt gemäß Formel (5).

Die beschriebenen Strategien innerhalb der Grundentscheidung (I) sind natürlich nicht beliebig kombinierbar mit den Strategien innerhalb der Grundentscheidung (II). Nur im Falle, dass die Referate in die Berechnung einbezogen werden, kommt eine Wahl zwischen den beiden Strategien K und E zum Tragen.

Ich greife die Eingangsbeispiele in 1.2.4 auf und kläre die unterschiedlichen Berechnungsverfahren unter Bezug auf die oben beschriebenen Strategien auf. Um die Unterschiede zwischen den Strategien im Licht der berechneten Noten besonders augenscheinlich zu machen, sei angenommen, dass im Falle der Einbeziehung der Note für das oder die Referate diese jeweils doppelt so schwer wie die mündlichen Fachnoten gewichtet werden.

- Anna hat für die mündliche Fachleistung die Einzelnoten 2,5, 1,0; 1,0; 1,5 erhalten; ein Referat hat sie nicht gehalten, was ihr nicht nachteilig angerechnet wird. Auf der Grundlage der Strategie 2 errechnet der Lehrer gemäß Formel (1):

$$\bar{M} = \frac{2,5 + 1,0 + 1,0 + 1,5}{4} = 1,5.$$

- Ben hat die gleichen Einzelnoten wie Anna erhalten, nämlich 2,5; 1,5; 2,0; 1,0, sich jedoch der Einbringung eines Referats verweigert. Die Lehrerin erachtet die Übernahme eines Referates als Pflichtleistung und gewichtet die Leistungssparte der Referate grundsätzlich doppelt so stark wie die Sparte der sonstigen mündlichen Fachnoten. Sie notiert für die nicht erbrachte Leistung die Note 6 und berechnet die mündliche Leistung nach Strategie 1 in Kombination mit Strategie K (Strategie 1K) wie folgt:

$$\bar{M} = \frac{1 \cdot [(2,5 + 1,0 + 1,0 + 1,5) / 4] + 2 \cdot 6,0}{1 + 2} = \frac{1 \cdot 1,5 + 2 \cdot 6,0}{3} = 4,5.$$

Strategie 1 in Kombination mit Strategie E (Strategie 1E) liefert hingegen (gemäß Formel (4) bzw. (5)):

$$\bar{M} = \frac{1 \cdot (2,5 + 1,0 + 1,0 + 1,5) + 2 \cdot 6,0}{1 \cdot 4 + 2 \cdot 1} = \frac{1 \cdot 4 \cdot 1,5 + 1 \cdot 2 \cdot 6,0}{6} = 3,0.$$

Der dramatische Unterschied rührt also in diesen beiden Berechnungsweisen von der Unterschiedlichkeit der Kategoriengewichtung im Vergleich mit der Einzelleistungsgewichtung.

- Clara erhielt nur eine mündliche Fachnote 5,0; sie hat jedoch mit zwei sehr ordentlichen Referaten aufgewartet, die mit 2,5 und 1,5 bewertet wurden. Sicher wird die Lehrerin auf der Grundlage von Strategie 2 im Einvernehmen mit Clara und unbedingt aufgrund von Strategie 1 und ebenso 3 die Referate mit der Gewichtung 2 in die Berechnung der mündlichen Gesamtnote einbeziehen. Allerdings kann dies wiederum unter Verwendung der Strategie K erfolgen:

$$\bar{M} = \frac{1 \cdot [5,0] + 2 \cdot [(2,5 + 1,5) / 2]}{1 + 2} = \frac{1 \cdot 5,0 + 2 \cdot 2,0}{3} = 3,0;$$

oder die Lehrerin verwendet die Strategie E und errechnet (gemäß Formel (4) bzw. (5)):

$$\bar{M} = \frac{1 \cdot 5,0 + 2 \cdot 2,5 + 2 \cdot 1,5}{1 \cdot 1 + 2 \cdot 2} = \frac{13,0}{5} = 2,6.$$

- Für David hat der Lehrer für die mündliche Fachleistung die Einzelnoten 2,5; 1,0; 1,0; 1,5 notiert, für ein eher misslungenes Referat die Note 4,5. Der Lehrer deklariert gemäß Strategie 2 die Referate als freiwillige Wahlleistung und rechnet sie nur dann in die Note ein, wenn David dies wünscht bzw. wenn die Note für das Referat die mündliche Gesamtnote deutlich aufwertet. In diesem Fall wird er die mündliche Gesamtnote wohl nach Formel (1) ohne die Note für das Referat berechnen und zu folgender Note gelangen:

$$\bar{M} = \frac{2,5 + 1,0 + 1,5 + 1,0}{4} = 1,5.$$

- Emilie hat die gleichen mündlichen Einzelnoten Noten wie David sowie für ein misslungenes Referat die Note 4,5. Die Lehrerin verfährt nach Strategie 1 und wertet das Referat als Pflichtleistung. Kombiniert sie mit Strategie K, so errechnet sie

$$\bar{M} = \frac{1 \cdot [(2,5 + 1,0 + 1,5 + 1,0) / 4] + 2 \cdot 4,5}{1 + 2} = \frac{1 \cdot 1,5 + 2 \cdot 4,5}{3} = 3,5;$$

kombiniert die Lehrerin Strategie 1 mit Strategie E, so errechnet sie (gemäß Formel (4) bzw. (5)):

$$\bar{M} = \frac{1 \cdot 2,5 + 1 \cdot 1,0 + 1 \cdot 1,5 + 1 \cdot 1,0 + 2 \cdot 4,5}{1 \cdot 4 + 2 \cdot 1} = \frac{15,0}{6} = 2,5.$$

In den jeweiligen Einzelfällen lässt sich die Notengebung durchaus legitimieren. Ist die Notengebung aber über diese Einzelfälle hinweg konsistent und gerecht?

Diese Beispiele fassen eine in der Praxis häufige Konstellation auf: In vielen Fächern wird erwartet, dass der Schüler ein Referat in den Unterricht einbringt. Durchaus ist die erörterte Grundfragestellung aber auch auf andere Leistungsaspekte übertragbar.

