

# Leistungen messen, beurteilen und ausweisen

## Situationsanalyse



Vera Husfeldt, Matthias Gut, Frank D. Bauer

Institut Forschung und Entwicklung  
Zentrum Bildungsorganisation und Schulqualität  
im Auftrag des Departements Bildung Kultur und Sport (BKS) des Kantons Aargau

Aarau, September 2008

---

<b>1.</b>	<b>Einleitung</b>	<b>3</b>
<b>2.</b>	<b>Situationsanalyse im Kanton Aargau</b>	<b>6</b>
<b>3.</b>	<b>Häufigkeit von Klassenprüfungen</b>	<b>8</b>
<b>4.</b>	<b>Festlegung der Noten für Prüfungen und Zeugnis</b>	<b>13</b>
4.1.	Prüfungen	13
4.2.	Zeugnis	16
<hr/>		
<b>5.</b>	<b>Verwendete Messinstrumente</b>	<b>19</b>
<b>6.</b>	<b>Leistungsergebnisse als Basis für Fördermassnahmen</b>	<b>24</b>
<b>7.</b>	<b>Zusammenarbeit im Bereich Leistungsbeurteilung</b>	<b>28</b>
<b>8.</b>	<b>Probleme und Herausforderungen der Leistungsbeurteilung</b>	<b>30</b>
8.1.	Probleme	30
8.2.	Herausforderungen	31
<hr/>		
<b>9.</b>	<b>Unterstützungsmassnahmen</b>	<b>33</b>
9.1.	Unterstützung seitens der Schulleitung	33
9.2.	Unterstützung seitens des BKS	33
<hr/>		
<b>10.</b>	<b>Zusammenfassung</b>	<b>35</b>
<b>11.</b>	<b>Literatur</b>	<b>37</b>
<b>12.</b>	<b>Abbildungsverzeichnis</b>	<b>38</b>
<b>13.</b>	<b>Tabellenverzeichnis</b>	<b>39</b>

# 1. Einleitung

Zur Bewertung und Ausweisung einer Schülerleistung sind eine Reihe von differenzierten Überlegungen notwendig, die in einem schrittweisen Prozess zum Tragen kommen. Die Leistung muss zunächst abgerufen werden, in einem nächsten Schritt gemessen und danach beurteilt werden, bevor eine Bewertung und ein Ausweisen der Leistungen geschehen kann. Abgerufen wird die Leistung in der Regel durch eine schriftliche oder mündliche Prüfung oder durch Impulse zur Mitarbeit im Unterricht.

Die *Leistungsmessung* beinhaltet das quantitative bzw. qualitative Feststellen der Leistung durch Codierung und Skalierung. Beim Codieren wird festgestellt, ob eine Aufgabe richtig oder falsch bzw. teilweise richtig gelöst wurde. Dieses Codieren ist beispielsweise bei Multiple-Choice-Aufgaben eines standardisierten Tests recht leicht zu bewältigen, während die Codierung z.B. freier Texte und Schreibproben wesentlich aufwändiger und schwieriger ist. Skalierung bedeutet, dass Aufgaben, die zum gleichen Kompetenzbereich gehören, gemeinsam durch einen Kennwert abgebildet werden. Es kann beispielsweise eine (Summen)-Skala für Leistungen in Geometrie abgebildet werden, indem die korrekten Antworten in einem Geometrietest ausgezählt werden. Die Summe der korrekten Antworten kann dann als Index für die Leistung in Geometrie verwendet werden.

Die *Leistungsbeurteilung* erfolgt im Abgleich zu einem Referenzsystem, das sich an der kriterialen, der sozialen oder der individuellen Bezugsnorm orientieren kann. Wird ein kriteriales Referenzsystem verwendet, so müssen konkrete Lerninhalte oder Kompetenzen festgelegt werden, mit denen die von den Schülerinnen und Schülern erreichten Leistungen verglichen werden können. Dies geschieht in der Regel auf der Grundlage von Leistungskriterien oder Kompetenzstufenmodellen. Beurteilungen, die auf einen kriterialen Vergleich abzielen, dienen in der Regel zur Orientierung über die Verortung der gemessenen Leistung in einem inhalts- oder kompetenzorientierten Bezugssystem. Bei der Verwendung eines sozialen Referenzsystems werden die gemessenen Leistungen eines Schülers oder einer Schülerin mit den Leistungen einer Gruppe verglichen. Problematisch ist dies, wenn die Vergleichsgruppe nicht repräsentativ ist, wenn beispielsweise die Leistungen einer sehr guten Schülerin im Vergleich zu einer noch besseren Klasse nur noch mittelmässig beurteilt werden. Vergleichsarbeiten, bei denen mehrere Klassen die gleichen Aufgaben bearbeiten und die gemittelten Leistungen mehrerer Klassen miteinander verglichen werden, können hilfreich sein, das Niveau einer einzelnen Klasse besser einschätzen zu können. Bei Verwendung der individuellen Bezugsnorm wird die gemessene Leistung einer Schülerin oder eines Schülers mit der zu einem früheren Zeitpunkt gemessenen Leistung derselben Person verglichen. Der Vergleich anhand der individuellen Bezugsnorm ermöglicht Aussagen über die Lernentwicklung.

Erst im letzten Schritt nach der Leistungsbeurteilung kann eine Leistungsbewertung vorgenommen werden. Im Rahmen der Leistungsbewertung wird festgelegt, ob eine gemessene Leistung ausreichend ist, bzw. es wird die Leistung in einem Benotungssystem verortet. Erst in diesem Schritt geht es darum, die gemessene Leistung mit den in der Regel normativ gesetzten Standards abzugleichen. Beispielsweise ermöglicht die Leistungsbewertung Aussagen darüber, ob eine gemessene Leistung die Kriterien eines Minimal-, Durchschnitts- oder Maximalstandards erreicht. Häufig werden die Begriffe Leistungsmessung und Bildungsstandards eng zusammengebracht, obwohl die Leistungsmessung unabhängig von der Festlegung der Standards ist. Nur die Bewertung

der Ergebnisse aus der Leistungsmessung orientiert sich an den Standards. Das *Ausweisen* der Leistung in Zeugnissen, Lernberichten etc. erfolgt auf der Grundlage der Leistungsbewertung.

Auch wenn die Diagnostik ein vielbeachteter Bereich in der Pädagogik ist und zu den wichtigsten Themen der Lehrerausbildung gehört, zeigt sich in der Praxis häufig, dass Beurteilungen und Bewertungen von verschiedenen Lehrpersonen in unterschiedlicher Weise vorgenommen werden.

Dass Beurteilungen bei Aufsätzen je nach Beurteilerin oder Beurteiler unterschiedlich ausfallen können, wurde durch zahlreiche Untersuchungen nachgewiesen (z.B. Coffman, 1971; Eberle & Metzger, 1988). Dass das gleiche Problem in ähnlichem Ausmass aber auch bei Mathematikarbeiten auftritt, ist zunächst vielleicht etwas überraschender, bei näherer Betrachtung jedoch ebenfalls nachvollziehbar. Karlheinz Ingenkamp verweist bereits 1971 in seinem vielbeachteten Buch "Die Fragwürdigkeit der Zensurengebung" auf diese Problematik, die nicht an einzelne Fächer gebunden zu sein scheint.

Häufig kommen nicht nur die Richtigkeit von Lösungen oder die Kompetenz in einem bestimmten Bereich, sondern eine Reihe von anderen Aspekten bei einer Beurteilung zum Tragen, die stark von der Persönlichkeit der beurteilenden Lehrperson abhängen. Die Objektivität von Beurteilungen ist ein Thema, dem bereits am Anfang des letzten Jahrhunderts grosse Bedeutung zukam. So führt beispielsweise Daniel Starch bereits 1913 verschiedene Ursachen für die Variabilität von Beurteilungen auf: 1. Unterschiede in den Anforderungen verschiedener Schulen; 2. Unterschiede im Ausbildungsstand verschiedener Lehrpersonen; 3. Unterschiedliche Betonung des relativen Wertes verschiedener Elemente in Prüfungsarbeiten durch verschiedene Lehrpersonen, einschliesslich Inhalt und Form; 4. Unterschiede, die auf blosse Unfähigkeit zurückzuführen seien, zwischen eng verwandten Gütegraden zu differenzieren. Ingenkamp (1971) problematisiert zusätzlich das klasseninterne Bezugssystem und eine weit gehende Subjektivität, die unter anderem dazu führt, dass, so Ingenkamp, der stärkste Faktor für Erfolg oder Misserfolg die Schichtzugehörigkeit der Schülerinnen und Schüler ist.

Zu ähnlichen Schlüssen kommen auch neuere Studien. Beispielsweise benennt Jörg Ziegenspeck (1999) erneut die subjektiven Fehlerquellen wie Halo-Effekte, Beharrlichkeitstendenzen, Reihungseffekte, Kontrasteffekte, Beurteilungstendenzen und Abschätzung von Konsequenzen. Insbesondere wird, wie schon bei Ingenkamp, das klasseninterne Bezugssystem als problematisch dargestellt, wenn es darum geht Leistungen klassen- und schulübergreifend valide zu bewerten. Weiterhin stellt Ziegenspeck aber auch fächerspezifische Unterschiede fest, die dadurch beeinflusst werden, wie stark, sich die Notengebung auf schriftliche, quantifizierbare Leistungen stützt und welchen Stellenwert verbale Anforderungen einnehmen.

Für die Schweiz kann Urs Vögeli-Mantovani in dem SKBF Trendbericht "Mehr fördern, weniger auslesen: Zur Entwicklung der schulischen Beurteilung in der Schweiz" (1999) ebenfalls bestätigen, dass die Bewertung von Leistungen in der Schule problematisch ist. Er zählt dafür sechs Gründe auf: 1. Verschiedene Lehrpersonen bewerten dieselbe Arbeit unterschiedlich. 2. Die Lehrpersonen beurteilen dieselbe Arbeit zu unterschiedlichen Zeitpunkten unterschiedlich. 3. Die Beurteilungskriterien sind häufig nicht transparent. 4. Es gibt eine Reihe von unerwünschten Nebeneffekten. 5. Noten sind zur Beurteilung bestimmter Sachverhalte unzulässig. 6. Notarithmetik ist mathematisch unzulässig. Vögeli-Mantovani zu Folge müssen die Noten zu viele Funktionen erfüllen, die jede

für sich genommen, zu wenig konkret definiert sind. Die Beurteilungsstrategien der Lehrpersonen seien ebenfalls nicht transparent. Erneut wird auf die Orientierung am Klassenbezugssystem eingegangen und herausgestellt, dass diese Orientierung sich ungünstig auf die Motivation, die Ursachenzuschreibung und die Selbstwahrnehmung der Schülerinnen und Schüler auswirken kann. Insbesondere die mathematische Durchschnittsbildung von Ziffernnoten wird als problematisch und trügerisch hervorgehoben.

Vor dem Hintergrund der Ergebnisse solcher empirischer Untersuchungen stellen Hans Brügelmann et al. (2006) in einer wissenschaftlichen Expertise im Auftrag des Grundschulverbandes unterschiedliche Schlussfolgerungen an, die sich ihres Erachtens für eine differenziertere Betrachtung der Bewertungen in den Schulen aussprechen. Im einzelnen werden folgende Punkte hervorgehoben:

- Ziffernnoten sind durch differenziertere Formen der Dokumentation und der Bewertung von Leistung zu ersetzen. Rückmeldungen und Bewertungen sollten dabei getrennt werden. Der Leistungsstand sollte bezogen auf konkrete Lernziele und die individuelle Entwicklung dargestellt werden.
- Leistungen sollten nicht nur bewertet, sondern differenziert beschrieben und schriftlich oder mündlich kommentiert werden.
- Vergleichende Leistungsbeurteilung sollte nur bezogen auf klassenübergreifende Stichproben z.B. zentrale Lernstandserhebungen angewendet werden. Diese sollten aber nicht zum heimlichen Lehrplan werden.
- Leistungsbewertung sollte sich stärker an den Lernzielen und am individuellen Lernfortschritt orientieren und nicht an der Sozialnorm.
- Leistungen sollten zu mehreren Zeitpunkten und in unterschiedlichen Aufgaben/Situationen erfasst werden. Eine Leistungsbewertung sollte von mehreren involvierten Personen vorgenommen werden.
- Es braucht eine Sensibilisierung der Lehrpersonen für die Schwierigkeiten beim Beurteilen und für die Erwartungen der Zielgruppen. Eine gezielte Aus-/Weiterbildung kann dieser Forderung Rechnung tragen und die Bewertungskompetenzen der Lehrpersonen stärken. Zusätzlich sind fachdidaktisch begründete Kriterien und ein kontinuierlicher kollegialer Austausch notwendig.
- Zunehmend sollte eine dialogische Form einer Verbindung von Selbst- und Fremdbeurteilungen erprobt werden.
- Der Förderauftrag muss Vorrang vor der Selektion haben.

Bei einzelnen dieser Schlussfolgerungen bleibt allerdings zu diskutieren, wie realistisch im Schulalltag damit umgegangen werden kann, inwiefern sie im Einklang mit der geltenden Promotionsverordnung stehen und ob sinnvollerweise auf sie zurückgegriffen werden sollte. Beispielsweise ist die Bewertung mit Ziffernnoten eine politische Vorgabe, die nicht auf der Ebene der Lehrpersonen verhandelbar ist.

## 2. Situationsanalyse im Kanton Aargau

Das Departement Bildung, Kultur und Sport (BKS) des Kantons Aargau hat das hier dargestellte Projekt mit dem Ziel aufgelegt, eine Situationsanalyse der Beurteilungspraxis von Aargauischen Lehrpersonen zu erstellen und auf dieser Grundlage Handlungsschritte zu erarbeiten, die die Lehrpersonen in ihrer Beurteilungspraxis zusätzlich unterstützen können. Es wird insbesondere der Frage nachgegangen, wie Lehrpersonen die Leistungen ihrer Schülerinnen und Schüler *messen, beurteilen* und *ausweisen*. Folgende Leitfragen sind zielführend für das hier dargestellte Projekt:

- Wie oft werden Klassenprüfungen durchgeführt?
- Wie legen die Lehrpersonen die Noten für die Prüfungen und das Schulzeugnis fest?
- Auf welche Messinstrumente greifen die Lehrpersonen neben den Klassenprüfungen zurück?
- Inwiefern werden die Leistungsergebnisse als Ausgangslage zu weiteren Fördermassnahmen genutzt?
- Wie findet die Zusammenarbeit innerhalb der Schulen in diesem Bereich statt?
- Wo sehen die Lehrpersonen die Probleme und Herausforderungen im Rahmen ihrer Leistungsbeurteilung?
- Welche Unterstützungsmassnahmen sind aus der Sicht der Lehrpersonen hilfreich?

Zur Bearbeitung dieser Fragestellungen wurde ein hypothesengeleiteter Fragebogen entwickelt, der explizit auf die Sichtweise der Lehrpersonen eingeht. Folgende Erwartungen und Hypothesen zu den Ergebnissen wurden im Vorfeld der Untersuchung zu den einzelnen Forschungsfragen formuliert:

- Die Promotionsverordnung sieht eine Mindestanzahl von Prüfungen vor, die der Anzahl der Wochenlektionen entsprechen. Wenn nur eine Lektion pro Woche erteilt wird, so sind mindestens 2 Prüfungen vorgesehen. Eine Hypothese lautet deshalb, dass die Lehrpersonen sich bei der Festlegung der Anzahl von Prüfungen an dieser Vorgabe orientieren, bzw. grob an einer Formel  $\text{Wochenlektionen} + 1$ .
- In der Literatur wird häufig dargestellt, dass Lehrpersonen in ihrer Beurteilungs- und Bewertungspraxis ein starkes Gewicht auf quantifizierbare und klassenintern vergleichbare Leistungsbeurteilungen legen (vgl. Kapitel 1). Aus diesem Grunde wird als Hypothese formuliert, dass die Noten für Prüfungen und Zeugnisse sich häufig an der sozialen Norm orientieren und aus mathematischen Berechnungen des Notendurchschnitts der Klassenprüfungen ergeben.
- Es wird aufgrund der dargestellten Literatur weiterhin erwartet, dass die Beurteilung der Leistungen zum Zeitpunkt der Erhebung zu grossen Teilen auf schriftlichen Klassenprüfungen und anderen schriftlichen Leistungen beruht.
- Lehrpersonen messen den Leistungsergebnissen vermutlich vor allem mit Blick auf die individuelle Förderung der Schülerinnen und Schüler ein hohes Gewicht bei.
- Die Forderung nach einer stärkeren kollegialen Zusammenarbeit innerhalb der Schulen wird in der Literatur stark betont. Es wird deshalb erwartet, dass auch diese Untersuchung bestätigt, dass die Kooperation in der Frage des Messens, Beurteilens und Ausweisens von Leistung noch optimierbar ist.