Beispiele:

- Der Lehrer sieht für ein Fach, in welchem die aktive Unterrichtsbeteiligung besonders relevant ist und in dem die mündliche Teilleistung gleich gewichtet wie die schriftliche Leistung in die Gesamtnote eingeht, für die Kategorie der mündlichen Note eine Leistungssparte „Förderung des Unterrichtsprozesses“ vor.
- Der Lehrer erwartet von jedem Schüler, dass er mindestens einmal einen Hausaufgabenaufsatz zur Bewertung vorlegt.
- Der Lehrer erwartet in seinem Fach mindestens eine praktische Leistung.

In allen solchen Situationen stehen gleichartige Entscheidungen für die Notenbildung an, wie sie oben für das Beispiel von Referaten analysiert wurden.

In den Beispielen dieses Unterabschnitts haben wir Pflichtleistungen, Wahlleistungen und Wahlpflichtleistungen im Zusammenhang mit der Notenbildung gegenüber gestellt. Ich setze im Folgenden die angestellten Überlegungen in Bezug zu ähnlichen Konzepten in der Psychologischen Diagnostik.

4.2 Adjunktives Modell (Oder-Modell) der Notenbildung

In der Aussagenlogik bezeichnet die *Adjunktion* $A \vee B$ („A oder B oder beides“; andere Bezeichnungen: *nicht-ausschließende Disjunktion* oder *inklusives Oder*) zweier Aussagen A und B eine zusammengesetzte Aussage, die wahr ist genau dann, wenn mindestens eine der beiden Aussagen A oder B wahr ist; sie ist also falsch genau dann, wenn sowohl A als auch B falsch sind.

Übertragen auf die Notenbildung im Zusammenhang mit schulischer Leistungsbewertung kann man unter dem *Adjunktiven Modell* folgendes verstehen:

4.2.1 Adjunktives Modell (Oder-Modell) der Notenbildung:

- (1) In die Berechnung der Gesamtleistung einer Kategorie werden mindestens eine Einzelleistung, aber nicht notwendig alle Einzelleistungen einbezogen.
- (2) Existieren in einer Leistungskategorie Unterkategorien, so werden mindestens eine Unterkategorie, aber nicht notwendig alle Unterkategorien in die Berechnung der Gesamtnote der Leistungskategorie einbezogen.

Beispiele:

- Ein Schüler hat mehrere „Versuche“, eine möglichst gute Leistung im sportlichen Weitsprung zu erzielen, dann kann er wählen, welche Leistung gewertet wird, wobei er dann i.d.R. die beste dieser Leistungen werten lassen wird.
- Ein Schüler hat eine sehr gute Durchschnittsleistung in der Kategorie der mündlichen Fachnoten, jedoch ein misslungenes Referat gehalten; nun kann er entscheiden, ob die schlechte Note für dieses Referat in die Berechnung eingehen oder ausgeschlossen werden soll, wobei er vermutlich die zweite Option wählen wird.

Das Adjunktive bzw. *Oder-Modell* der Notenbildung ist eine extreme Variante kompensatorischer Notenbildung. Es ist nicht notwendig, die Note aus allen Teilnoten zu berechnen, sondern für eine gute Gesamtnote genügen gute Einzelnoten in *einer* Kategorie, d.h. die Leistung kann auf die eine oder andere Weise erbracht werden.

Zum o.g. Beispiel: Ein Schüler erhält eine sehr gute mündliche Gesamtnote, weil er sich bei gezielter Befragung wenige Male mit einer sehr guten Einzelleistung hervorgetan, ansonsten aber keinen förderlichen Beitrag zum Unterrichtsprozess geleistet hat.

Umgekehrt: Ein Schüler bringt kontinuierlich konstruktive Beiträge in den Unterrichtsprozess ein. Bei den gezielten Befragungen schneidet er allerdings schlecht ab. Der Lehrer lässt jedoch diese schwachen Leistungen bei der Festsetzung der mündlichen Gesamtnote unberücksichtigt.

Einem solchen Verfahren der Notenbildung liegt also ein adjunktives Modell der Notenbildung zugrunde. Ob dies so zu legitimieren ist, ist eine grundsätzliche fachliche oder fachdidaktische Frage. Amelang und Schmidt-Atzert (2006, 400) bemerken dazu:

„Kompensatorische Strategien sind immer dort dysfunktional, wo in jedem Teilbereich bestimmte Mindestleistungen unabdingbar vorliegen müssen, um eine Tätigkeit erfolgreich ausführen zu können. Beispielsweise kann ein Chirurg nicht mangelnde feinmotorische Kompetenz durch Intelligenz kompensieren.“

Ein Adjunktives bzw. *Oder-Modell* der Notenbildung ist im diskutierten Beispiel also dann nicht angemessen, wenn es gute Gründe für die Auffassung gibt, ein Schüler müsse seine Leistungsfähigkeit im mündlichen Bereich sowohl in seinen mündlichen Beiträgen als auch darin zeigen, dass er förderlich zum Unterrichtsprozess beiträgt. In diesem Falle können die einzelnen Leistungskomponenten nicht wahlweise unberücksichtigt bleiben.

4.3 Konjunktives Modell (*Und-Modell*) der Notenbildung

Eine *Konjunktion* $A \wedge B$ (*A und B*) bezeichnet in der Aussagenlogik eine aus zwei Aussagen A und B zusammengesetzte Aussage, die genau dann wahr ist, wenn sowohl die Aussage A als auch die Aussage B wahr ist; andernfalls ist die Aussage $A \wedge B$ falsch.

Sinngemäß übertragen auf die Notenbildung bedeutet die *Konjunktive Notenbildung*:

4.3.1 *Konjunktives Modell (Und-Modell) der Notenbildung:*

- (1) In die Berechnung der Gesamtleistung einer Kategorie werden bedingungslos alle Einzelleistungen einbezogen.
- (2) Existieren in einer Leistungskategorie Unterkategorien, so werden alle Unterkatego-

rien bedingungslos in die Berechnung der Gesamtnote der Leistungskategorie einbezogen.

Beispiele:

- Ein Schüler kann im mündlichen Leistungsbereich nicht die Note „sehr gut“ erhalten, wenn er sich nicht in Eigeninitiative fördernd in den Unterricht eingeschaltet hat.
- Eine mündliche Fachleistung setzt sich zusammen aus den Einzelaspekten: Reproduktion, Anwendung, Transfer, etc.
- Die Fachnote setzt sich zwingend zusammen aus der mündlichen und der schriftlichen Gesamtleistung – u.U. auch noch aus anderen, wie etwa den praktischen Leistungen.

In allen diesen Fällen, wo in jedem Teilbereich, in jeder Kategorie bestimmte Mindestleistungen erbracht werden müssen, ist eine konjunktive Notenbildung zwingend.

4.4 Allgemeine Logikmodelle der Notenbildung

Pflichtleistungen gemäß der Strategie 1 in 4.1 können dem Konjunktiven Modell der Notenbildung wie in 4.3 beschrieben zugeordnet werden. Hierbei werden alle Einzelleistungen einer Kategorie bzw. alle Unterkategorien einer Kategorie in die Bildung einer Gesamtnote einbezogen. Dagegen erfasst das Adjunktive Modell aus 4.2 nicht genau das Vorgehen in 4.1 im Falle von Wahl- oder Wahlpflichtleistungen. Das dort gewählte Beispiel, dass ein Referat zusätzlich zu sonstigen mündlichen Leistungen in die Bildung der mündlichen Gesamtleistungen entweder einbezogen werden kann oder nicht einbezogen werden kann, beinhaltet eine Asymmetrie im Adjunktiven Modell: In jedem Falle soll nämlich der Durchschnitt der mündlichen Fachnoten in die Notenbildung einbezogen werden und nur wahlweise zusätzlich die Note aus der Unterkategorie der Referate.

Zur Charakterisierung verschiedener Logikmodelle der Notenbildung werden folgende Bezeichnungen eingeführt²⁰:

- X_1, X_2, \dots, X_k denotieren Kategorien mit einer darin enthaltenen unbestimmten Anzahl von Einzelnoten
- X_i stehe für $i = 1, \dots, k$ zugleich für die Aussage: „Die Kategorie X_i wird in die Notenbildung (die Berechnung der Gesamtnote) einbezogen.“

Dann können wir mit Hilfe logischer Operatoren folgende *Logikmodelle der Notenbildung* charakterisieren:

²⁰ Die folgende formale Inkorrektheit der gleichen Bezeichnung für unterschiedliche Entitäten sei der besseren Lesbarkeit ohne die Befürchtung von Missverständnissen geschuldet.

4.4.1 Modell der Pflichtleistungen (Konjunktives Modell):

$$X_1 \wedge X_2 \wedge \dots \wedge X_k$$

Alle Kategorien werden in die Notenbildung einbezogen.

4.4.2 Modell der freien Wahlleistung:

$$X_1 \vee X_2 \vee \dots \vee X_k$$

Es besteht Wahlfreiheit, welche Kategorien zur Notenbildung herangezogen werden.

4.4.3 Modell der relativen (bedingten) Wahlleistung:

Stehen X_1, \dots, X_r ($1 \leq r \leq k - 1$) für die Aussagen, dass die Leistungen der entsprechenden Kategorien auf jeden Fall in die Notenbildung einbezogen werden sollen, und $D := X_1 \wedge \dots \wedge X_r$ für die Konjunktion dieser Aussagen; darüber hinaus mögen die Aussagen X_{r+1}, \dots, X_k beinhalten, dass die entsprechenden Kategorien wahlweise, aber nicht verpflichtend in die Notenbildung einbezogen werden sollen²¹. Dann lässt sich dieses Notenbildungsmodell durch folgende Aussage charakterisieren:

$$D \vee (D \wedge X_{r+1}) \vee (D \wedge X_{r+2}) \vee \dots \vee (D \wedge X_k),$$

was logisch äquivalent ist zu

$$D \wedge (D \vee X_{r+1} \vee \dots \vee X_k).$$

4.4.4 Modell der bedingten Wahlpflichtleistung:

Mit den gleichen Bezeichnungen wie oben beinhaltet die Aussage

$$D \wedge (X_{r+1} \vee X_{r+2} \vee \dots \vee X_k),$$

dass die Leistungen in den Kategorien X_1, \dots, X_r auf jeden Fall in die Notenbildung einbezogen werden, darüber hinaus verpflichtend aber mindestens noch eine oder mehrere der anderen Kategorien.

Da für die Aussagenlogik die Satzoperatoren \wedge , \vee zusammen mit der Negation (Symbol: \neg) ein vollständiges System bilden, mit welchem sämtliche aussagenlogischen Ausdrücke gebildet werden können, können wir mit den obigen logikbasieren Modellen der Notenbildung sämtliche Modelle der pflichtmäßigen oder wahlweisen Einbeziehung bestimmter Leistungen in die Notenbildung abbilden.

²¹ Ggf. durch Umbenennung lassen sich die Kategorien in dieser Reihenfolge ordnen.