- Die Objektivität von Leistungsmessungen und das damit verbundene Gewicht der Verantwortung ist in der Literatur stark gewichtet (vgl. Kapitel 1) und wird erwartungsgemäss von den Lehrpersonen ebenso wahrgenommen.
- Die Wünsche der Lehrpersonen zu Unterstützungsmassnahmen beziehen sich erwartungsgemäss zum grossen Teil auf Angebote standardisierter und halbstandardisierter Instrumente zur Leistungsmessung, auf klare und verbindliche Richtlinien in unterschiedlichen Bereichen sowie auf Fort- und Weiterbildungen zum Thema.

Es wurden im Februar 2008 mittels eines Online-Fragebogens die Lehrpersonen jener Klassen befragt, die unmittelbar vor dem Übertritt in eine weiterführende Schulstufe stehen. Dies sind die Lehrpersonen der 5. und 9. Klassen. Eine Einladung zur Teilnahme an der Befragung haben alle Lehrpersonen dieser Klassen im Kanton Aargau über ihre Schulleitung bekommen. Insgesamt sind 251 Lehrpersonen auf die Einladung zur Teilnahme an der Befragung eingegangen, wobei die Verteilung von Frauen und Männern mit 50,4 Prozent und 49,6 Prozent so gut wie ausgewogen ist. Von allen Befragten unterrichten im ersten Semester des Schuljahres 2007/2008 39 Prozent in der 5. Klasse und 61 Prozent in der 9. Klasse. Davon gaben 82 Prozent der Lehrerinnen und Lehrer an, Klassenlehrperson zu sein. Die Aufteilung der befragten Lehrpersonen von 9. Klassen nach Schulstufe zeigt, dass der Grossteil dieser Befragungsteilnehmenden in der Bezirksschule unterrichtet, gefolgt von der Sekundarschule und der Realschule (vgl. Tabelle 1).

**Tabelle 1 (Schulstufe/ Schultyp, an der die Befragten im 1. Semester 2007/2008 hauptsächlich unterrichten.)**

<b>Schultyp</b>	<b>Prozent</b>
Primarschule	38,2
Kleinklasse Primarstufe	0,4
Bezirksschule	23,1
Sekundarschule	21,1
Realschule	13,5
Kleinklasse Oberstufe	2,4
Berufswahljahr	1,2
<b>Gesamt</b>	<b>100,0</b>

Der Mittelwert der Anzahl der Arbeitsjahre im Schuldienst beträgt 18,5 Jahre (SD = 11,9), wobei die Spannweite der Arbeitsjahre von einem Jahr bis zu 44 Jahren reicht. Das Durchschnittsalter der befragten Lehrerinnen und Lehrer liegt bei 44,5 Jahren. Die älteste Lehrperson ist 64 und die jüngste 23 Jahre alt. Weiterhin wurden die Befragten gebeten, anzugeben, mit wie vielen Stellenprozenten sie unterrichten. Im Durchschnitt sind es ca. 80 Stellenprozent. Ein signifikanter Zusammenhang zwischen dem Arbeitspensum nach Stellenprozenten und den Dienstjahren der Lehrpersonen besteht nicht.

In den folgenden Abschnitten werden die Ergebnisse der statistischen Datenanalysen dargestellt, indem jeweils auf die oben genannten Leitfragen eingegangen wird.

### 3. Häufigkeit von Klassenprüfungen

Unter Klassenprüfungen wurden in der vorliegenden Untersuchung alle zeugnisrelevanten schriftlichen, mündlichen oder praktischen Prüfungen zusammengefasst, an denen die Schülerinnen und Schüler teilnehmen. Betrachtet man für beide Klassenstufen die Unterrichtsfächer mit den häufigsten Klassenprüfungen, so ist eine Schwerpunktsetzung auf die Sprachenfächer sowie auf Mathematik erkennbar.

In der 5. Klasse sind Prüfungen in der Regel in den Fächern Mathematik, Deutsch, Realien, Bildnerisches Gestalten, Werken, Musik und Sport üblich<sup>1</sup>. Die 68 Lehrpersonen, die in der 5. Klasse Ethik und Religion unterrichten, führen hingegen keine Prüfungen durch. Am häufigsten wird im Fach Deutsch geprüft (vgl. Tabelle 2), im Durchschnitt sind es hier in der 5. Klasse 12,1 Prüfungen. Es zeigt sich im Fach Deutsch jedoch auch eine im Gegensatz zu den anderen Fächern hohe Standardabweichung von 4,7. Das heisst, dass die Angaben nicht sehr homogen ausfallen, sondern die Anzahl der Deutschprüfungen je nach Lehrperson relativ stark schwankt. Tatsächlich führen 88 Prozent der befragten Deutschlehrpersonen in den 5. Klassen mehr als 8 Prüfungen pro Semester durch (vgl. auch Abbildung 1).

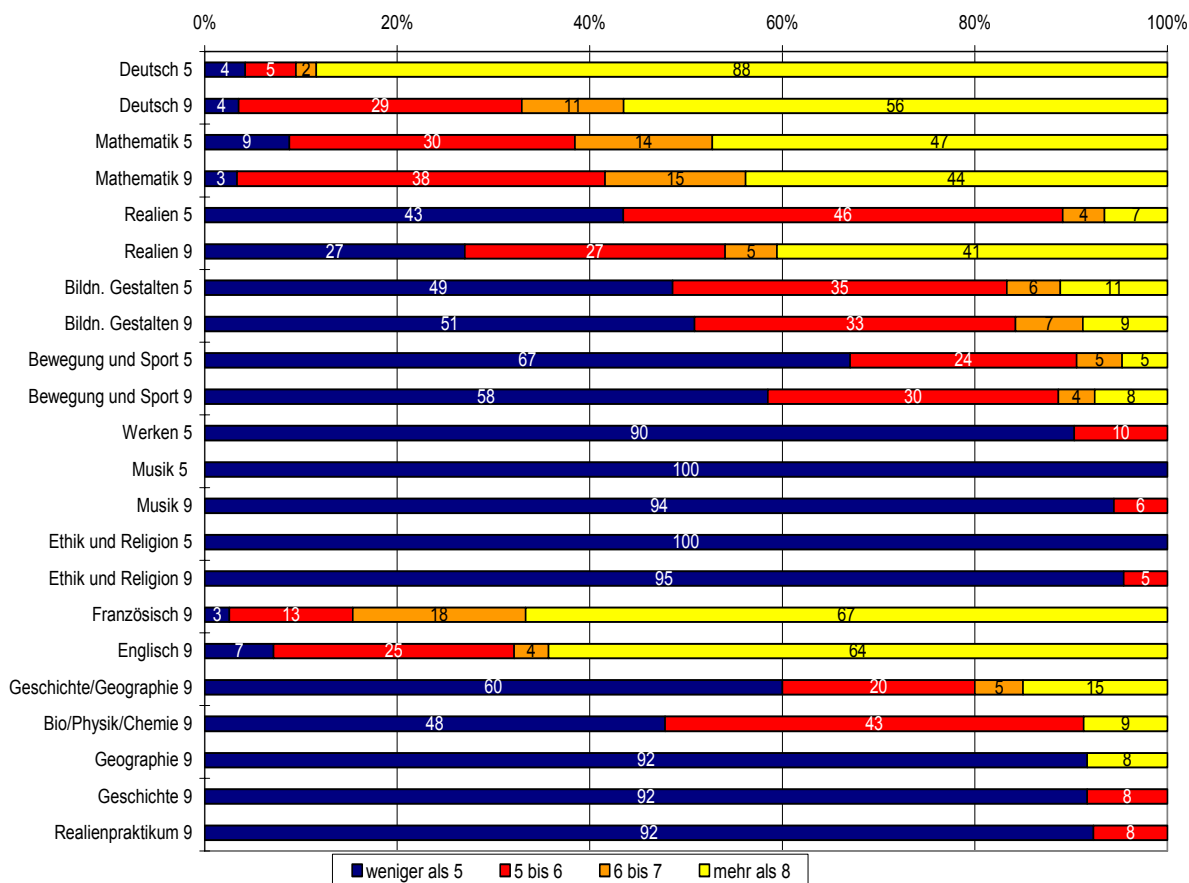
**Tabelle 2 (Durchschnittliche Anzahl Klassenprüfungen pro Semester nach Unterrichtsfächern in der 5. Klasse und in der 9. Klasse. Anzahl der Lehrpersonen, Mittelwert und Standardabweichung, Mindestanzahl Prüfungen laut Promotionsverordnung)**

Unterrichtsfach	Anzahl LP 5. Klasse	Mittelwert Anzahl Klassenprüfungen 5. Klasse (Standardabweichung)	Mindestanzahl Prüfungen laut Promotionsordnung 5. Klasse	Anzahl LP 9. Klasse	Mittelwert Anzahl Klassenprüfungen 9. Klasse (Standardabweichung)	Mindestanzahl Prüfungen laut Promotionsordnung 9. Klasse <sup>2</sup>
Deutsch	95	12,1 (4,7)	5-6	85	8,9 (4,6)	5
Mathematik	91	7,4 (2,4)	5	89	8 (3,9)	5
Realien	92	4,9 (1,5)	4	37	7,6 (4,5)	7
Bildnerisches Gestalten	72	4,4 (3,2)	2	57	4,4 (2,6)	2
Bewegung und Sport	85	3,8 (3,2)	3	53	4,2 (2,3)	3
Werken	62	2,5 (1,6)	2			
Musik	87	1,7 (1,4)	2	36	2,2 (1,6)	2
Ethik und Religion	68	0,0 (0,0)	2	22	0,6 (1,3)	2
Französisch				39	8,8 (2,7)	3
Englisch				28	7,8 (2,9)	3
Geschichte/Geographie				40	4,6 (2,1)	5
Biologie/Physik/Chemie				23	4,3 (1,9)	3
Geographie				12	2,9 (2,6)	2
Geschichte				12	2,8 (1,1)	2
Realienpraktikum				26	1,1 (1,8)	2

<sup>1</sup> Fächer, die von weniger als 10 Personen angeführt wurden, werden hier nicht aufgelistet.

<sup>2</sup> Für die 9. Klasse unterscheiden sich diese Vorgaben im Bereich der Realien. Es sind hier die Vorgaben der Realschule für das Fach Realien, die der Sekundarschule für Geschichte/Geographie und Biologie/Physik/Chemie, die der Bezirksschule für Geschichte und Geographie verwendet worden. (Da Geographie laut Stundentafel in der 4. Klasse der Bezirksschule nicht regulär unterrichtet wird, wurden in diesem Fall die Angaben aus der 3. Klasse verwendet.)

Das bedeutet, dass in diesen Klassen mindestens in jeder zweiten Woche eine Prüfung stattfindet. Mehr als 10 Prozent der befragten Deutschlehrpersonen der 5. Klassen geben sogar an, mehr als 16 Prüfungen, also ca. eine Prüfung pro Woche durchzuführen. Im Fach Mathematik kommen im Durchschnitt 7,4 Prüfungen zum Einsatz, wobei hier die Schwankungsbreite niedriger ist als im Fach Deutsch. (vgl. Tabelle 2 und Abbildung 1).



**Abbildung 1 (Anzahl Klassenprüfungen pro Semester nach Unterrichtsfächern in der 5. Klasse und in der 9. Klasse in Prozent)**

In der 9. Klasse liegt die Anzahl der Prüfungen in den Fächern Deutsch, Mathematik und Realien genau wie in der 5. Klasse recht hoch. Es kommen hier aber auch noch die Fremdsprachen dazu, die ebenfalls häufig geprüft werden. Auffällig ist, dass in Italienisch und Latein (beide weniger als 10 Nennungen) weniger häufig geprüft wird als in Englisch und Französisch, was darauf hindeutet, dass die Begründung der Prüfungshäufigkeit nicht in der Materie des Faches, sondern in seiner Bedeutung als Haupt- oder Nebenfach liegt. In den Promotionsfächern scheinen Prüfungen am häufigsten stattzufinden.

Die Promotionsverordnung sieht vor, dass die Anzahl der Prüfungen mindestens der Anzahl der Wochenlektionen entsprechen sollte. In Fällen, in denen nur eine Wochenlektion unterrichtet wird, sollen mindestens 2 Prüfungen durchgeführt werden. In Tabelle 2 ist die nach dieser Vorgabe berechnete Mindestanzahl angegeben. Es fällt auf, dass im Durchschnitt in fast allen Fächern die laut Promotionsverordnung vorgegebene Anzahl von Prüfungen als Mindestmass eingehalten wird. Lediglich in Musik, Ethik und

Religion, Geschichte/Geographie und Realienpraktikum wird im Schnitt weniger geprüft, als es die Promotionsverordnung vorsieht.

Eine neue Variable, die angibt, ob eine Lehrperson mehr, genauso häufig oder weniger prüft als laut der Promotionsverordnung mindestens vorgesehen, erlaubt es, Zusammenhänge zu anderen Kontextvariablen zu analysieren. So interessiert es, ob die Klasse, die Rolle als Klassen- oder Fachlehrperson, der Stellenprozentanteil, die Jahre im Schuldienst, das Geschlecht oder das Alter im Zusammenhang mit der Festlegung der Anzahl von Prüfungen eine Rolle spielen. Es zeigt sich, dass über alle Fächer hinweg hinsichtlich der genannten Variablen kein Unterschied zwischen den "Mehrprüfenden" und den "Wenigerprüfenden" besteht. Bei der Betrachtung einzelner Fächer wird jedoch deutlich, dass signifikante Effekte im Bereich der am stärksten vertretenen Fächer Deutsch und Mathematik auftreten. Im Fach Deutsch sind die "Mehrprüfenden" signifikant häufiger Lehrpersonen aus den 5. Klassen als Lehrpersonen aus den 9. Klassen und ebenfalls signifikant häufiger Klassenlehrpersonen als Fachlehrpersonen und häufiger weiblich als männlich. Sie haben weiterhin einen signifikant höheren Stellenprozentanteil und eine höhere Anzahl von Dienstjahren. In Fach Mathematik lassen sich diese Zusammenhänge nur zum Stellenprozentanteil und zum Dienstalter nachweisen, wobei auch hier ein hoher Stellenprozentanteil und hohes Dienstalter tendenziell mit häufigeren Prüfungen einhergehen.

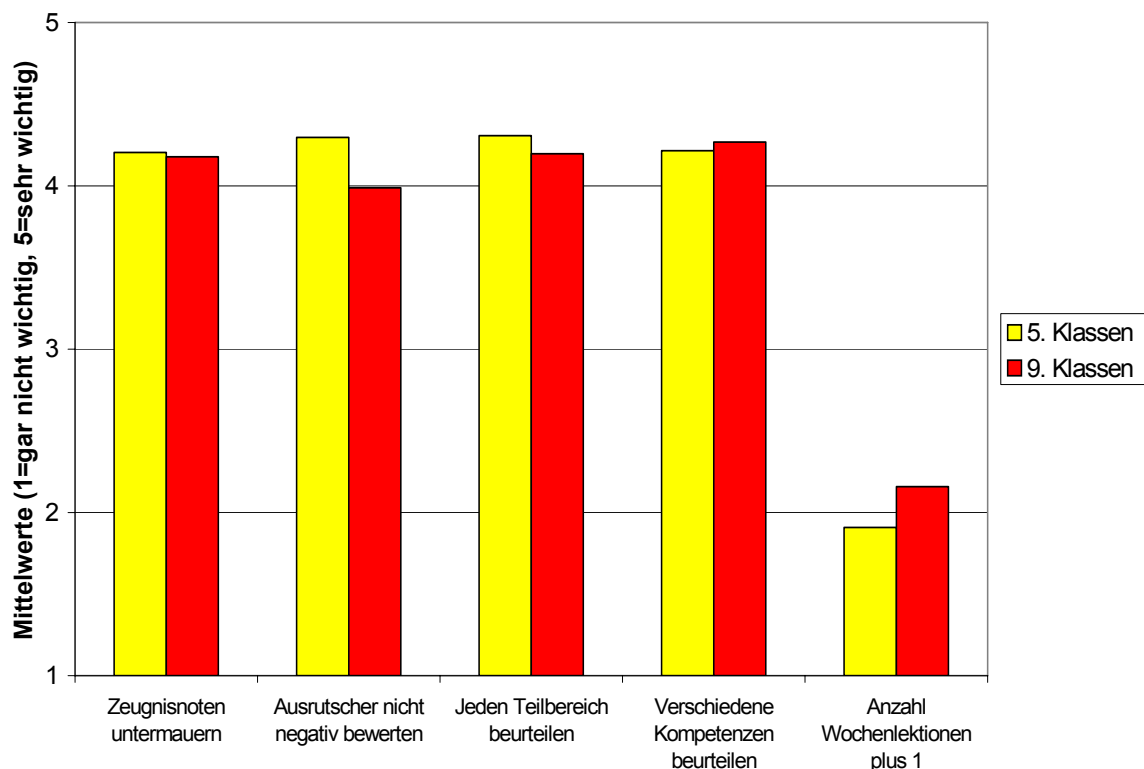
Bei der hohen Spannweite der Prüfungshäufigkeit über die Fächer stellt sich die Frage nach dem Grund, den die Lehrpersonen für die Festlegung der Anzahl von Prüfungen anführen. Nach den Schlussfolgerungen aus der Expertise von Brügelmann et al. (2006) sollten die Leistungen zu mehreren Zeitpunkten und in unterschiedlichen Aufgaben oder Situationen erfasst werden. Es ist die Aufgabe der Lehrpersonen das richtige Mass dabei zu finden, um einerseits genügend Informationen für eine möglichst objektive Bewertung zu erhalten und andererseits die Schülerinnen und Schüler nicht zu überfordern und ihnen neben der Prüfungszeit auch genügend Lernzeit zu lassen.