Der Vollständigkeit halber sei im Folgenden noch eine allgemeine Berechnungsformel für die Notenbildung im Falle von logischen Mischmodellen angegeben. Wir benötigen hierzu eine Funktion H , die jeder Kategorie X_i eine 0 oder eine 1 zuweist in Abhängigkeit davon, ob diese Kategorie in die Notenbildung einbezogen werden soll oder nicht, also für jede Kategorie X_i :

$$H(X_i) =: \begin{cases} 0 & \text{falls die Kategorie } X_i \text{ in die Notenbildung einbezogen werden soll} \\ 1 & \text{falls die Kategorie } X_i \text{ nicht in die Notenbildung einbezogen werden soll} \end{cases}$$

wobei gelten soll, dass nicht alle $H(X_i)$ den Wert 0 annehmen. Dann lautet eine allgemeine Formel für die Errechnung eines gewichteten Gesamtwertes \bar{X}_g aus k Kategorien mit den Durchschnitten $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ und den Gewichten Φ_1, \dots, Φ_k , wobei die Kategorien gemäß einem vorgegebenen Notenbildungsmodell in die Berechnung einbezogen werden sollen, mit Rückgriff auf Formel (8):

$$(11) \quad \bar{X}_g = \frac{\sum_1^k \Phi_i H_i \bar{x}_i}{\sum_1^k \Phi_i H_i} = \frac{\Phi_1 H_1 \bar{x}_1 + \dots + \Phi_k H_k \bar{x}_k}{\Phi_1 H_1 + \dots + \Phi_k H_k}$$

In die Berechnung der Gesamtnote gehen damit nur solche Kategorien ein, die nach dem vorgegebenen Logikmodell einbezogen werden sollen und somit den H -Wert 1 erhalten. Entsprechend der Bemerkung in 3.5.2 ist in dieser Formel auch die Formel (5) enthalten, nämlich für den speziellen Fall, dass die Kategoriengewichte von der Form $\Phi_i = \gamma_i s_i$ sind, wobei die γ_i die vorgegebene Bedeutsamkeitsgewichtung der Einzeldaten in der Kategorie X_i und die s_i die Anzahlgewichtungen der Kategorie X_i entsprechend der Anzahl der in der Kategorie X_i vorhandenen Einzeldaten darstellen.

4.5 Zusammenfassung und Schlussfolgerung

Die Betrachtungen in diesem Abschnitt 4 führen zu folgenden Schlussfolgerungen:

- In der Notenbildung im Sinne der Berechnung einer Gesamtnote aus Einzelnoten bzw. einer Gesamtnote in einer Kategorie mit Unterkategorien sind zwei Grundmodelle zu unterscheiden: Das *Adjunktive Modell* bietet eine Option, nicht alle, sondern nur bestimmte Einzelleistungen oder Unterkategorien in die Berechnung der Gesamtnote einzubeziehen; das *Konjunktive Modell* erfordert, dass alle Einzelleistungen bzw. alle Unterkategorien in die Berechnung der Gesamtnote einbezogen werden.
- Aus dem Adjunktiven und dem Konjunktiven Grundmodell lassen sich logische *Mischmodelle* konstituieren, die in ihren Charakteristika alle in der Praxis auftretende Modelle der Einbeziehung von Pflicht-, Wahl- und Wahlpflicht-Leistungen in die Notenbildung erfassen.
- Prinzipiell lassen sich alle Logikmodelle der Notenbildung unter Bezug auf die praktischen Gegebenheiten und Erfordernisse legitimieren.

- Die *Gerechtigkeit* in der Notenbildung erfordert jedoch grundsätzlich eine Gleichbehandlung aller Schüler im gegebenen Kontext, wobei ein Prinzip der *Billigkeit* u.U. die besondere Behandlung im Einzelfall rechtfertigen könnte. In jedem Fall erfordert jedoch der Grundsatz der *Transparenz*, dass der Lehrer seine Verfahrensweise bei der Notenbildung gegenüber den Schülern (ggf. den Eltern, Kollegen etc.) offenlegt.

5 Dimensionen und Abhängigkeiten von Leistungen

In den vorangegangenen Betrachtungen zur Notenbildung habe ich weitestgehend unterstellt, dass die zu einer Gesamtnote zusammenfassenden Leistungsaspekte voneinander „unabhängig“ sind. Bei der Integration von Leistungsaspekten zu einer Note oder bei der Integration mehrerer Noten zu einer zusammenfassenden Note ist es jedoch sehr wichtig zu unterscheiden, ob die einzelnen Leistungsaspekte bzw. die Einzelnoten voneinander unabhängig sind oder ob sie voneinander abhängen. „Unabhängig“ sind Leistungsaspekte dann, wenn die zugrunde liegenden Aufgaben oder Anforderungen keinen sachlichen Zusammenhang aufweisen, indem z.B. die eine Aufgabe oder Anforderung die andere als Teilaspekt oder als Voraussetzung beinhaltet²².

Die Problematik wird in Beispiel 1.2.3, angesprochen: Wenn Schüler Markus immer wieder in den Unterricht sehr anspruchsvolle Beiträge einbringt, die das erworbene Wissen reflektiert und ideenreich auf neue Sachverhalte anwenden, dann kann man unterstellen, dass er – wie Lisa – auf Anforderung auch das zugrunde liegende Basiswissen wiedergeben oder diese Kenntnisse auf bekannte Aufgabenstellungen anwenden kann. Kann der Lehrer also darauf verzichten, bei Markus das Basiswissen abzufragen?

Hinsichtlich der Abhängigkeiten von Leistungen unterscheide ich im Folgenden drei Grundmodelle der Notenbildung:

- (1) das *Dimensionen-Modell* mit unabhängigen Leistungsdimensionen
- (2) das *Taxonomie-Modell* mit hierarchisch geordneten Leistungsstufen
- (3) Modelle mit komplexen Leistungsabhängigkeiten.

5.1 Das Dimensionen-Modell

Mit dem Verständnis bestimmter Leistungsaspekte als *Dimensionen*²³ verbindet man das Merkmal, dass die Leistungsaspekte in den verschiedenen Dimensionen *unabhängig* voneinander variieren können.

²² Ich verzichte hier auf eine genaue Definition der „Unabhängigkeit“. Gemeint ist eine „lokale stochastische Unabhängigkeit“ etwa im Sinne der Probabilistischen Testtheorie. Eine Korrelation zweier unabhängiger Leistungen ergibt sich dann lediglich aus dem Umstand, dass ein guter Schüler sowohl die eine wie die andere Leistungsanforderung gut bewältigt ebenso wie ein schlechter Schüler eher bei beiden Anforderungen schlecht abschneidet als dass er die eine gut, die andere schlecht bewältigt.