Welche Kriterien die Lehrpersonen zur Festlegung der Anzahl ihrer Prüfungen innerhalb eines Semesters heranziehen, macht Abbildung 2 deutlich. Demnach wird in beiden Klassenstufen, entgegen unserer ursprünglichen Hypothese, kaum von einer Formel *Anzahl Wochenlektionen plus 1* Gebrauch gemacht. Von rund 50 Prozent der Lehrpersonen beider Klassen wird sie auf einer fünfstufigen Antwortskala (1 = ist mir gar nicht wichtig bis 5 = ist mir sehr wichtig) sogar als gar nicht wichtig bewertet.

Vielmehr wird die Prüfungsanzahl unter Berücksichtigung der Bewertung unterschiedlicher Schülerkompetenzen und der verschiedenen inhaltlichen Teilbereiche eines Unterrichtsfaches festgelegt. Auch werden häufige Prüfungen als Absicherung verwendet, um zu vermeiden, dass sich einzelne Ausrutscher nicht zu negativ auswirken. Betrachtet man die Antworten zu dieser Fragestellung nach Klassenstufen getrennt, liegen die wichtigsten Einschätzungen über die Anzahl der Semesterprüfungen für die Lehrpersonen der 5. Klassenstufen mit  $M = 4,3$  ( $SD = 0,65$ ) bei der Aussage *Ich lege die Anzahl der Prüfungen so fest, dass jeder Teilbereich des Faches beurteilt werden kann* und mit  $M = 4,3$  ( $SD = 0,84$ ) bei *Ich will genügend Noten haben, um sicher zu sein, dass sich ein einzelner „Ausrutscher“ nicht zu negativ auswirkt*. Die Lehrerinnen und Lehrer der 9. Klassen legen ihre Semesterprüfungsanzahl vor allem auf der Grundlage fest, *dass verschiedene Kompetenzen der Schülerinnen und Schüler beurteilt werden können* ( $M = 4,3$ ,  $SD = 0,84$ ) und *dass jeder Teilbereich des Faches beurteilt werden kann* ( $M = 4,2$ ,  $SD = 0,90$ ). Wichtig ist den Lehrpersonen aus beiden Klassen aber auch, ge-

nügend Noten zu haben, *um die Zeugnisnoten untermauern zu können*. Die Unterschiede zwischen beiden Gruppen sind nicht signifikant.

Zusammenfassend zeigen die Aussagen der Lehrpersonen, dass die Festlegung der Anzahl von Prüfungen mit dem Bestreben verknüpft ist, eine möglichst hohe Zuverlässigkeit der Beurteilung zu erreichen. Dies entspricht der Schlussfolgerung von Brügelmann et al. (2006), die darauf hinweisen, dass Leistungen zu mehreren Zeitpunkten und in unterschiedlichen Aufgaben/Situationen erfasst werden sollte. Die Anwendung von Vorgaben oder Rezepten scheint für die Lehrpersonen eine eher untergeordnete Rolle zu spielen. Es muss jedoch festgehalten werden, dass hier lediglich die Einschätzungen der Lehrpersonen berücksichtigt wurden, die sich möglicherweise aufgrund der sozialen Erwünschtheit von der tatsächlichen Praxis unterscheiden kann, gleichzeitig aber auch die Idealvorstellung der Lehrpersonen widerspiegeln.



**Abbildung 2 (Beweggründe zur Festlegung der Anzahl von Klassenprüfungen nach Klassenstufen)**

Die Lehrpersonen konnten zusätzlich in Freitextnennungen notieren, welche weiteren Aspekte für sie zur Festlegung der Prüfungsanzahl entscheidend sind. Insgesamt 82 Personen gaben eine Antwort dazu ab. Die Aussagen wurden kategorisiert, um sie in einem Folgeschritt quantifizieren zu können. 44 Lehrpersonen betonten, eine Prüfung pro Lektion bzw. Fachthema als Lernzielkontrolle durchzuführen. 20 Personen gaben an, dass sie die Prüfungsanzahl so anpassen, dass sie den Lernfortschritt und die Kompetenzen gut aufnehmen können. Weiter war es 10 Personen wichtig, die Dosierung für die Schülerinnen und Schüler erträglich zu halten. Manchmal werden die Prüfungen hingegen gerade als Training gegen Prüfungsangst und Leistungsdruck eingesetzt (10 Nennungen). Jeweils 6 Lehrpersonen orientieren sich an einer Regelmässigkeit (z.B. einmal pro Woche) an der Motivation der Schülerinnen und Schüler oder am aktuellen Klassendurchschnitt. Auch das Pensum der Lehrperson und der Vergleich mit anderen Klassen spielen für je 4 Befragte eine Rolle. Jeweils zweimal wurden ausserdem die

folgenden Aspekte genannt: Anzahl der Prüfungen gemäss einem Beurteilungskonzept, gemäss der jeweiligen Themenvielfalt oder gemäss einer Absprache mit dem Kollegium. Ebenfalls je zwei Nennungen gab es für die Festlegung der Prüfungsanzahl entsprechend dem Bedarf bei der Anpassung der Förderplanung, bei der Hausaufgabenkontrolle, bei Selektion oder bei Feststellung der Aufarbeitung von nachzuholendem Stoff.

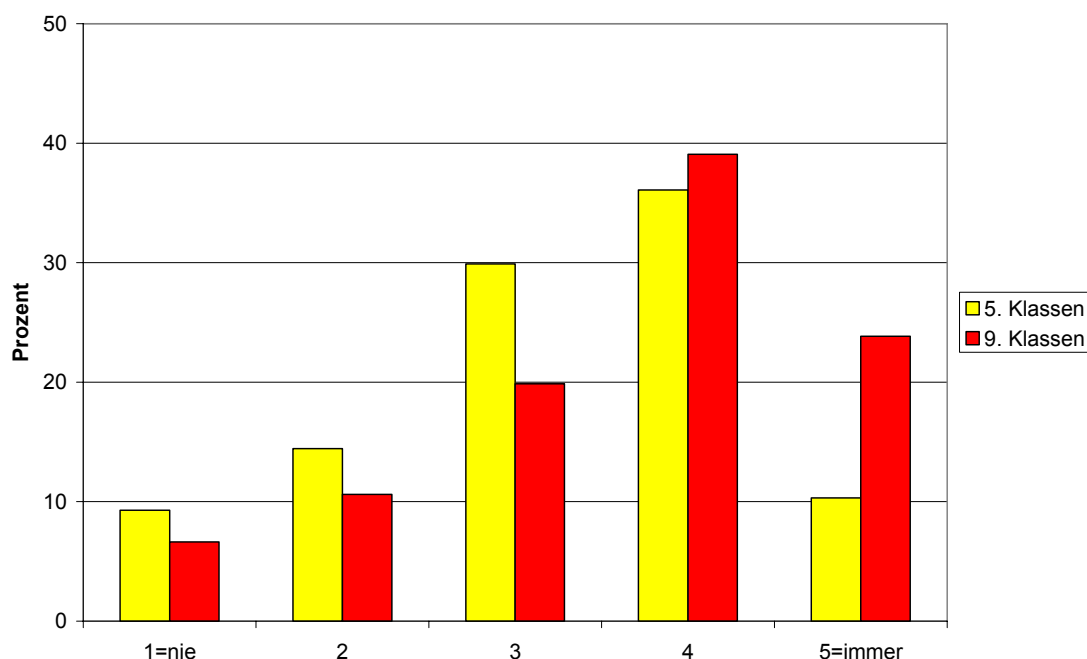
## 4. Festlegung der Noten für Prüfungen und Zeugnis

Die Frage, wie Lehrerinnen und Lehrer die Beurteilungen für das Zeugnis ihrer Schülerinnen und Schüler vornehmen, konnte anhand von insgesamt 16 Items bearbeitet werden, wobei die Fragen zu beiden Sachverhalten – Prüfungen und Zeugnisse – separat gestellt wurden, um sie ebenso getrennt auswerten zu können. Die vorgegebenen Aussagen konnten die Lehrpersonen auf einer fünfstufigen Skala von 1 = nie bis 5 = immer beantworten.

### 4.1. Prüfungen

Um festzustellen, nach welcher Bezugsnorm die befragten Lehrpersonen die Prüfungen ihrer Schülerinnen und Schüler beurteilen, wurden 10 Fragen in den Fragebogen aufgenommen, die je Aussagen zur kriterialen (4 Items), zur sozialen (3 Items) und zur individuellen Bezugsnorm (3 Items) beinhalten.

Entgegen der Hypothese, dass die Entscheidung für Prüfungsnoten im Wesentlichen auf Grund eines Vergleichs mit den Leistungen anderer Schülerinnen und Schüler aus der Klasse getroffen wird, wird aus der Analyse der Daten deutlich, dass sowohl in den 5. als auch in den 9. Klassen die Prüfungsnoten am stärksten durch eine am Kriterium orientierte Norm bestimmt werden.



**Abbildung 3 (Angaben zu der Aussage "Ich entwickle vor der Durchführung der Prüfungen ein Kriterienraster, in dem auch die Noten für die jeweiligen zu erwartenden Leistungen bereits festgelegt sind" in Prozent nach Klassenstufen.)**

Ein wesentlicher Indikator für die Orientierung an einer kriterialen Norm stellt im Allgemeinen die Verwendung von Kriterien dar, die bereits vor der Durchführung der Prüfung entwickelt werden (vgl. Abbildung 3). Auf einer Skala von 1=nie bis 5=immer gaben nur 24 Prozent der Lehrkräfte aus den 5. Klassen an, selten oder gar nicht (Antwortkategorie 1 oder 2) ein vor den Prüfungen entwickeltes Kriterienraster zur Leistungsbewertung einzusetzen. Dagegen nutzen 46 Prozent ihrer Kolleginnen und

Kollegen ein solches Raster immer oder zumindest häufig (Antwortkategorien 4 oder 5). In den 9. Klassenstufen nutzen mit 63 Prozent die Mehrheit der Lehrpersonen häufig oder immer einen Kriterienkatalog zur Beurteilung der Schülerleistungen. Nur 17 Prozent geben an, selten oder nicht davon Gebrauch zu machen.

Der Aussage, dass die gesteckten Ziele in der Beurteilung eine wesentliche Rolle spielen und wer diese Ziele erreicht, auch eine gute Note erhält, konnten 87 Prozent der Lehrpersonen der 5. Klassen und 89 Prozent der Lehrpersonen in den 9. Klassen zustimmen. Weiterhin gaben 80 Prozent der Lehrpersonen aus den 5. Klassen und 90 Prozent ihrer Kolleginnen und Kollegen aus den 9. Klassen an, dass ihre Schülerinnen und Schüler genau wissen, welche Leistungen in den Klassenprüfungen von ihnen erwartet werden und an welchen Kriterien sie gemessen werden. Ferner meint jeweils die Mehrheit der Befragten, dass es durchaus vorkommen kann, dass alle Schülerinnen und Schüler innerhalb einer Prüfung eine überdurchschnittlich gute Note bekommen, wenn die Leistungen der gesamten Klasse dementsprechend gut sind: Eine solche Aussage machen 80 Prozent der Lehrpersonen der 5. und 77 Prozent der Lehrpersonen der 9. Klassen.

In Tabelle 3 sind die Mittelwerte und Standardabweichungen zu den Fragen dieses Blocks insgesamt und getrennt nach Klassenstufen aufgeführt. Hinsichtlich der kriterialen Bezugsnorm zeigt sich, dass die Lehrpersonen der 9. Klassen häufiger angeben, eigene Kriterienvorgaben und -raster zu entwickeln als die Lehrpersonen der 5. Klassen. Ebenfalls meinen die Lehrpersonen der 9. Klassen häufiger, dass ihre Schülerinnen und Schüler genau wissen, an welchen Kriterien ihre Leistung gemessen wird und wie diese bewertet wird. Auch wenn hinsichtlich dieser beiden Aspekte signifikante Unterschiede zwischen den zwei Klassenstufen auftreten, ist dennoch zu betonen, dass insgesamt bei der Leistungsbeurteilung in beiden Klassenstufen, nach Aussage der Lehrpersonen, grosses Gewicht auf den Vergleich mit der kriterialen Norm gelegt wird.

Im Gegensatz zu den Items, die auf die Individual- und Sozialnorm bezogen sind, ist bei den Mittelwerten zu Items der kriterialen Norm eine wesentlich höhere Zustimmung zu beobachten. An dieser Stelle muss erneut betont werden, dass die Zustimmung zur kriterialen Norm noch kein Beweis für die tatsächliche Beurteilungspraxis ist. Der Abgleich an der kriterialen Norm ist gewissermassen für eine standardorientierte Beurteilung erwünscht und insofern könnten die Aussagen der Lehrpersonen in diese Richtung beeinflusst sein.

Hinsichtlich der sozialen Bezugsnorm scheint es für Lehrpersonen der 9. Klassen etwas unwichtiger zu sein, die Noten am Durchschnitt der gesamten Klasse abgleichen zu können. Sie geben deshalb noch seltener als die Lehrpersonen der 5. Klassen an, die Noten erst festzulegen, wenn sie wissen, wie alle anderen Schülerinnen und Schüler abgeschnitten haben. Aus der Umfrage wird nicht deutlich warum dieser Unterschied besteht, es ist jedoch zu vermuten, dass die Lehrpersonen in den 9. Klassen sich stärker an den Kriterien orientieren, deren Erreichen für den anstehenden Einstieg in den Beruf oder für die Übergang in die weiterführenden Schulen wichtig ist.

Hinsichtlich die individuellen Norm zeigen sich hingegen keine Unterschiede zwischen den Klassen. Die Aussagen der Lehrpersonen zur Häufigkeit der Anwendung dieser Methoden liegt in beiden Klassen im mittleren Skalenbereich zwischen "nie" und "immer".