²³ im Sinne des psychologisch-diagnostischen Konzeptes von Leistungsdimensionen

In einem fremdsprachlichen Fach mögen z.B. die Aspekte „Aussprache“ und „Wortschatz“ relativ unabhängig voneinander variieren. Eine Korrelation der beiden Aspekte wird sich nur durch den Umstand ergeben, dass z.B. ein guter Schüler wahrscheinlich sowohl in der Aussprache wie auch im Wortschatz gut ist, ein schlechter Schüler wird dagegen eher in beiden Aspekten schlecht sein. In der Art der Leistungsanforderung lässt sich aber vermutlich kaum eine systematisch-sachliche Abhängigkeit begründen. Diese beiden Leistungsaspekte könnte man in diesem Sinne als unabhängige *Dimensionen* der mündlichen Leistungsnote verstehen.

Für solche in diesem Sinne unabhängige Leistungsaspekte können dann entweder Einzelnoten gegeben werden, oder sie können in begründeter Gewichtung in eine Einzelnote integriert werden. Auch beruht die in der Beurteilungspraxis übliche Unterscheidung zwischen „mündlichen Noten“ und „schriftlichen Noten“ auf der nicht abwegigen Unterstellung, dass die mündlichen Leistungen und die in schriftlichen Prüfungen erbrachten Leistungen zwei voneinander unabhängigen Leistungsdimensionen zuzuordnen sind. Unabhängigkeit wird in der Praxis auch für die aufeinanderfolgenden Leistungen eines einzelnen Schülers angenommen; er hat immer wieder eine „neue Chance“ der Leistungserbringung, obgleich testtheoretisch bzw. statistisch solche Daten eher als „abhängig“ gelten würden.

Die in Abschnitt 3 diskutierten Gewichtungsverfahren eignen sich zunächst einmal für die Integration solcher voneinander *unabhängiger Einzelleistungen* bzw. *voneinander unabhängiger Leistungsdimensionen*.

5.2 Taxonomie-Modelle

Von unabhängigen Leistungsdimensionen grundsätzlich zu unterscheiden sind *Taxonomien* mit Leistungsstufen und Leistungsabhängigkeiten. Sie stellen hierarchische Modelle für die zunehmende Schwierigkeit oder Komplexität von Lernzielen und Lernleistungen dar. Die bekanntesten Beispiele von Taxonomien sind sicherlich die *Taxonomie von Bloom (1956) für kognitive Lernziele*²⁴ und die revidierte Taxonomie von Anderson und Krathwohl (2001). Eine einfache Taxonomie in drei Stufen beschreibt aber auch das im schulischen Kontext weit verbreitete *Drei-Stufen-Modell* mit drei gestuften Anforderungsbereichen, wie es beispielsweise in Paradies, Wester & Greving (2005) dargestellt wird:

Der **Anforderungsbereich I** umfasst die Wiedergabe von Sachverhalten aus einem abgegrenzten Gebiet im gelernten Zusammenhang und die Verfügbarkeit der notwendigen inhaltlichen und methodischen Kenntnisse in einem begrenzten Gebiet und einem wiederholenden Zusammenhang. Dazu gehören: ...

Der **Anforderungsbereich II** umfasst das selbstständige Erfassen und Einordnen bekannter Sachverhalte, das Strukturieren und gedankliche wie sprachlich angemessene Be- und Verarbeiten dieser Sachverhalte sowie das selbstständige Anwenden und Übertragen des Gelernten auf vergleichbare Sachverhalte. Dazu gehören: ...

Der **Anforderungsbereich III** umfasst das planmäßige Verarbeiten und die eigenständige Reflexion komplexer Gegebenheiten mit dem Ziel, zu selbstständigen Begründungen, Folgerungen, Deutungen und Wertungen zu gelangen und eigene Lösungsansätze zu entwickeln. Dazu gehören: ...

²⁴ Darüber hinaus existieren auch für die anderen Lernzieldimensionen eigene Taxonomien. In der Gegenwart kursieren verschiedene Überarbeitungen und Weiterentwicklungen dieser Taxonomien.

Auch die PISA-Tests arbeiten mit Taxonomiestufen, die zumeist empirisch aus den Daten rekonstruiert werden, wobei Bemühungen in Richtung einer theoriegeleiteten Aufgabenkonstruktion auf verschiedenen Niveaustufen gehen.

Für eine Taxonomie grundlegend ist die folgende

Hauptthese: *Das Erreichen von Zielen bzw. die Bewältigung einer Anforderung auf höherer Taxonomiestufe setzt jeweils das Erreichen von Zielen bzw. die Bewältigung der Anforderungen auf allen vorausgehenden Taxonomiestufen voraus.*

Kurz: In der Leistungsanforderung einer höheren Taxonomiestufe sind die Leistungsanforderungen aller niedrigeren Stufen im Sinne von Voraussetzungen oder Komponenten enthalten.

In der praktischen Anwendung von Taxonomie-Modellen ist sicherlich zwischen *Lernsituationen* und *Leistungssituationen* zu unterscheiden. Für die Organisation des Lehrangebots in Lernsituationen leisten Taxonomie-Modelle wertvolle, unverzichtbare Dienste. Zuerst müssen Anforderungen auf einer unteren Ebene bewältigt werden können, dann kann die nächste Anforderungsstufe angegangen werden. Jede Sequenzierung von Lernanforderungen und Lernzielen, jedes Curriculum berücksichtigt diese Lernabhängigkeiten zumindest als ein notwendiges Leitprinzip. Ungenauigkeiten in der Modellierung der Anforderungsabhängigkeiten führen vielleicht zu einem „Stolpern“ im Lernprozess, lassen aber in der Regel keine Unfälle befürchten.

Strenger sind die definatorischen Merkmale von Taxonomie-Modellen einzufordern, wenn diese in Leistungssituationen als Prüfungsmodelle eingesetzt werden. Hier muss die strenge Forderung gelten:

- (1) *Jede Taxonomiestufe muss durch repräsentative Aufgaben und Anforderungen vertreten sein.*
- (2) *Jede Aufgabe / Anforderung einer höheren Taxonomiestufe muss Anforderungen der unteren Taxonomiestufen beinhalten oder voraussetzen.*

Der kritische Punkt eines solchen Taxonomie-Modells im Zusammenhang einer Leistungsprüfung liegt auf der Hand: Es muss gelingen, jede Taxonomiestufe durch repräsentative Aufgaben zu operationalisieren. Auf diese Schwierigkeiten, die sich nur fachspezifisch von den jeweiligen Fachdidaktiken bearbeiten lassen, gehe ich an dieser Stelle nicht ein.