**Tabelle 3 (Antworten zur Festlegung der Prüfungsnoten in den 5. und 9. Klassen. Anzahl Lehrpersonen (N), Mittelwerte (M) und Standardabweichungen (SD). 1=nie, 5=immer, \*\*= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,01$ .)**

Item	Gesamt			5. Klasse			9. Klasse		
	N	M	SD	N	M	SD	N	M	SD
<b>Kriteriale Norm</b>									
In meiner Bewertung spielen die gesteckten Ziele die wesentliche Rolle. „Wer die gesteckten Ziele erreicht, erhält eine gute Note.“	251	4,2	0,8	98	4,1	0,8	153	4,2	0,8
Meine Schülerinnen und Schüler wissen genau, an welchen Kriterien ihre Leistungen gemessen werden und wie diese bewertet werden.	251	4,3	0,7	98	4,1**	0,8	153	4,4**	0,7
Ich entwickle vor der Durchführung der Prüfungen ein Kriterienraster, in dem auch die Noten für die jeweiligen zu erwartenden Leistungen bereits festgelegt sind.	248	3,5	1,2	97	3,2**	1,1	151	3,6**	1,2
Es kann bei mir auch vorkommen, dass alle Schülerinnen und Schüler innerhalb einer Prüfung eine überdurchschnittlich gute Note bekommen, wenn die Leistungen der gesamten Klasse dementsprechend gut sind.	248	4,0	1,1	97	4,0	0,9	151	4,0	1,2
<b>Soziale Norm</b>									
Meine Notengebung lege ich in Abwägung der Leistungen der gesamten Klasse fest. „Wer im Vergleich zum Klassendurchschnitt besser abschneidet, erhält eine gute Note.“	249	2,6	1,2	98	2,7	1,1	151	2,5	1,2
Ich achte bei der Benotung darauf, dass innerhalb der Klasse das Spektrum möglicher Noten gut ausgeschöpft ist, damit die Leistungsunterschiede der Schülerinnen und Schüler darin zum Ausdruck kommen.	250	2,4	1,3	97	2,6	1,2	153	2,3	1,2
Ich lege die Noten erst fest, wenn ich weiss, wie alle Schülerinnen und Schüler der Klasse abgeschnitten haben.	250	2,7	1,3	98	3,0**	1,2	152	2,6**	1,3
<b>Individuelle Norm</b>									
Ich lege die Noten in Abwägung des Lernfortschritts der Schülerin / des Schülers fest: "Wer viel Fortschritt macht, erhält eine gute Note".	248	2,4	1,0	96	2,2	0,9	152	2,5	1,0
Wenn ich weiss, dass ein Schüler oder eine Schülerin sich gegenüber der bisherigen Leistung verbessert, bekommt er/sie auch eine bessere Note.	248	2,5	1,1	96	2,4	1,0	152	2,5	1,2
Wenn ich sehe, dass sich eine Schülerin oder ein Schüler angestrengt hat, ist das für mich bei der Notengebung wichtiger als das Leistungsergebnis oder der Vergleich mit anderen.	250	2,5	1,0	97	2,5	0,9	153	2,4	1,0

Es scheint also, als orientierten die Lehrpersonen sich in der Festlegung der Noten weitgehend an der kriterialen Norm, indem sie den Leistungsstand an konkreten Lernzielen abgleichen und Kriterien verwenden. Die von Vögeli-Mantovani (1999) geäußerte Kritik, die Beurteilungskriterien seien häufig nicht transparent und die Beurteilung orientiere sich häufig am Klassenbezugssystem, wird damit von den Lehrpersonen zurückgewiesen. Über die fachdidaktisch begründete Qualität der verwendeten Kriterien gibt die vorliegende Untersuchung jedoch keine Anhaltspunkte.

## 4.2. Zeugnis

Auf die Frage, wie die Noten für das Zeugnis festgelegt werden, zeigt sich für beide Klassenstufen deutlich die Berücksichtigung der Selbstkompetenz wie z.B. Fleiss, Mitarbeit oder Motivation. Danach erst folgen andere Aspekte wie z.B. auch die Berechnung des mathematisch gerundeten Notendurchschnittes, die insbesondere von Vögeli-Mantovani (1999) kritisiert wird.

Tabelle 4 gibt die gültigen Fälle, die Mittelwerte und die Standardabweichungen zu den diesbezüglichen Aussagen wieder.

**Tabelle 4 (Antworten zur Festlegung der Zeugnisnoten in den 5. und 9. Klassen. Gültige Fälle, Mittelwerte und Standardabweichungen. 1=nie, 5=immer; \*\*= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0.01$ .)**

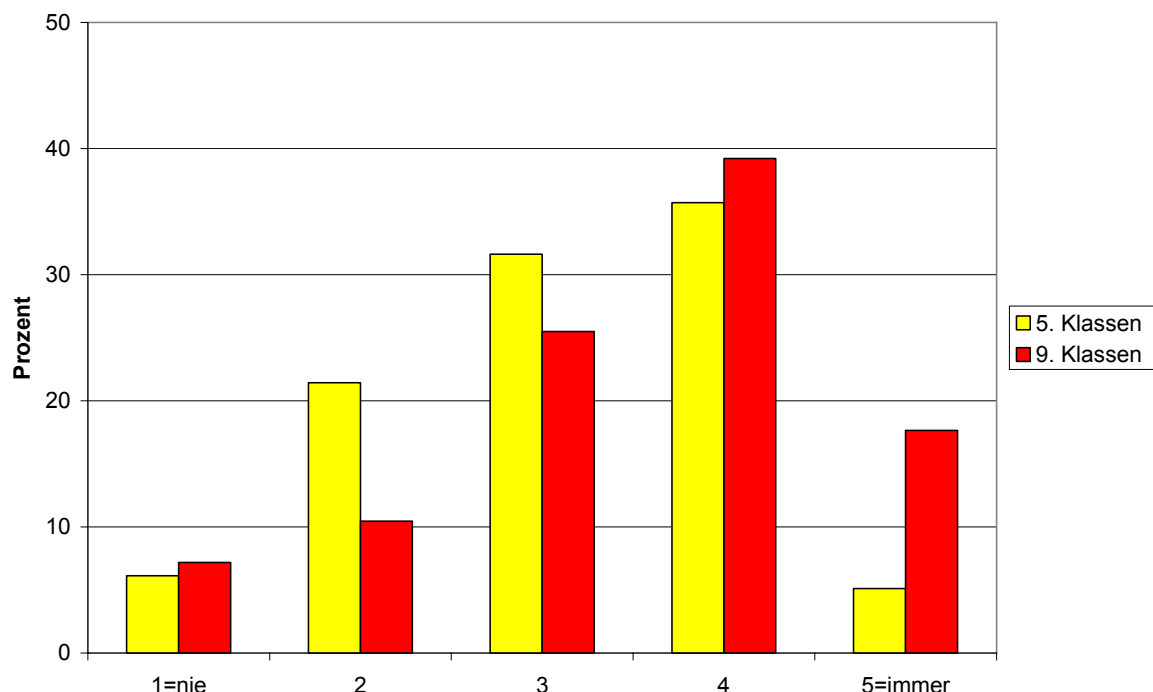
Item	Gesamt			Klassenstufe 5			Klassenstufe 9		
	N	M	SD	N	M	SD	N	M	SD
Wenn ich die Zeugnisnote festlege, runde ich den Durchschnitt der Klassenprüfungsnoten streng mathematisch (z.B. aufrunden bei 4,75 auf 5).	251	3,4	1,1	98	3,1**	1,0	153	3,5**	1,1
Wenn eine Schülerin oder ein Schüler zwischen zwei Noten steht, lasse ich zu meiner Entscheidung für die Zeugnisnote auch meine Beobachtungen über die Sozialkompetenz mit einfließen.	250	2,9	1,4	98	2,8	1,4	152	3,0	1,3
Wenn eine Schülerin oder ein Schüler zwischen zwei Noten steht, lasse ich zu meiner Entscheidung für die Zeugnisnote auch meine Beobachtungen über die Selbstkompetenz (z.B. Motivation und Fleiss) mit einfließen.	250	3,9	1,1	98	3,9	1,0	152	4,0	1,1
Wenn eine Schülerin oder ein Schüler zwischen zwei Noten steht, berücksichtige ich für die Zeugnisnote auch die Konsequenzen hinsichtlich der Promotion bzw. Remotion.	249	3,1	1,1	98	3,4**	1,1	151	3,0**	1,2
Ich gewichte für die Zeugnisnote die Noten je nach Wichtigkeit der zugrundeliegenden Leistungen, indem ich gewisse Prüfungen / Leistungen doppelt zähle, andere nur halb.	249	3,3	1,5	98	3,2	1,4	151	3,4	1,4
Die Selbst- und Sozialkompetenz bewerte ich während des Semesters und lasse sie wie eine Prüfungsnote in die Zeugnisnote mit einfließen.	249	2,2	1,3	96	1,9**	1,0	153	2,5**	1,4

Die Grundlage für die Notengebung im Zeugnis ist nicht in allen Fällen für die zwei Klassen vergleichbar. Während in den 9. Klassen häufiger der mathematisch gerundete No-

tendurchschnitt ausschlaggebend ist und teilweise auch Selbst- und Sozialkompetenz während des Semesters separat bewertet und in die Zeugnisnote integriert wird, scheinen in den 5. Klassen häufiger auch die Konsequenzen für die Remotion oder Promotion herangezogen zu werden. In den 5. Klassen wird generell eher selten die Selbst- oder Sozialkompetenz während des Semesters bewertet, um sie dann *wie eine Prüfungsnote* in die Zeugnisnote mit einfließen zu lassen. Dennoch ist auffällig, dass die Beobachtungen über die Sozialkompetenz wie z.B. Motivation und Fleiss nach Angaben der Lehrpersonen in beiden Klassenstufen recht häufig in die Zeugnisnote einfließen, *wenn eine Schülerin oder ein Schüler zwischen zwei Noten steht*. Es ist in diesem Zusammenhang zu vermuten, dass diese Beobachtungen einen grossen subjektiven Spielraum zulassen. Dieser scheint in den 5. Klassen aufgrund der selteneren bewussten Bewertungen der Selbst- und Sozialkompetenz während des Semesters höher zu sein als in den 9. Klassen.

Im Vorfeld der Untersuchung wurde die Hypothese aufgestellt, dass vorwiegend der Notendurchschnitt der Klassenprüfungen eines Semesters der Zeugnisnote entspricht, also einen direkten Einfluss auf die Zeugnisnote hat. Um diese Hypothese testen zu können, wurde bei der Konstruktion des Messinstrumentes eine entsprechende Itemformulierung bedacht. So konnte die Aussage *Wenn ich die Zeugnisnote festlege, runde ich den Durchschnitt der Klassenprüfungsnoten streng mathematisch* wieder anhand einer Fünferskala von 1 für *nie* bis 5 für *immer* beantwortet werden.

Das Ergebnis für die Gesamtstichprobe (5. und 9. Klassen zusammen, N = 251) zeigt im Durchschnitt eine mittlere Tendenz von  $M = 3,4$  mit einer Streuung von einer Skaleneinheit ( $SD = 1,1$ ). Die Prozentangaben der Lehrpersonen zu dieser Aussage sind in Abbildung 4 im Detail aufgeführt.



**Abbildung 4 (Angaben zu der Aussage "Wenn ich die Zeugnisnote festlege, runde ich den Durchschnitt der Klassenprüfungsnoten streng mathematisch" in Prozent nach Klassenstufen.)**

Fasst man die Antworten 1 (nie) und 2 sowie 4 und 5 (immer) ohne Mittelkategorie zusammen, wird deutlich, dass gut die Hälfte der befragten Lehrerinnen und Lehrer praktisch immer das arithmetische Mittel der Klassenprüfungsnoten des Semesters zur Festlegung der Schulzeugnisnoten nutzen. Nur etwas mehr als 20 Prozent aller Befragten geben an, sich kaum oder gar nicht dieser Art der Notenvergabe für Schulzeugnisse zu bedienen. In der Gegenüberstellung der beiden Klassen lässt sich eine ähnliche Antwortverteilung wie die der Gesamtstichprobe erkennen. Diejenigen Lehrpersonen, die den Notendurchschnitt immer runden, unterrichten jedoch häufiger in den 9. Klassen. Vögeli-Mantovani (1999) hebt gerade die mathematische Durchschnittsbildung von Ziffernoten deutlich als problematisch und trügerisch hervor und auch Brügelmann et al. (2006) empfiehlt, die Ziffernoten durch differenziertere Formen der Dokumentation und Bewertung von Leistung zu ersetzen. Hinsichtlich dieses Aspektes scheint noch eine relativ grosse Uneinigkeit bei vielen Lehrpersonen zu bestehen. Die mathematische Durchschnittsbildung scheint auf den ersten Blick logisch und objektiv, dazu kommt, dass dieses Vorgehen einfach zu realisieren ist, während die Forderung nach differenzierteren Formen der Dokumentation und Bewertung zunächst abstrakt und schwer einlösbar erscheint. Wenn es das Ziel sein soll, Bewertungen differenzierter zu begründen, als dies durch die mathematische Durchschnittsbildung der Prüfungsnoten geschieht, werden die Lehrpersonen eine stärkere Unterstützung in diesem Bereich benötigen.

## 5. Verwendete Messinstrumente

Ein Teil der hier dargestellten Untersuchung beinhaltete die Frage, inwiefern unterschiedliche Messinstrumente in der Praxis Anwendung finden. Im Vorfeld der Untersuchung stand die Hypothese, dass die Notengebung hauptsächlich auf schriftlichen Klassenprüfungen und anderen schriftlichen Leistungen und weniger auf mündlichen Leistungsmessungen oder der Messung von Selbst- oder Sozialkompetenz beruht. Zur Überprüfung dieser Hypothese wurde ein Fragenblock entwickelt, bei dem auf einer fünfstufigen Antwortskala (5 = mehrmals wöchentlich, 4 = mehrmals im Monat, 3 = mehrmals in diesem Semester, 2 = einmal in diesem Semester, 1 = in diesem Semester nicht eingesetzt) angegeben werden konnte, welche Arten der Leistungsmessung wie häufig eingesetzt werden. Explizit wurde nach folgenden Formen der schriftlichen und mündlichen Leistungsmessung sowie nach Formen der Beurteilung von Selbst- und Sozialkompetenz gefragt:

*Sachkompetenz:* Schriftliche Leistungen gemessen mit:

- selbstentwickelten schriftlichen Klassenprüfungen,
- selbstentwickelten schriftlichen Vergleichsarbeiten innerhalb der Schule,
- schriftlichen Vergleichsarbeiten zum Vergleich mit anderen Schulen (z.B. Check 5),
- schriftlichen Arbeiten (z.B. Aufsätze, Projektarbeiten, ...),
- schriftlichen Gruppenarbeiten,
- schriftlicher Bearbeitung von Hausaufgaben,
- Heftführung,
- Arbeits- oder Lernjournalen,
- Selbstbeurteilungsbögen der Schülerinnen und Schüler zu ihrer schriftlichen Leistung.

*Sachkompetenz:* Mündliche Leistungen gemessen mit:

- selbstentwickelten mündlichen Klassenprüfungen,
- selbstentwickelten mündlichen Vergleichsarbeiten innerhalb der Schule,
- Präsentationen von Einzelarbeiten,
- Präsentationen von Gruppenarbeiten,
- Präsentationen von Hausaufgaben,
- Beobachtungslisten für die mündliche Mitarbeit,
- Selbstbeurteilungsbögen der Schülerinnen und Schüler zu ihrer mündlichen Leistung.

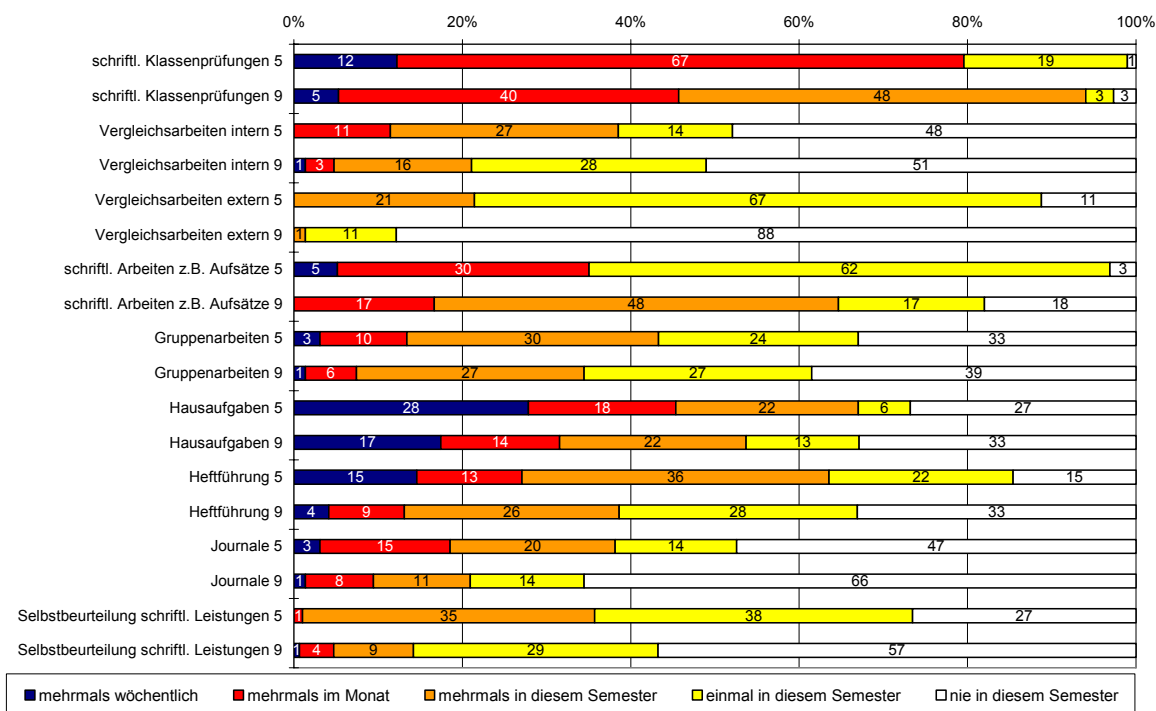
*Selbst- und Sozialkompetenz:* gemessen mit:

- Beobachtungslisten zur Selbstkompetenz,
- Beobachtungslisten zur Sozialkompetenz,
- Selbstbeurteilung der Schülerinnen und Schüler zu ihrer Selbstkompetenz,
- Selbstbeurteilungsbögen der Schülerinnen und Schüler zu ihrer Sozialkompetenz.

Über beide Klassen hinweg zeigt sich tatsächlich, dass schriftliche Klassenprüfungen als Messinstrumente neben schriftlichen Arbeiten wie Projektarbeiten oder Aufsätzen und schriftlichen Hausaufgaben am häufigsten sind (vgl. Abbildung 5). Fast alle Lehrpersonen setzen schriftliche Klassenprüfungen mindestens einmal im Semester ein und fast 90 Prozent verwenden mindestens einmal schriftliche Arbeiten wie z.B. Aufsätze, Projektarbeiten etc., um die Leistung der Schülerinnen und Schüler zu messen. Auch schriftliche Hausaufgaben und die Heftführung werden relativ häufig als Beurteilungs-

grundlage verwendet. Fast 60 Prozent aller Lehrpersonen beurteilen die Hausaufgaben mehrmals in einem Semester und fast 50 Prozent kontrollieren mehrfach die Heftführung. Selbstbeurteilungen zur schriftlichen Leistung sowie Arbeits- und Lernjournale werden sehr viel seltener zur Beurteilung herangezogen. Bei der Hälfte der Lehrpersonen spielen weiterhin interne Vergleichsarbeiten zum Vergleich innerhalb der Schule mindestens einmal pro Semester eine Rolle. Externe Vergleichsarbeiten sind in den 9. Klassen sehr selten, werden aber in den 5. Klassen häufig genutzt.

Beim Vergleich der Verwendung von Messinstrumenten für schriftliche Leistungen über die zwei Klassen fällt auf, dass diese fast durchgängig in den 5. Klassen häufiger Verwendung finden als in den 9. Klassen. 80 Prozent der Lehrpersonen der 5. Klassen setzten mehrmals im Monat schriftliche Klassenprüfungen ein, während dies in den 9. Klassen nicht einmal die Hälfte der Lehrpersonen so häufig tun. Besonders auffällig ist der Unterschied zwischen den Klassen auch bei den Selbstbeurteilungen zur schriftlichen Leistung. Während über 70 Prozent der Lehrpersonen in den 5. Klassen dieses Instrument mindestens einmal im Semester einsetzen, benutzen nicht einmal die Hälfte der Lehrpersonen der 9. Klassen es so häufig zur Messung der schriftlichen Leistungen der Schülerinnen und Schüler. Dass externe Vergleichsarbeiten weit häufiger in den 5. Klassen anzutreffen sind, als in den 9. Klassen, ist nicht weiter verwunderlich, da der Check 5 für die Lehrpersonen der 5. Klassen ein externes Instrument anbietet, dass vielfach genutzt wird. Nach den Angaben der Lehrpersonen aus dieser Stichprobe, kann davon ausgegangen werden, dass ca. 90 Prozent der Lehrpersonen dieser Klassen den Check 5 einsetzen.

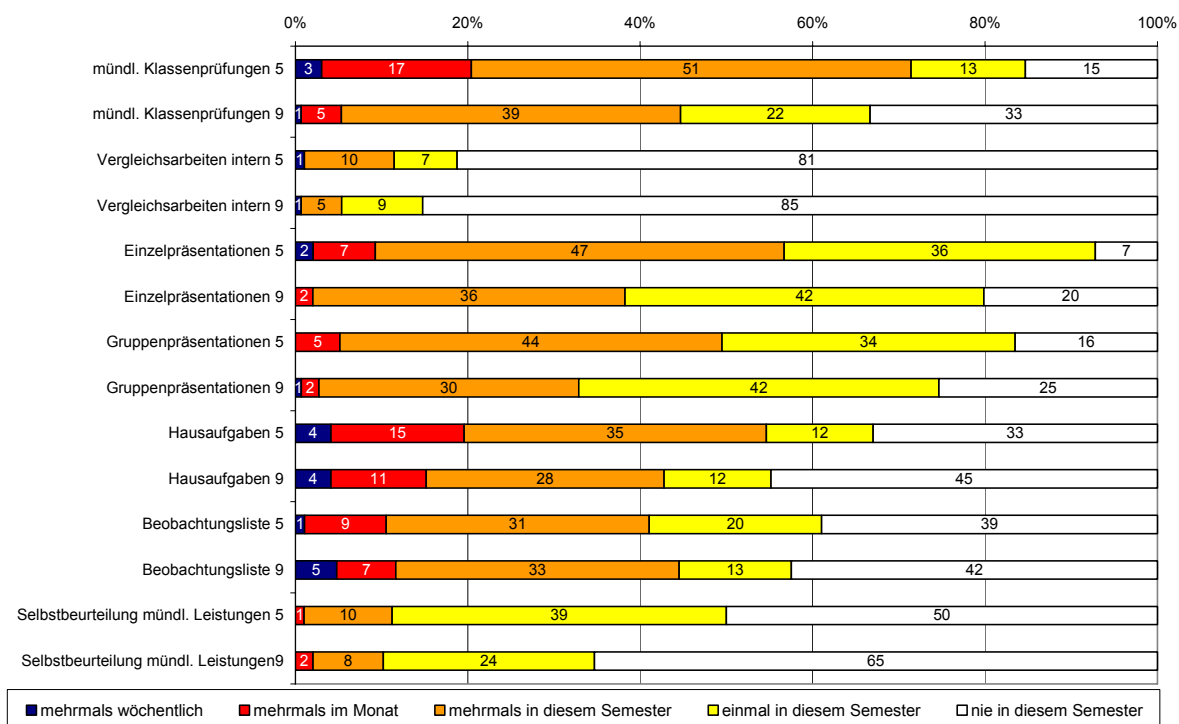


**Abbildung 5 (Angaben zu der Frage "Welche Leistungsbeurteilungen haben Sie im 1. Semester des Schuljahres 2007/2008 zur Feststellung der schriftlichen Leistungen pro Klasse wie häufig eingesetzt?" in Prozent nach Klassenstufen.)**

Messinstrumente kommen für mündliche Leistungen wesentlich seltener zum Einsatz als für schriftliche Leistungen (Abbildung 6). Dabei sind mündliche Klassenprüfungen, Einzelpräsentationen und Gruppenpräsentationen am häufigsten. Jeweils zwischen 40 und 50 Prozent der Lehrpersonen misst die mündliche Leistung mehrmals im Semester mit Hilfe von Einzelpräsentationen, Gruppenpräsentationen, Hausaufgaben und Be-

obachtungslisten. Interne mündliche Vergleichsarbeiten wie z.B. gemeinsam durchgeführte Sprechtests im Fremdsprachenunterricht spielen fast überhaupt keine Rolle, während Selbstbeurteilungen bei mündlichen Leistungen zwar nicht so häufig vorkommen wie bei schriftlichen Leistungen, aber dennoch von über 40 Prozent der Lehrpersonen mindestens einmal im Semester für die Beurteilung berücksichtigt werden.

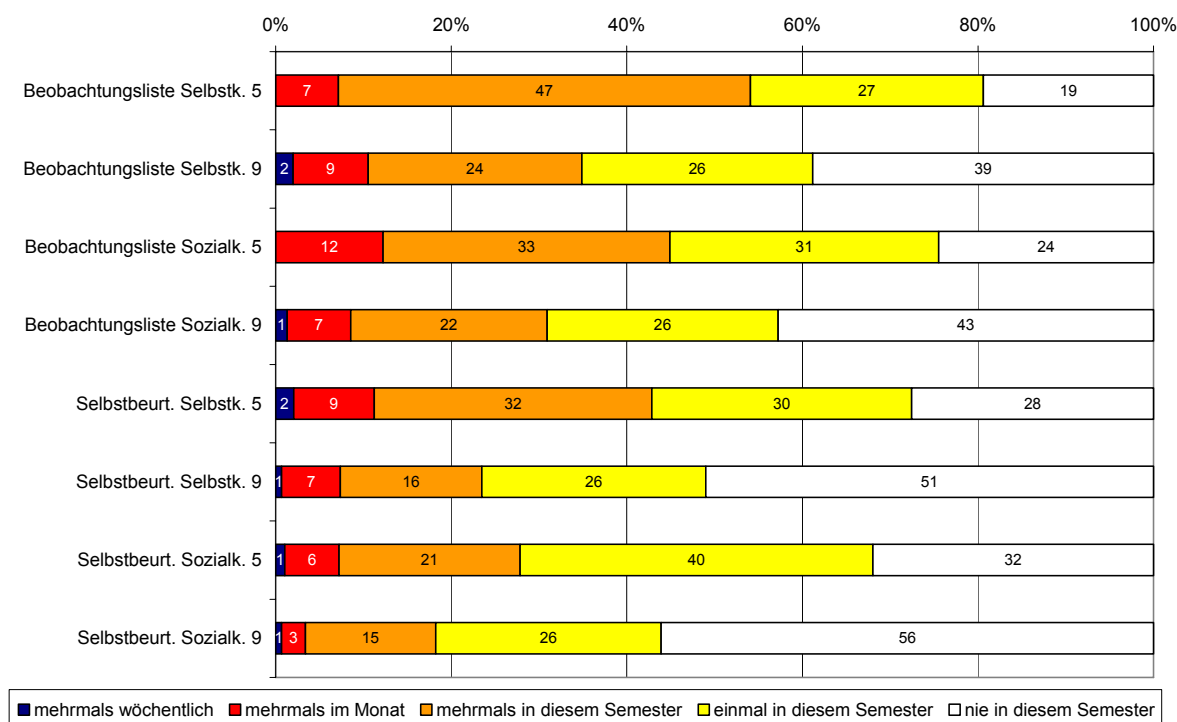
Betrachtet man die in den 5. und 9. Klassen verwendeten Messinstrumente für mündliche Leistungen, zeigt sich ein ganz ähnliches Bild wie hinsichtlich der schriftlichen Leistungen. Fast in allen Fällen wird jeweils in den 5. Klassen häufiger auf die Messinstrumente zurückgegriffen. Eine Ausnahme bilden die Beobachtungslisten für mündliche Leistungen. In beiden Klassenstufen werden diese von ca. 60 Prozent der Lehrpersonen mindestens einmal pro Semester und von etwas mehr als 40 Prozent mehrmals im Semester verwendet. Ein Unterschied zwischen den Klassen ist hier kaum erkennbar.



**Abbildung 6 (Angaben zu der Frage "Welche Leistungsbeurteilungen haben Sie im 1. Semester des Schuljahres 2007/2008 zur Feststellung der mündlichen Leistungen pro Klasse wie häufig eingesetzt?" in Prozent nach Klassenstufen.)**

Generell scheinen Selbstbeurteilungen vor allem im Bereich der Selbstkompetenz eine Rolle zu spielen (vgl. Abbildung 7). Bei über 30 Prozent der Lehrpersonen kommen diese Selbstbeurteilungen mehrmals im Semester zum Tragen und fast 60 Prozent verwenden sie mindestens einmal. Die Selbstbeurteilungen zur Sozialkompetenz sind etwas seltener. Nur etwas mehr als 20 Prozent der Lehrpersonen greift mehrmals im Semester darauf zurück und knapp 45 Prozent verwenden sie nie. Die Lehrpersonen scheinen sich auch hinsichtlich der Selbst- und Sozialkompetenz mehr auf ihre eigenen Beobachtungen zu verlassen, als auf die Einschätzungen der Schülerinnen und Schüler. Teilweise werden dazu Beobachtungslisten verwendet. Rund 40 Prozent der Lehrpersonen verwendet mehrmals im Monat Beobachtungslisten zur Beurteilung der Selbst- und Sozialkompetenz. Nur etwas mehr als je 30 Prozent geben an, nie solche Beobachtungslisten zu verwenden. Wie auch bei der Verwendung der Beurteilungsinstrumente für schriftliche und mündliche Leistungen fällt auf, dass in den 5. Klassen

insgesamt häufiger darauf zurückgegriffen wird. Dieser Befund steht in einem gewissen Widerspruch zu den Angaben, der Lehrpersonen dazu, wie sie die Noten für die Zeugnisse festlegen. Für die 5. Klassen wurde in diesem Zusammenhang festgestellt, dass die Beobachtungen der Sozialkompetenz zwar Eingang in das Zeugnis finden, *wenn eine Schülerin oder ein Schüler zwischen zwei Noten steht*, dies aber eher selten auf der Grundlage spezifischer Bewertungen über das Semester geschieht. Beobachtungslisten zur Selbst- und Sozialkompetenz werden also nach den Angaben der Lehrpersonen recht häufig geführt. Es scheinen daraus aber in der Regel keine spezifischen semesterbegleitenden Bewertungen abgeleitet zu werden, die wie eine eigene Prüfungsnote in das Zeugnis eingehen (vgl. Abschnitt 4.2). Statt dessen werden diese Kompetenzen erst am Ende des Semesters bei Bedarf verrechnet.



**Abbildung 7 (Angaben zu der Frage "Welche Leistungsbeurteilungen haben Sie im 1. Semester des Schuljahres 2007/2008 zur Feststellung der Selbstkompetenz und der Sozialkompetenz pro Klasse wie häufig eingesetzt?" in Prozent nach Klassenstufen.)**

Es scheinen insgesamt relativ viele unterschiedliche Formen der Leistungsbeurteilung zum Tragen zu kommen, wie es beispielsweise auch von Brügelmann et al. (2006) vorgeschlagen wird. Auffällig ist jedoch, dass Vergleichsarbeiten (mit Ausnahme der externen schriftlichen Vergleichsarbeiten in der 5. Klasse) recht wenig genutzt werden. Schriftliche Vergleichsarbeiten sind dabei noch etwas häufiger als mündliche. Nach der Einschätzung vieler Experten ist die Orientierung am Klassenbezugssystem höchst problematisch (z.B. Ingenkamp, 1971; Ziegenspeck, 1999; Vögeli-Mantovari, 1999; Brügelmann et al., 2006) und eine Möglichkeit die Sensibilität für diese Problematik zu stärken, ist der klassenübergreifende Vergleich. Eine Leistungsbewertung sollte nach den Einschätzungen von Brügelmann et al. (2006) von mehreren involvierten Personen vorgenommen werden, ebenfalls wird die Notwendigkeit eines kollegialen Austauschs betont. Ein häufigerer Einsatz von internen, aber auch externen Vergleichsarbeiten könnte in diese Richtung wirksam sein. Es ist dabei jedoch einerseits zu bedenken, dass der gemeinsame Austausch über die Leistungsbewertungen zusätzliche zeitliche Ressourcen bindet, die bei der Forderung nach einheitlichen Beurteilungen von vornherein eingeplant werden sollten. Andererseits mag der Austausch mit anderen Fachkolleginnen

und Fachkollegen aber auch die Nutzung von Synergien ermöglichen (beispielsweise gemeinsame Prüfungen, oder gemeinsam genutzte Unterrichtsmaterialien), die wiederum zu einer Zeiteinsparung führen könnten.

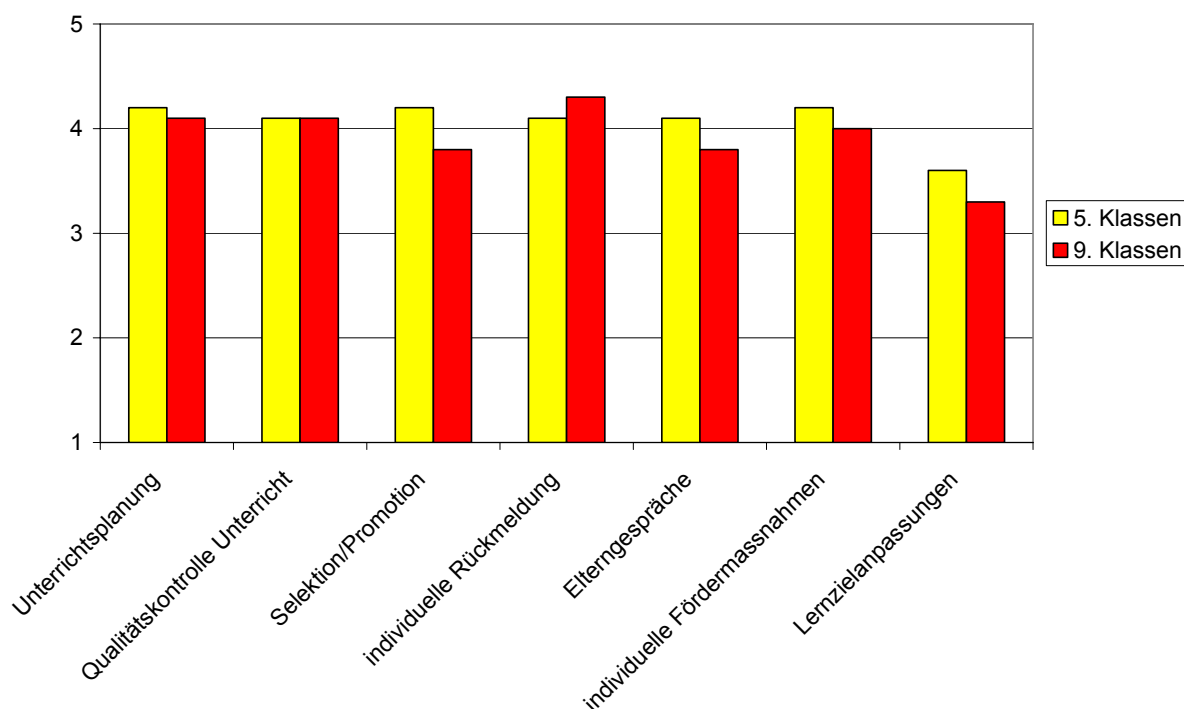
Neben den vorgegebenen Antworten wurden die Befragten gebeten, frei zu nennen, welche anderen Beurteilungsmöglichkeiten sie für schriftliche und mündliche Leistungen sowie für die Beurteilung von Selbst- und Sozialkompetenz benutzen. Die Auswertung zeigt, dass als Grundlage für die weitere schriftliche Leistungsbeurteilung hauptsächlich Gespräche über erbrachte Leistungen (im Klassenverband oder mit einzelnen Schülern), praktische Arbeiten (Selbstgestaltung von Lektionen, Geschichtsspiel) und das Portfolio benutzt werden. Weiterhin wurden als Beurteilungskriterium für schriftliche Leistungen unterschiedliche Formen der Selbstbeurteilung der Schüler genannt.