Eine andere Problematik im Zusammenhang mit Taxonomie-Modellen soll hier jedoch näher betrachtet werden: Leider ist mit den meisten gängigen Taxonomien schulischen Wissens kein adäquates Modell der Notenbildung verbunden. Denn dass eine Leistung auf einer höheren Ebene Leistungen auf unteren Ebenen voraussetzt, begründet eine Abhängigkeit dieser Leistungen, die auf die Notenbildung Einfluss nehmen muss. Die auf solchen Abhängigkeiten aufsetzende Notenbildung wirft jedoch einige Fragen auf:

- (1) Wie werden Leistungen auf den unterschiedlichen Taxonomiestufen auf die Notenskala abgebildet? Im Beispiel unter 1.2.3 wird diese Frage aufgeworfen.
- (2) Wie wird aus Leistungen auf den unterschiedlichen Taxonomiestufen eine Gesamtnote gebildet?

Im Unterabschnitt 5.4 werden Verfahren der Notenbildung bei zugrundeliegenden Taxonomien diskutiert.

5.3 Modelle mit komplexen Leistungsabhängigkeiten

Taxonomie-Modelle sind bei Weitem nicht die komplexesten Modelle von Abhängigkeiten. Abhängigkeiten zwischen zwei Aufgaben oder Anforderungen *a* und *b* der Art

Wer die Aufgabe a lösen kann, der kann auch eine Aufgabe b lösen bzw.

Wer die Anforderung a bewältigt, der kann auch die Anforderung b bewältigen

kann man als sog. „Ordnungsrelation“ verstehen und in Gestalt von *Partiellen Ordnungen* oder *Quasiordnungen* modellieren. In solchen Partiellen oder Quasiordnungen wird noch nicht einmal gefordert, dass stets zwei Aufgaben oder Anforderungen hinsichtlich ihrer Lösungserwartungen vergleichbar sind. In diesen komplexen Leistungszusammenhängen wurde m.W. ein rational begründbares Modell der Notenbildung noch nicht entwickelt. Im folgenden Abschnitt gehe ich nur auf die Notenbildung im Falle einer taxonomischen Ordnung von Leistungsaspekten ein.

5.4 Notenbildung für taxonomisch geordnete Leistungen

Das Beispiel in 1.2.3 spricht die Problematik der Notenbildung für den Fall an, dass von zwei Schülern eine ungleiche Anzahl von Leistungen auf taxonomisch verschiedenen Stufen erbracht werden. Betrachten wir dieses Beispiel näher.

5.4.1 Beispiel

Auf der Grundlage einer dreifach gestuften Anforderungshierarchie zeigt Tabelle 3 die Notenbilder der Schüler Markus und Lisa aus dem Beispiel in 1.2.3. Von Markus liegen gute Leistungen nur in den Anforderungsbereichen II und III vor, von Lisa nur in den Anforderungsbereichen I und II.

Tabelle 3. Notenbilder der Schüler/in Markus und Lisa in drei Anforderungsbereichen einer Taxonomie

	Noten im Anforderungsbereich I			Noten im Anforderungsbereich II			Noten im Anforderungsbereich III		
Markus				2,5			1,5	2,0	1,0
Lisa	1,5	2,0	1,0	2,5					

Die einfachste Strategie der Notenberechnung würde alle Leistungen, ungeachtet des Anforderungsniveaus, auf dem sie erbracht wurden, mit gleicher Einzelleistungsgewichtung zusammenfassen und käme zu folgenden Berechnungen:

$$\text{Markus: } \bar{N} = \frac{2,5 + 1,5 + 2,0 + 1,0}{4} = \frac{7,0}{4} = 1,75$$

$$\text{Lisa: } \bar{N} = \frac{1,5 + 2,0 + 1,0 + 2,5}{4} = \frac{7,0}{4} = 1,75$$

Offenbar würde die in Beispiel 5.4.1 angewandte Bewertungsstrategie dem Sinn einer Anforderungsstufung widersprechen, denn die Leistungen von Markus, die überwiegend auf der Anforderungsstufe III erbracht wurden, sind insgesamt besser zu bewerten als die Leistun-

gen von Lisa, die überwiegend der Anforderungsstufe I zuzuordnen sind. Wie können diese Leistungen auf unterschiedlichen Anforderungsstufen durch ein geeignetes Notenbildungsverfahren angemessen gewürdigt werden? Auf die Lösung dieser Frage weist folgende Idee: Wir müssen die Leistungen in den verschiedenen Anforderungsbereichen offenbar verschiedenen Kategorien zuordnen und ihnen jeweils besonderes Gewicht beimessen.

Sicherlich überzeugend ist die pragmatische Klärung der Frage der Zuordnung von Leistungen auf verschiedenen Taxonomiestufen: Sie werden unterschiedlichen Leistungskategorien zugeordnet, wie dies bereits durch Tabelle 3 nahe gelegt wird. Damit kann man nichts falsch machen, denn bei Bedarf ist die Zusammenlegung verschiedener Kategorien unter Berücksichtigung vorgegebener Gewichtungen stets möglich. Wir ordnen also die Leistungen, die auf n verschiedenen Taxonomiestufen erbracht wurden, n verschiedenen Kategorien zu.

Nun stehen vor dem Hintergrund der vorangegangenen Abschnitte folgende weitere Strategie-Entscheidungen an:

- Grundentscheidung (I):** Welches Logikmodell ist für die Einbeziehung der verschiedenen Anforderungsstufen bzw. Leistungskategorien in die Notenberechnung angemessen?
- Grundentscheidung (II):** In welchen Gewichtungen werden ggf. Leistungen auf verschiedenen Anforderungsstufen der Taxonomie bzw. aus verschiedenen Leistungskategorien in die Berechnung einer Gesamtnote einbezogen?