Ähnliche Freitextnennungen gab es auch im Zusammenhang mit der mündlichen Leistung. Dabei stehen Gespräche mit den Schülerinnen und Schülern und praktische Arbeiten wie Vorträge, Vorsingen, Moderation und Diskussionen im Vordergrund. Im Zusammenhang mit der Leistungsmessung im Bereich der Selbst- und Sozialkompetenzen legen die Freitextnennungen ebenfalls besonderen Wert auf Gespräche mit den Schülerinnen und Schülern. Auch werden teilweise Rücksprachen mit dem Klassenrat in die Beurteilung der Selbst- und Sozialkompetenz mit eingebracht. Ferner werden Arbeits- und Lernjournale sowie das Lerntagebuch als Messinstrument dieser Kompetenzbereiche genutzt.

Diese Rückmeldungen sowie auch die Verwendung von Selbstbeurteilungen für schriftliche und mündliche Leistungen sowie für die Selbst- und Sozialkompetenz lassen darauf schließen, dass in einigen Klassen bereits viele unterschiedliche Formen der Beurteilung verwendet werden. Brügelmann et al. (2006) schlagen weiterhin vor, eine dialogische Form einer Verbindung von Selbst- und Fremdbeurteilung zu erproben. Für ca. die Hälfte der Klassen kann aufgrund der Ergebnisse festgestellt werden, dass Selbstbeurteilungen mindestens einmal pro Semester eine Rolle spielen. Inwiefern die Selbstbeurteilung jedoch dialogisch mit der Fremdbeurteilung verkoppelt ist, geht aus dieser Untersuchung nicht hervor.

## 6. Leistungsergebnisse als Basis für Fördermassnahmen

Zur Untersuchung der Frage, wie mit Leistungsergebnissen der Schülerinnen und Schüler umgegangen wird und wozu sie konkret genutzt werden, wurden 7 Fragen formuliert, die sich einerseits auf Förderungsmassnahmen für die Schülerinnen und Schüler sowie andererseits auf Unterrichtsqualität und Unterrichtsplanung beziehen. Die Befragten hatten hierzu wieder die Möglichkeit, ihre Antworten anhand einer Fünferskala von 1 für *Das ist mir gar nicht wichtig* bis 5 für *Das ist mir sehr wichtig* abzustufen.



**Abbildung 8** (Angaben zu der Frage "Wie wichtig sind Ihnen die Leistungsergebnisse?" nach Klassenstufen. Mittelwerte der fünfstufigen Antwortskala. 1=gar nicht wichtig, 5=sehr wichtig.)

Bezüglich der Frage, inwiefern Leistungsergebnisse als Ausgangslage zu weiteren Förderungsmassnahmen oder Unterrichtsentwicklung genutzt werden, wird für beide Klassenstufen deutlich, wie wichtig den Lehrpersonen die Leistungsmessungen in diesem Zusammenhang sind. Der Nutzen der Ergebnisse der Leistungsmessung wird für alle abgefragten Aspekte als sehr hoch eingeschätzt (vgl. Abbildung 8). Dies entspricht der Einschätzung von Vögeli-Mantovani (1999), der darauf hinweist, dass Noten sehr viele Funktionen zu erfüllen haben, die jede für sich genommen, zu wenig konkret definiert sind. Besonders hoch schätzen die Lehrpersonen die Möglichkeiten der individuellen Förderung und der individuellen Rückmeldung an die Schülerinnen und Schüler ein. Selektion und Promotion sind wichtige Themen in den 5. Klassen, wie auch aus Tabelle 5 ersichtlich ist. Ebenso fällt auf, dass die Leistungsmessungsergebnisse in den 5. Klassen noch stärker als in den 9. Klassen als wichtige Grundlage für die Gespräche mit den Eltern angesehen werden. Die Mittelwertsunterschiede hinsichtlich der beiden letztgenannten Aspekte ist zwischen den beiden Klassen signifikant ( $\alpha < 0,01$  Selektion/Promotion;  $\alpha < 0,05$  Elterngespräche). Auffallend ist, dass in beiden Klassen die Bedeutung der Leistungsergebnisse für die Lernzielanpassung am geringsten beurteilt wird.

**Tabelle 5 (Nutzung der Leistungsmessungsergebnisse in den 5. und 9. Klassen. Gültige Fälle, Mittelwerte und Standardabweichungen. 1=gar nicht wichtig, 5=sehr wichtig, \*\*/\*= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,01 / \alpha < 0,05$ ). )**

Wie wichtig sind Ihnen die Leistungsergebnisse ...	Gesamt			5. Klasse			9. Klasse		
	N	M	SD	N	M	SD	N	M	SD
... als Grundlage für die weitere Planung meines Unterrichts?	251	4,1	0,8	98	4,2	0,7	153	4,1	0,9
... für die Qualitätskontrolle meines Unterrichts?	250	4,1	0,7	97	4,1	0,6	153	4,1	0,8
... als Grundlage für Promotions- und Selektionsentscheide?	251	4,0	1,1	98	4,2**	0,7	153	3,8**	1,2
... als Grundlage für individuelle Rückmeldungen an die Schülerinnen und Schüler?	250	4,2	0,8	97	4,1	0,8	153	4,3	0,7
... als Grundlage für die Gespräche mit den Eltern?	251	3,9	0,9	98	4,1*	0,7	153	3,8*	1,0
... als Grundlage für individuelle Fördermassnahmen für einzelne Schülerinnen und Schüler?	250	4,1	0,9	97	4,2	0,8	153	4,0	0,9
... als Grundlage für Lernzielanpassungen für einzelne Schülerinnen und Schüler?	251	3,4	1,2	98	3,6	1,1	153	3,3	1,2

Die Bedeutung der Noten für die individuelle Rückmeldung und die Elterngespräche lässt darauf schliessen, dass Lehrpersonen sich darum bemühen, eine differenzierte Beschreibung und Kommentierung der Leistungen abzugeben. Problematisch scheint eher der Anspruch an die Lehrpersonen, die Noten für sehr unterschiedliche Funktionen nutzbar zu machen. Die Lehrpersonen scheinen in einigen Fällen den Noten Funktionen zuzuschreiben, die über den Förderauftrag hinausgehen und teilweise schwer mit diesem in Einklang zu bringen sind. Vermutlich ist die Ursache dafür nicht in den Kompetenzen der Lehrpersonen, sondern vielmehr in den Anforderungen zu sehen, die an sie gestellt sind, in Verbindung mit den Umständen, unter denen sie diese Anforderungen erfüllen müssen. Die individuelle Förderung, die Beurteilung der Leistungen für Selektionsentscheide und die Entwicklung des eigenen Unterrichts sind Anforderungen, die sich untereinander sehr stark unterscheiden. Ebenso widerspricht der Anspruch nach Selektion der Vorstellung, jeder Schüler und jede Schülerin könne durch adäquate individuelle Förderung die geforderten Leistungen erbringen.

Auch wenn den Noten, wie die vorliegende Untersuchung zeigt, insgesamt unterschiedliche Funktionen zugeordnet werden, scheinen die Lehrpersonen sich jedoch recht deutlich bewusst zu sein, welche Funktionsüberschneidungen es dabei gibt. Das zeigen die Korrelationen zwischen den Einschätzungen zu unterschiedlichen Formen des Umgangs mit den Leistungsmessungen, die in Tabelle 6 dargestellt sind. Insbesondere diejenigen Lehrpersonen, die besonderen Wert auf die individuelle Förderung und Lernzielanpassung für die Schülerinnen und Schüler legen, betrachten den Nutzen für die Promotion und Selektion sowie für die Qualitätskontrolle ihres Unterrichts als weniger wichtig. Individuelle Förderung und Lernzielanpassung sind hingegen Funktionen, die nach Ansicht der Lehrpersonen sehr eng verknüpft sind.

**Tabelle 6 (Korrelationen der Einschätzungen zu den Formen des Umgangs mit der der Leistungsmessung.)**

	Planung Unterricht	Qualitätskontrolle	Promotion/Selektion	Ind. Rückmeldung	Elterngespräche	Ind. Förderung	Lernzielanpassung
Planung Unterricht	1	0,26**	0,18**	0,27**	0,26**	0,42**	0,2**
Qualitätskontrolle	0,26**	1	0,24**	0,16*	0,21**	n.s.	n.s.
Promotion/Selektion	0,18**	0,24**	1	n.s.	0,43**	n.s.	n.s.
Ind. Rückmeldung	0,27**	0,16*	n.s.	1	0,34**	0,32**	n.s.
Elterngespräche	0,26**	0,21**	0,43**	0,34**	1	0,25**	0,16*
Ind. Förderung	0,42**	n.s.	n.s.	0,32**	0,25**	1	0,57**
Lernzielanpassung	0,23**	n.s.	n.s.	n.s.	0,16*	0,57**	1

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

n.s. nicht signifikant.

Neben der Frage nach dem Umgang mit den Leistungsmessungen, wurden die Lehrpersonen auch gebeten, Auskunft darüber zu geben, welche Form der Rückmeldung sie wählen. In beiden Klassenstufen geben die Lehrpersonen in der Mehrheit an, den Schülerinnen und Schülern jeweils eine Rückmeldung zum relativen Leistungsstand in der Klasse zu geben (Tabelle 7).

Auch ausführliche mündliche Rückmeldungen werden recht häufig erteilt ( $M = 3,3$ ), während schriftliche Rückmeldungen zum Leistungsstand eher seltener sind ( $M = 2,4$ ). Tendenziell erteilen die Lehrpersonen in den 9. Klassenstufen eher eine ausführliche mündliche Rückmeldung als diejenigen der 5. Klassen. Die Mittelwertsunterschiede sind statistisch signifikant ( $\alpha < 0,05$ )

**Tabelle 7 (Rückmeldungen der Leistungsmessungsergebnisse in den 5. und 9. Klassen. Gültige Fälle, Mittelwerte und Standardabweichungen. 1=nie, 5=immer, \*= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,05$ .)**

Ich erteile den Schülerinnen und Schülern ...	Gesamt			Jg.-Stufe 5			Jg.-Stufe 9		
	N	M	SD	N	M	SD	N	M	SD
... jeweils eine Rückmeldung zum Stand in der Klasse (Klassendurchschnitt, Rangliste, Verteilung o.ä.)	251	3,4	1,42	98	3,3	1,51	153	3,5	1,37
... jeweils eine ausführliche mündliche Rückmeldung zu ihrer Leistung.	251	3,3	1,00	98	3,1*	0,93	153	3,4*	1,03
... neben der Note jeweils eine ausführliche schriftliche Rückmeldung zu ihrer Leistung.	249	2,4	0,95	98	2,6	0,86	151	2,4	1,00

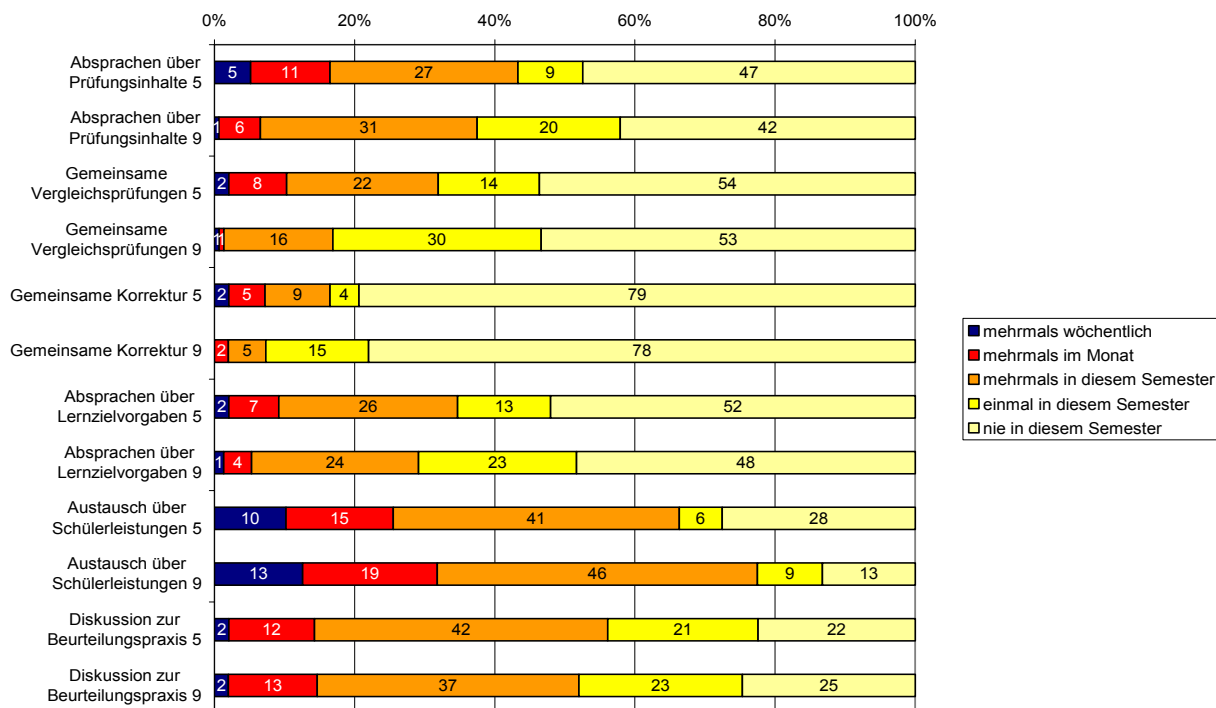
Dass ausführliche mündliche Rückmeldungen zu den Leistungen erteilt werden, entspricht den Vorschlägen von Brügelmann et al. (2006) nach denen die Leistungen nicht nur bewertet, sondern differenziert beschrieben und schriftlich oder mündlich kommentiert werden sollte. Aussagen darüber, wie differenziert die Rückmeldungen der hier befragten Lehrpersonen ausfallen, lassen sich aufgrund der vorliegenden Daten nicht machen. Es ist anzunehmen, dass die Lehrpersonen aufgrund beschränkter Zeitressourcen grösstenteils auf eine ausführliche schriftliche Rückmeldung verzichten und eher mündlich kommentieren. Dabei geben sie häufig, obwohl sie tendenziell eher kriterienorientiert prüfen (s.o.), den Schülerinnen und Schülern eine Rückmeldung zum relativen Stand in der Klasse. Die Orientierung am Klassenbezugssystem ist offensichtlich noch vorherrschend, obwohl die Lehrpersonen angeben, ihre Leistungsbeurteilungen bewusst kriterienorientiert vorzunehmen.

## 7. Zusammenarbeit im Bereich Leistungsbeurteilung

Die Kooperation und der Austausch zwischen den Lehrpersonen bezüglich der Leistungsbeurteilung ihrer Schülerinnen und Schüler war ein weiterer Schwerpunkt des vorliegenden Projektes. Um einen genaueren Überblick darüber zu erhalten, wurden fünf Aussagen für den Fragebogen konstruiert. Die im Folgenden genannten Aussagen wurden von den befragten Lehrpersonen bewertet, wobei diese die Auswahl zwischen *mehrmals wöchentlich*, *mehrmals im Monat*, *mehrmals in diesem Semester*, *einmal in diesem Semester* und *in diesem Semester nicht eingesetzt* hatten.

- Wir sprechen die Prüfungsinhalte untereinander ab.
- Wir führen gemeinsame Vergleichsprüfungen durch.
- Wir korrigieren unsere Prüfungen gemeinsam.
- Wir sprechen die Lernzielvorgaben untereinander ab.
- Wir tauschen uns über die Leistungen der einzelnen Schülerinnen und Schüler aus.
- Wir diskutieren über die Beurteilungspraxis an unserer Schule.

Abbildung 9 zeigt die prozentualen Angaben der Lehrpersonen zu diesen Fragen im Vergleich der beiden Klassenstufen. Sowohl beim Lehrpersonal der 5. Klassen als auch der 9. Klassen ist im Bereich der Leistungsbeurteilung eine geringe Zusammenarbeit innerhalb der jeweiligen Kollegien erkennbar. Die häufigste Form der Zusammenarbeit ist der Austausch über die Leistungen der einzelnen Schülerinnen und Schüler. Auch wird häufiger über die Beurteilungspraxis an den Schulen diskutiert. Ca. 50 Prozent der Lehrpersonen gaben an, dies mehrmals im Semester zu tun. Auffällig selten kommt es vor, dass Lehrpersonen ihre Prüfungen gemeinsam korrigieren. Dafür wird aber von der Hälfte der Befragten angegeben, dass sie mindestens einmal im Semester Vergleichsprüfungen durchführen und sich über Prüfungsinhalte und Lernzielvorgaben abstimmen.



**Abbildung 9 (Zusammenarbeit zwischen Lehrpersonen zu Aspekten der Leistungsbeurteilung nach Klassenstufen. Angaben in Prozent.)**

Hinsichtlich der Zusammenarbeit im Kollegium ist kaum ein Unterschied zwischen den Klassenstufen erkennbar. Lediglich der Austausch über einzelne Schülerinnen und Schüler findet in der 9. Klasse etwas häufiger statt als in der 5. Klasse.

Die hier dargestellten Ergebnisse stehen in Einklang mit den Ergebnissen zu den Formen der Leistungsmessungen. Bereits im Kapitel 5 dieses Berichts, konnte festgestellt werden, dass Vergleichsarbeiten, insbesondere mündliche Vergleichsarbeiten (z.B. kriterienorientierte Sprechtests im Fremdsprachenunterricht) relativ selten sind. Die Zusammenarbeit bei der Leistungsbeurteilung ist nicht sehr stark ausgeprägt, auch wenn sie eine ganz fundamentale Voraussetzung für einheitliche Bewertungen darstellt. Auf der Grundlage dieses Befundes verwundert es nicht mehr, dass das Klassenbezugssystem immer noch eine wesentliche Rolle spielt und trotz einer proklamierten Tendenz zur Kriterienorientierung weiterhin einen starken Einfluss ausübt.

Es ist in diesem Zusammenhang auf die Schlussfolgerungen von Brügelmann et al. (2006) hinzuweisen, nach denen eine Leistungsbewertung von mehreren involvierten Personen vorgenommen werden sollte und ein kollegialer Austausch notwendigerweise erfolgen sollte. Die Ergebnisse zeigen zwar, dass eine kleine Gruppe von Lehrpersonen tatsächlich sehr stark um Zusammenarbeit bemüht ist, tendenziell kommt der Kooperation im Kollegium jedoch eine relativ geringe Bedeutung zu.

Zusätzlich zur Beantwortung der geschlossenen Fragen konnten die Lehrpersonen wieder in Freitextantworten Auskunft darüber geben, welche anderen Formen der Zusammenarbeit in Bezug auf die Beurteilungspraxis gepflegt werden. Die am häufigsten genannten Arten der Zusammenarbeit waren dabei Besprechungen mit einzelnen Kolleginnen oder Kollegen bzw. mit Fachpersonen wie beispielsweise Experten der Schulischen Heilpädagogik. Ebenfalls wurden Besprechungen im Kollegium genannt, um sich über die Beurteilungspraxis auszutauschen und abzustimmen. Neben diesen Einzel- und Gruppenbesprechungen ist für einzelne Personen auch der Austausch von Bewertungsbögen, Prüfungskriterien und Materialien als eine Form der kollegialen Zusammenarbeit relevant. Einige Personen nannten weiterhin Tandemprojekte, die Zusammenarbeit im Rahmen standardisierten Vergleichsprüfungen und ähnliches.

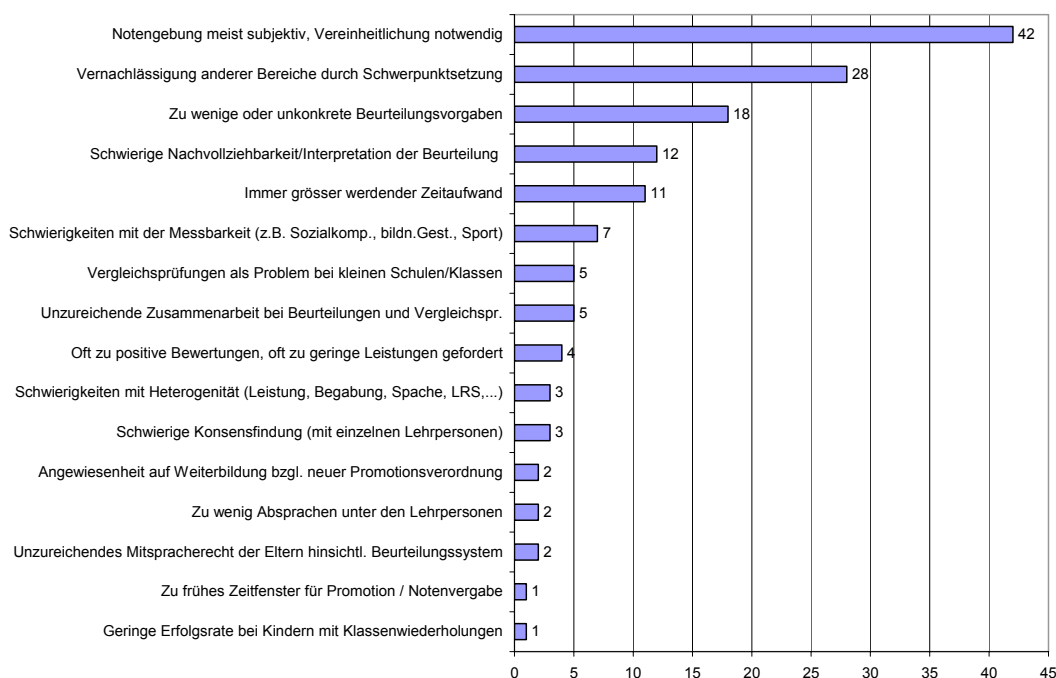
## 8. Probleme und Herausforderungen der Leistungsbeurteilung

Zusätzlich zu den Beurteilungen der Antwortvorgaben wurde den befragten Lehrerinnen und Lehrern die Möglichkeit gegeben, ihre Meinungen darüber zu notieren, welche Probleme und Herausforderungen Sie im Zusammenhang mit der Beurteilungspraxis in der Schule sehen. Abschliessend wurden die Lehrpersonen dazu befragt, welche Unterstützung hinsichtlich dieser Thematik sie sich von der jeweiligen Schulleitung und vom Departement Bildung, Kultur und Sport (BKS) wünschen.

Durch diese offenen Fragen konnte eine Vielzahl von Nennungen und Meinungen gewonnen werden. Diese wurden nach inhaltlichen Gesichtspunkten gebündelt, sortiert und kategorisiert, um sie in einem Folgeschritt quantifizieren zu können. Somit lassen sich konsistentere Aussagen als gegenüber einer Einzeldarstellung der Antworten gewinnen. In den beiden nachstehenden Abschnitten werden die kategorisierten Nennungen dargestellt, die nach Häufigkeit absteigend geordnet sind und bei denen somit die für das Lehrpersonal wichtigsten Punkte in den Abbildungen an oberster Stelle stehen.

### 8.1. Probleme

Die Frage zu den Problemen haben ca. drei Viertel aller in dieser Studie Befragten beantwortet. Insgesamt konnten aus den Freitextnennungen zur Frage nach den Problemen der Leistungsbeurteilung seitens der Lehrerinnen und Lehrer 17 Kategorien herauskristallisiert werden, die in Abbildung 10 dargestellt sind.



**Abbildung 10 (Kategorien der Freitextnennungen zu Problemen im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.)**

Die grösste Schwierigkeit liegt den Angaben der Lehrpersonen nach in der Subjektivität der Leistungsbeurteilung. So gibt fast jede dritte Lehrperson, die diese Frage beantwortet hat, zu, in ihren Bewertungen oft nach eigenem Ermessen zu handeln. Begründet wird dieser Tatbestand mit einer fehlenden Vereinheitlichung und Standardisierung der Beurteilungen und dass man oft nicht wisse, welche Kriterien konkret zur Leistungsbeur-

teilung herangezogen werden sollten. Dazu meinen 18 Personen, dass zu wenig oder nur unkonkrete Beurteilungsvorgaben vorhanden seien. Hier müsse vor allem die Frage geklärt werden, welche Lernziele in welchem Masse für eine Promotion von den Schülerinnen und Schülern erreicht werden müssen. 22 Personen sprechen ferner von einer Vernachlässigung anderer Bereiche durch eine Schwerpunktsetzung in der Leistungsbeurteilung hinsichtlich Promotion und Remotion. Dazu zählten vor allem die (individuelle) Förderung, das Eingehen auf den Lernfortschritt der Schülerinnen und Schüler und sogar eine Vernachlässigung einiger Unterrichtsfächer. Weiterhin bemängeln diese Lehrpersonen eine unzureichende Förderung von Schülerinnen und Schülern mit Legasthenie und Dyskalkulie.

## 8.2. Herausforderungen

Die Frage zu den Herausforderungen haben ca. 70 Prozent aller in dieser Studie Befragten beantwortet. Aus den Nennungen zu den Herausforderungen bezüglich der Leistungsbeurteilung der Lehrpersonen konnten 15 Kategorien gebildet werden. Wie die Abbildung 11 zeigt, korrespondieren die Aussagen mit jenen zu den Problemen (s.o.). Die grösste Herausforderung besteht für das Lehrpersonal in den noch nicht konkretisierten und standardisierten Beurteilungsvorgaben und Beurteilungskriterien (28 Nennungen). Hierzu bedarf es dringend einer besseren Informationsgrundlage für die Lehrerinnen und Lehrer, um einschätzen zu können, welche Lernziele in welchem Masse für eine Promotion ihrer Schülerinnen und Schüler erreicht werden müssen.



**Abbildung 11 (Kategorien der Freitextnennungen zu Herausforderungen im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.)**

Weiterhin sprechen 20 Personen von Schwierigkeiten beim individuellen und förderorientierten Unterrichten. Problematisiert wird dabei der Gegensatz zwischen Individualität und Förderung auf der einen Seite und Standardisierung von Leistung auf der anderen Seite. Die Transparenz der Beurteilungen (14 Nennungen) und die Objektivität (7 Nen-

nungen) sowie das Bestreben, richtige und gerechte Empfehlungen für den Schulübertritt zu geben, sind ebenfalls wichtige Themen für viele Lehrpersonen.

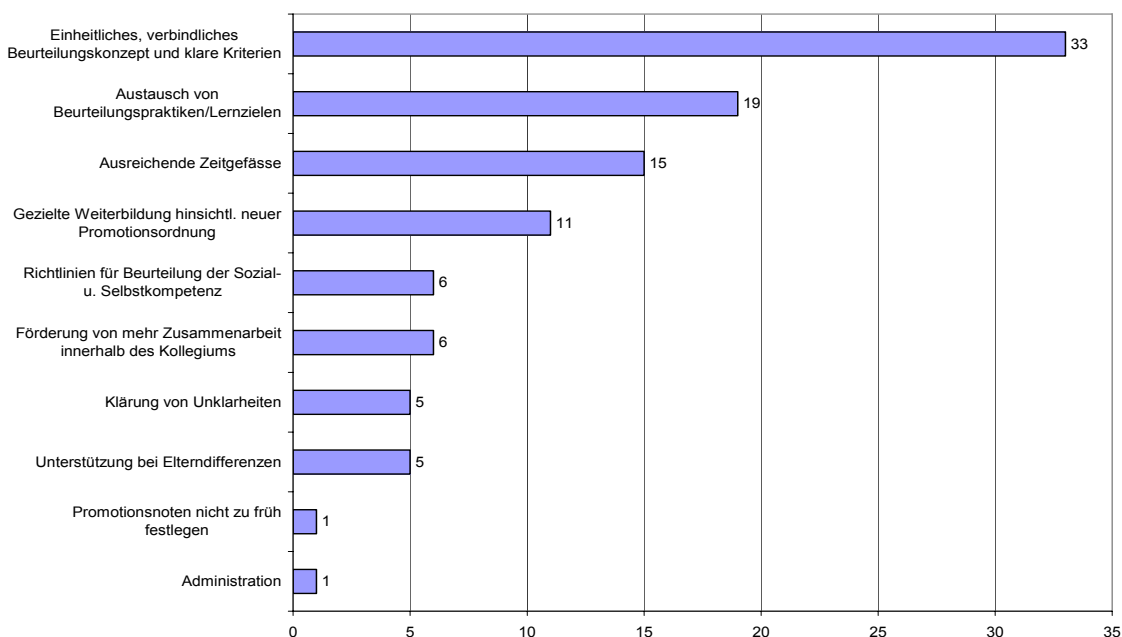
Als Herausforderung wird weiterhin auch die Zusammenarbeit im Team genannt, wobei gleichzeitig ebenfalls der Zeit- und Arbeitsaufwand betont wird. Es scheint auf der Grundlage der aus dieser Studie erzielten Ergebnisse tatsächlich eine grosse Herausforderung zu sein, die Möglichkeiten einer fruchtbaren Kooperation zwischen den Lehrpersonen so auszubauen, dass damit der Arbeitsaufwand nicht erhöht sondern reduziert wird.

## 9. Unterstützungsmassnahmen

In einer abschliessenden Freitextfrage wurde den Lehrpersonen die Möglichkeit gegeben, ihre Ansicht über nützliche und sinnvolle Unterstützungsmassnahmen von seiten der Schulleitung und des Departements Bildung, Kultur und Sport (BKS) zu äussern. Wie in den beiden folgenden Abschnitten dargestellt, sind den Lehrpersonen dabei vor allem transparente und handhabbare Instrumente und die Möglichkeit zu umfassender Information wichtig.

### 9.1. Unterstützung seitens der Schulleitung

Die Frage zur Unterstützung seitens der Schulleitung haben ca. 60 Prozent aller in dieser Studie Befragten beantwortet. Es fallen bei der Auswertung dieser Frage erneut besonders die Unsicherheiten zu den Beurteilungen auf: 33 Personen wünschen sich klare Beurteilungsrichtlinien und ein einheitliches und für alle Lehrpersonen verbindliches Beurteilungskonzept. Als weitere Unterstützung wäre den Lehrpersonen der Austausch von Beurteilungspraktiken und den zugrundeliegenden Lernzielen sehr wichtig. Damit ist die Kooperation sowohl mit der Schulleitung als auch innerhalb des Kollegiums und auch mit Kolleginnen und Kollegen anderer Schulen gemeint. Ein nicht zu vernachlässigender Punkt ist die zeitliche Ressource: 15 Personen wünschen sich eine Unterstützung durch ausreichende Zeitgefässe für die neue Beurteilungspraxis. 11 Personen sprechen sich darüber hinaus für gezielte Fort- und Weiterbildungen zu der neuen Promotionsverordnung aus (vgl. Abbildung 12).