Widmen wir uns zunächst der Grundentscheidung (I). Hier wird schnell ersichtlich, dass es unangemessen wäre, die drei Anforderungsbereiche im Sinne eines Adjunktiven Modells (eines Wahl- oder Wahlpflichtmodells – siehe 4.4) in die Notenbildung einzubeziehen. Würde nämlich beispielsweise der eine Schüler ausschließlich Leistungen auf dem untersten Anforderungsniveau erbringen, der andere Schüler ausschließlich auf dem obersten Anforderungsniveau, dann könnte selbst eine unterschiedliche Gewichtung keine Differenzierung in der Notenberechnung bewirken.

Die Notenberechnung auf der Grundlage eines Taxonomie-Modells muss somit stets einem *Konjunktiven Modell* der Notenbildung folgen. Vom Schüler sind in jedem Anforderungsbereich Pflichtleistungen zu erbringen, und die Lehrperson hat für eine Leistungserbringung die Gelegenheit anzubieten. Wird von einem Schüler in einem der Anforderungsbereiche keine Leistung erbracht oder eine solche verweigert, dann ist dieser Anforderungsbereich entsprechend geltender Regelung mit der Note „ungenügend“ zu bewerten.

Löst diese Festlegung bereits die Frage nach der Differenzierung hinsichtlich der Anforderungsstufen? Wir greifen unser Beispiel 5.4.1 auf und unterstellen, dass die beiden Schüler trotz gegebener Gelegenheit in bestimmten Leistungsbereichen keine Leistung erbracht haben. Es gibt dann die uns bekannten zwei Möglichkeiten, diese nicht erbrachten Leistungen in die Notenbildung einzubeziehen, wobei wir zunächst mit einer Gleichheit aller Gewichtungen rechnen:

1. Möglichkeit: In dem nicht durch eine Leistung belegten Anforderungsbereich wird eine Einzelnote 6,0 eingetragen. Die Gesamtnote wird sodann als Durchschnitt aller (gleich gewichteter) Einzelnoten berechnet. Tabelle 4 folgt dieser Berechnung.

Tabelle 4. Notenbilder von Markus und Lisa, wobei nicht erbrachte Leistungen auf einzelnen Anforderungsstufen durch die Einzelnote 6,0 belegt werden. Die letzte Spalte zeigt die sich ergebende Gesamtnote.

	Noten im Anforderungsbereich I			Noten im Anforderungsbereich II			Noten im Anforderungsbereich III			\bar{N}
Markus	6,0			2,5			1,5	2,0	1,0	2,6
Lisa	1,5	2,0	1,0	2,5			6,0			2,6

2. **Möglichkeit:** Der gesamte Anforderungsbereich wird mit der Note 6,0 bewertet. Die Gesamtnote wird dann aus den Durchschnittsnoten der einzelnen Anforderungsbereiche ermittelt. In Tabelle 5 findet sich diese Berechnung.

Tabelle 5. Notenbilder von Markus und Lisa, wobei nicht erbrachte Leistungen auf einzelnen Anforderungsstufen durch die Kategoriennote 6,0 belegt werden. Die letzte Spalte zeigt die sich ergebende Gesamtnote.

	Noten im Anforderungsbereich I			\emptyset	Noten im Anforderungsbereich II			\emptyset	Noten im Anforderungsbereich III			\emptyset	\bar{N}
Markus				6,0	2,5			2,5	1,5	2,0	1,0	1,5	3,3
Lisa	1,5	2,0	1,0	1,5	2,5			2,5				6,0	3,3

Erwartbar ergeben sich bei den beiden Berechnungsstrategien wieder recht unterschiedliche Gesamtnoten. Allerdings wird nicht erreicht, dass die auf unterschiedlichen Anforderungsstufen erbrachten Leistungen der beiden Schüler auch zu einer unterschiedlichen Gesamtbewertung führen. Um diese Differenzierung in der Notenbildung wirksam werden zu lassen, müssen *unterschiedliche Gewichtungen* für die Anforderungsbereiche eingeführt werden, womit eine Strategiewahl bei der Grundentscheidung (II) ansteht.

Es soll nun der Entscheidung der Lehrperson nicht vorweggenommen werden, welche Gewichtungen für die einzelnen Anforderungsbereiche einer Taxonomie als angemessen anzusehen sind; dass aber eine Differenzierung in der Gesamtleistung durch unterschiedliche Gewichtungen zu erreichen ist, kann leicht am obigen Beispiel gezeigt werden. Nehmen wir beispielsweise an, dass Leistungen im Anforderungsbereich I einfach, im Anforderungsbereich II zweifach und im Anforderungsbereich III dreifach gewichtet werden, dann liefert die Notentabelle (Tabelle 3)

unter Verwendung der Einzelleistungsgewichtung

$$\text{Markus: } \bar{N} = \frac{1 \cdot 6,0 + 2 \cdot 2,5 + 3 \cdot (1,5 + 2,0 + 1,0)}{1 \cdot 1 + 2 \cdot 1 + 3 \cdot 3} = \frac{24,5}{12} = 2,0$$

$$\text{Lisa: } \bar{N} = \frac{1 \cdot (1,5 + 2,0 + 1,0) + 2 \cdot 2,5 + 3 \cdot 6,0}{1 \cdot 3 + 2 \cdot 1 + 3 \cdot 1} = \frac{27,5}{8} = 3,4$$

- unter Verwendung der Kategoriengewichtung

$$\text{Markus: } \bar{N} = \frac{1 \cdot 6,0 + 2 \cdot 2,5 + 3 \cdot 1,5}{1 + 2 + 3} = \frac{11,0}{6} = 1,8$$

$$\text{Lisa: } \bar{N} = \frac{1 \cdot 1,5 + 2 \cdot 2,5 + 3 \cdot 6,0}{1 + 2 + 3} = \frac{24,5}{6} = 4,0 .$$

Die unterschiedliche Gesamtbewertung ist dramatisch – ibs. im Falle der Kategoriengewichtung. Immerhin wird aber nun deutlich, dass nur durch eine unterschiedliche Gewichtung von Leistungen auf unterschiedlichen Anforderungsstufen die gewünschte Differenzierung in der Leistungsbewertung erreicht werden kann.

Eignet sich zur differenzierten Leistungsbewertung auf der Grundlage eines Taxonomiemodells jede Gewichte-Verteilung? Nein. Tatsächlich kann hinter einer ungünstig gewählten Gewichtung die Differenzierung von Leistungen nach unterschiedlichen Anforderungsniveaus auch wieder verschwinden. Würde man in den Notenbildern der Tabellen 4 und 5 etwa die Gewichtungen 3, 1 und 3 für die Anforderungsstufen 1, 2 und 3 wählen, dann würden die Unterschiede sowohl bei der Einzelnotengewichtung (s.o. Möglichkeit 1) als auch bei Kategoriengewichtung (s.o. Möglichkeit 2) verschwinden. Leicht nachweisbar ist eine solche Einebnung oder Kompensation von Leistungen auf unterschiedlichen Taxonomiestufen vermeidbar durch die Forderung, dass die Gewichtungsfaktoren eine streng monoton steigende oder streng monoton fallende Folge entsprechend der Taxonomiestufen bilden; diese Folge muss allerdings nicht linear steigend oder fallend sein. Wir halten das Resultat unserer Betrachtungen in folgender Aussage fest:

Leistungsbewertung in Taxonomiemodellen

Sollen Leistungen auf unterschiedlichen Anforderungsstufen eines Taxonomiemodells in einer durch Notenintegration gemäß dem Modell 3.2.1 ermittelten Gesamtbewertung sichtbar werden, so müssen die Leistungen auf allen unterschiedlichen Anforderungsstufen des Taxonomiemodells in die Bewertung einbezogen und mit einer streng monotonen Folge von Gewichtungsfaktoren gewichtet werden.

5.5 Zusammenfassung und Schlussfolgerung

Die Notenbildung auf der Grundlage von Leistungstaxonomien mit unterschiedlichen Anforderungsstufen bedarf einer besonderen Aufmerksamkeit, um die auf unterschiedlichen Anforderungsstufen erbrachten Leistungen differenziert in die Berechnung der Gesamtnote einzubringen. Die Überlegungen fasse ich in folgenden Resultaten zusammen:

- Leistungen auf unterschiedlichen Anforderungsstufen müssen entsprechenden Kategorien zugeordnet werden.
- Die den Anforderungsstufen entsprechenden Kategorien müssen auf der Grundlage des Konjunktiven Modells in die Notenbildung einbezogen werden, d.h. Leistungen auf den

unterschiedlichen Anforderungsstufen sind als Pflichtleistungen zu bewerten. Dabei muss jeder Schüler Gelegenheit erhalten, solche Leistungen zu erbringen. Nicht erbrachte oder verweigerte Leistungen sind nach geltender Regelung mit entsprechender Note zu bewerten.

- Die Leistungsdifferenzierung auf den unterschiedlichen Anforderungsstufen wird nur sichtbar, wenn die entsprechenden Kategorien bzw. die darin befindlichen Einzelleistungen mit unterschiedlicher Gewichtung in die Berechnung der Gesamtnote eingebracht werden, wobei eine streng monotone Folge von Gewichtungsfaktoren entsprechend der Abfolge der Taxonomiestufen unbedingt zu empfehlen ist, um Einebnungseffekten vorzubeugen. Ob für die Gewichtung eine Kategorien- oder Einzelleistungsgewichtung zu verwenden ist, sollte unter Bezug auf gewichtige Gründe entschieden werden.

6 Kleines Fazit

So scheinbar einfach sich in der schulischen Praxis eine Gesamtbeurteilung auf der Grundlage des Durchschnittes von Einzelnoten berechnen lässt, so tückisch, in Einzelfällen geradezu dramatisch können sich kleine Abweichungen vom „richtigen Weg“ in der Notenbildung zur Geltung bringen. Im vorliegenden Papier sollte aufgezeigt werden, wie sensibel sich alltägliche Entscheidungen über die Zuordnung von Einzelleistungen zu einer Kategorie, unterschiedliche Arten der Berechnung gewichteter Durchschnitte, verschiedene Strategien der wahlweisen oder pflichthaften Einbeziehung bestimmter Einzelleistungen in die Gesamtbeurteilung oder der Gesamtbeurteilung auf der Grundlage taxonomisch geordneter Einzelleistungen in der Notenbildung auswirken können. Ich hoffe, in diesen genannten Aspekten der Beurteilungspraxis mit dem vorliegenden Papier ein wenig zur Verfahrensklärung beigetragen zu haben. Viele weitere Probleme in der schulischen Beurteilungspraxis bleiben sicherlich bestehen.

Literatur

- Amelang, M., & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer Medizin.
- Anderson, L. W. & Krathwohl, D. R. (Hrsg.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, handbook I: Cognitive domain*. New York: Longmans, Green.
- Bohl, T. (2005). *Prüfen und Bewerten im Offenen Unterricht*. Weinheim: Beltz.
- Bohl, T., & Kucharz, D. (2010). *Offener Unterricht heute*. Weinheim: Beltz.
- Bortz, J. (1979). *Lehrbuch der Statistik. Für Sozialwissenschaftlicher*. Berlin: Springer.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden*. Weinheim, Basel: Beltz.
- Jürgens, E. (2000). *Leistung und Beurteilung in der Schule*. Sankt Augustin: Academia Verlag.
- Jürgens, E., & Lissmann, U. (2015). *Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule*. Weinheim: Beltz.
- Paradies, L., Wester, F., & Greving, J. (2005). *Leistungsmessung und –bewertung*. Berlin: Cornelsen Scriptor. S. 51-53
- Lehrerfreund (2017). *Punkte-Noten-Rechner (mit Knick)*. Abgerufen von <http://www.lehrerfreund.de/schule/1s/notenschluesselrechner/3608> [3.6.2017]