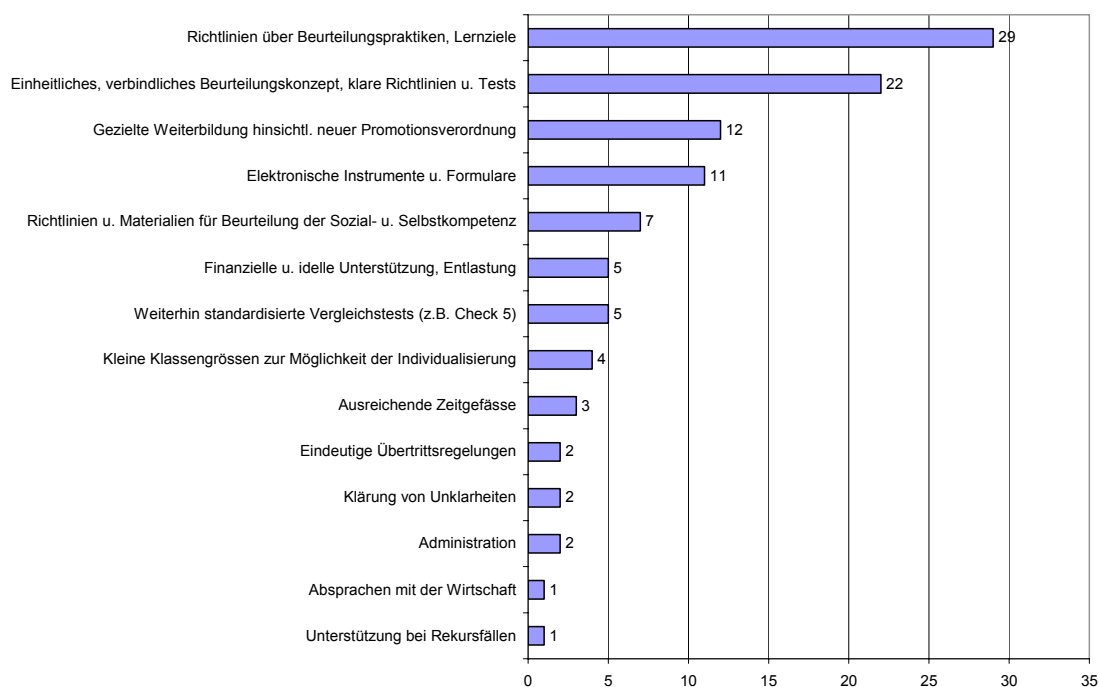


**Abbildung 12 (Kategorien der Freitextnennungen zur gewünschten Unterstützung seitens der Schulleitung im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.)**

### 9.2. Unterstützung seitens des BKS

Die Frage zur Unterstützung seitens des Departements Bildung, Kultur und Sport (BKS) haben ca. zwei Drittel aller in dieser Studie Befragten beantwortet. Neben Unterstützungen durch die Schulleitungen sind auch solche durch das BKS gewünscht. Diese

korrespondieren zum Teil mit denen, der in Abschnitt 9.1 genannten Unterstützungsleistungen. So werden Richtlinien über Beurteilungspraktiken und die zu beurteilenden Lernziele von 29 Personen erwartet. Diese sollten möglichst klar und verständlich sein und nach Möglichkeit auch in einheitliche – standardisierte – Leistungstests münden. 22 Personen würden sich von Seiten des BKS besonders darin unterstützt sehen, wenn sie ein einheitliches und verbindliches Beurteilungskonzept und klare Richtlinien sowie Tests zur Verfügung gestellt bekämen. Auch eine gezielte Weiterbildung zur neuen Promotionsverordnung wurde von 12 Personen als Wunsch geäußert. Mit den oben angesprochenen zeitlichen Ressourcen korrespondiert auch der Wunsch nach Vereinfachung der Beurteilungsdurchführung: 11 Personen sprechen sich für eine elektronische Erfassung durch computergestützte Instrumente und Formulare der Beurteilungen aus (vgl. Abbildung 13).



**Abbildung 13 (Kategorien der Freitextnennungen zur gewünschten Unterstützung seitens des BKS im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.)**

Wie sich zeigt, bestand zum Erhebungszeitpunkt im Februar 2009 bei den Lehrpersonen hinsichtlich der neuen Promotionsverordnung grösstenteils noch Informationsbedarf. Vor allem jedoch in der Anwendbarkeit, der für die neue Beurteilungspraxis zur Verfügung stehenden Zeitgefässe und den noch fehlenden klaren und eindeutigen, standardisierten Bewertungsrichtlinien. Diese Unsicherheiten und Fragen, so der Wunsch aus den Reihen der Lehrerinnen und Lehrer, sollten durch Schulungen bzw. Fort- und Weiterbildungen verringert werden.

## 10. Zusammenfassung

Die Situationsanalyse der Beurteilungspraxis von Aargauischen Lehrpersonen zeigt einige interessante Aspekte darüber auf, wie Lehrpersonen die Leistungen ihrer Schülerinnen und Schüler messen, beurteilen und ausweisen.

Es konnte festgestellt werden, dass in den 5. und 9. Klassen die Sprachen und die Mathematik zu den meistgeprüften Unterrichtsfächern gehören. In den künstlerisch-gestalterischen Fächern sowie Schulsport und Ethik- und Religion wird am wenigsten geprüft. Die Lehrpersonen orientieren sich vor allem an dem Anspruch, möglichst viele unterschiedliche Bereiche und Kompetenzen abzudecken, um die Zeugnisnoten zu untermauern. Ausrutscher sollen ebenfalls nicht zu negativ bewertet werden. Die Lehrpersonen orientieren sich in der überwiegenden Anzahl der Fälle an der laut Promotionsverordnung vorgegebenen Anzahl der Prüfungen als Mindestmass. Dieses Mindestmass wird nur in seltenen Fällen unterschritten, in der Regel prüfen die Lehrpersonen häufiger.

Die Festlegung der Prüfungsnoten erfolgt nach den Aussagen der Lehrpersonen meist kriterienorientiert. Es kann jedoch vermutet werden, dass in der Praxis ebenfalls die Orientierung am Klassenbezugssystem eine nicht unwichtige Rolle spielt, da die Rückmeldungen an die Schülerinnen und Schüler meist die Darstellung der Leistungen im Spektrum der Klassenleistung beinhaltet (vgl. Kapitel 6). Sehr häufig werden vor allem in den 9. Klassen die Prüfungsnoten streng mathematisch gerundet zu einer Zeugnisnote zusammengefasst, was einer realistisch unwahrscheinlichen identischen Wertung aller Prüfungen gleichkommt. Bei der Vergabe der Zeugnisnoten spielt neben dem Durchschnitt der Prüfungsnoten häufig auch die Selbstkompetenz eine Rolle. Diese wird eher selten semesterbegleitend bewertet, sondern kommt in der Regel erst am Ende zum Tragen, wenn ein Schüler oder eine Schülerin zwischen zwei Noten steht. In der 5. Klasse ist zusätzlich das Gewicht für die Promotion und Remotion von einiger Bedeutung.

Als Messinstrument für schulische Leistungen steht die schriftliche Klassenprüfung an oberster Stelle. Daneben werden vor allem schriftliche Arbeiten für die Benotung herangezogen, wozu Aufsätze, Projektarbeiten und das Bearbeiten von Hausaufgaben zählen. Mündliche Prüfungen kommen seltener zum Einsatz. Es ist ein Gefälle bezüglich der Klassenstufen erkennbar. In den 5. Klassen wird generell häufiger geprüft als in den 9. Klassen.

Bezüglich der Frage, inwiefern Leistungsergebnisse als Ausgangslage zu weiteren Fördermassnahmen genutzt werden, wird für beide Klassenstufen deutlich, dass generell der Nutzen der Ergebnisse aus Leistungsmessungen hoch eingeschätzt wird. Tendenziell werden den Leistungsmessungen viele unterschiedliche Funktionen zugeschrieben, jedoch besteht bei den Lehrpersonen eine deutliche Sensibilität dafür, einzelne Funktionen nicht zu vermischen. Lehrpersonen, denen die individuelle Förderung und die Lernzielanpassung wichtig sind, legen in der Regel weniger Wert auf Promotion und Selektion sowie auf eine Qualitätskontrolle ihres Unterrichts. Häufig geben die Lehrpersonen eine Rückmeldung zum jeweiligen Leistungsstand in der Klasse. Ausführliche mündliche Gespräche finden dabei häufiger Anwendung als schriftliche Rückmeldungen.

Für die Zusammenarbeit innerhalb der Schulen im Bereich Leistungsbeurteilung ist eine geringe Ausprägung zu erkennen: in beiden untersuchten Klassenstufen findet sie ab-

gesehen vom Austausch über einzelne Schüler nur selten statt. Selten wird ausserdem über die Beurteilungspraxis in den Schulen diskutiert.

Aufgrund dieser Befunde erklären sich deutlich die Probleme, Herausforderungen und Unterstützungswünsche, die die Lehrpersonen formulieren. Sie problematisieren häufig die unklaren Beurteilungskriterien und wünschen sich Unterstützung in Form klarer Hinweise auf Kompetenzbeurteilungen und Standards. Weiterhin wird der Wunsch nach handhabbaren, transparenten Instrumenten zur Leistungsmessung und der Bedarf nach Weiterbildung geäussert.

Der wohl eindrücklichste Befund dieser Untersuchung ist die innerhalb der Schulen schwach ausgeprägte Zusammenarbeit in Fragen der Leistungsmessung und -beurteilung. Schon seit langer Zeit ist bekannt, dass Beurteilungen heterogen ausfallen können. Daniel Starch hat in diesem Zusammenhang bereits 1913 ausgeführt, dass für heterogene Bewertungen vor allem Unterschiede zwischen Anforderungen verschiedener Schulen, Unterschiede im Ausbildungsstand verschiedener Lehrpersonen, die unterschiedliche Betonung des relativen Wertes verschiedener Elemente in Prüfungsarbeiten und Unterschiede aufgrund blosser Unfähigkeit verantwortlich seien. Dazu kommt nach Ingenkamp (1971) das klasseninterne Bezugssystem und die Subjektivität der Beurteilungen. All diese Schwächen können durch einen konstruktiven Austausch und die Zusammenarbeit bei der Definition von Beurteilungskriterien, bei der Durchführung von Prüfungen und Vergleichsarbeiten und bei anderen Fragen der Beurteilungspraxis verringert werden. Die von Brügelmann et al. (2006) ausgesprochenen Vorschläge zu einer Sensibilisierung der Lehrpersonen, nach fachdidaktisch begründeten Kriterien und nach einem kontinuierlichen kollegialen Austausch, wird aufgrund der Befunde dieser Studie am stärksten gestützt. Die befragten Lehrpersonen äussern neben dem Wunsch nach einer besseren kollegialen Zusammenarbeit auch den Wunsch nach weiteren Zeitressourcen. Es scheint, als wäre ein grosses Potenzial bei den Lehrpersonen vorhanden, das durch eine gezielte Unterstützung und konkrete Massnahmen nutzbar gemacht werden kann. Die wichtigste dieser Massnahmen sollte die Unterstützung bei der kollegialen Zusammenarbeit sein. Zeitressourcen spielen dabei sicherlich eine wichtige Rolle, wobei ebenfalls zu bedenken ist, dass Zusammenarbeit längerfristig auch über die Nutzung von Synergien zu einer Zeitersparnis führen kann. Weiterhin bedarf es konkreter verständlicher Beurteilungsvorgaben wie Kompetenzraster, Beobachtungslisten oder Instrumente zur Selbstbeurteilung. Auch einheitliche Prüfungs- und Übungsaufgaben sowie Aufgaben für interne Tests oder Vergleichsarbeiten sind mögliche Unterstützungen für Lehrpersonen in ihrer Beurteilungspraxis. Nicht zuletzt könnten, wie es u.a. auch Brügelmann et al. (2006) empfehlen, gezielte Weiterbildungen den Lehrpersonen helfen, ihre Beurteilungs- und Bewertungskompetenzen weiter auszubauen. In dieser Hinsicht würden Bestrebungen um die Bereitstellung von standardisierten und halbstandardisierten Instrumenten zum Messen, Beurteilen und Ausweisen von Leistung in die Richtung weisen, die auch von Lehrpersonen eingefordert wird.

## 11. Literatur

- Brügelmann, H., Backhaus, A., Brinkmann, E., Coelen, H., Franzcowiak, T., Knorre, S., Müller Naendrup, B., Oser, E., Roth, S. (2006). Sind Noten nützlich - und nötig?: Ziffernzensuren und ihre Alternativen im empirischen Vergleich. Frankfurt: Grundschulverband.
- Coffman, W. E. (1971). Essay examinations. In: Thorndike, R. L. (Hg.) *Educational Measurement*. Washington: American Council on Education.
- Eberle, F., Metzger, C. (1988). Gültigkeit und Zuverlässigkeit von Aufsatzkorrekturen am Beispiel der Zulassungsprüfung für ausländische Studienbewerber der Hochschule St. Gallen. In: *Studien und Berichte des IWP*, Heft 7, St. Gallen: Institut für Wirtschaftspädagogik.
- Ingenkamp, K. (Hg.)(1971). *Die Fragwürdigkeit der Zensurengebung: Texte und Untersuchungsberichte*. Weinheim, Basel: Beltz.
- Starch, D. (1913). The Reliability and distribution of grades. In: *Science*, 38, 630-636.
- Vögeli-Mantovani, U. (1999). Mehr fördern, weniger auslesen. Zur Entwicklung der schulischen Beurteilung in der Schweiz. (SKBF Trendbericht, Nr. 3) Aarau: SKBF.
- Ziegenspeck J. (1999). Handbuch Zensur und Zeugnis in der Schule. Historischer Rückblick, allgemeine Problematik, empirische Befunde und bildungspolitische Implikationen. Ein Studien- und Arbeitsbuch. Bad Heilbrunn/Obb: Klinkhardt.

## 12. Abbildungsverzeichnis

Abbildung 1 (Anzahl Klassenprüfungen pro Semester nach Unterrichtsfächern in der 5. Klasse und in der 9. Klasse in Prozent) .....	9
Abbildung 2 (Beweggründe zur Festlegung der Anzahl von Klassenprüfungen nach Klassenstufen).....	11
Abbildung 3 (Angaben zu der Aussage "Ich entwickle vor der Durchführung der Prüfungen ein Kriterienraster, in dem auch die Noten für die jeweiligen zu erwartenden Leistungen bereits festgelegt sind" in Prozent nach Klassenstufen.) .....	13
Abbildung 4 (Angaben zu der Aussage "Wenn ich die Zeugnisnote festlege, runde ich den Durchschnitt der Klassenprüfungsnoten streng mathematisch" in Prozent nach Klassenstufen.).....	17
Abbildung 5 (Angaben zu der Frage "Welche Leistungsbeurteilungen haben Sie im 1. Semester des Schuljahres 2007/2008 zur Feststellung der schriftlichen Leistungen pro Klasse wie häufig eingesetzt?" in Prozent nach Klassenstufen.).....	20
Abbildung 6 (Angaben zu der Frage "Welche Leistungsbeurteilungen haben Sie im 1. Semester des Schuljahres 2007/2008 zur Feststellung der mündlichen Leistungen pro Klasse wie häufig eingesetzt?" in Prozent nach Klassenstufen.).....	21
Abbildung 7 (Angaben zu der Frage "Welche Leistungsbeurteilungen haben Sie im 1. Semester des Schuljahres 2007/2008 zur Feststellung der Selbstkompetenz und der Sozialkompetenz pro Klasse wie häufig eingesetzt?" in Prozent nach Klassenstufen.).....	22
Abbildung 8 (Angaben zu der Frage "Wie wichtig sind Ihnen die Leistungsergebnisse?" nach Klassenstufen. Mittelwerte der fünfstufigen Antwortskala. 1=gar nicht wichtig, 5=sehr wichtig.) .....	24
Abbildung 9 (Zusammenarbeit zwischen Lehrpersonen zu Aspekten der Leistungsbeurteilung nach Klassenstufen. Angaben in Prozent.) .....	28
Abbildung 10 (Kategorien der Freitextnennungen zu Problemen im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.) .....	30
Abbildung 11 (Kategorien der Freitextnennungen zu Herausforderungen im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.).....	31
Abbildung 12 (Kategorien der Freitextnennungen zur gewünschten Unterstützung seitens der Schulleitung im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.).....	33
Abbildung 13 (Kategorien der Freitextnennungen zur gewünschten Unterstützung seitens des BKS im Zusammenhang mit der Beurteilungspraxis in der Schule. Angaben in Prozent der Nennungen.).....	34

## 13. Tabellenverzeichnis

Tabelle 1 (Schulstufe/ Schultyp, an der die Befragten im 1. Semester 2007/2008 hauptsächlich unterrichten.) .....	7
Tabelle 2 (Durchschnittliche Anzahl Klassenprüfungen pro Semester nach Unterrichtsfächern in der 5. Klasse und in der 9. Klasse. Anzahl der Lehrpersonen, Mittelwert und Standardabweichung, Mindestanzahl Prüfungen laut Promotionsverordnung).....	8
Tabelle 3 (Antworten zur Festlegung der Prüfungsnoten in den 5. und 9. Klassen. Anzahl Lehrpersonen (N), Mittelwerte (M) und Standardabweichungen (SD). 1=nie, 5=immer, **= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,01$ ).).....	15
Tabelle 4 (Antworten zur Festlegung der Zeugnisnoten in den 5. und 9. Klassen. Gültige Fälle, Mittelwerte und Standardabweichungen. 1=nie, 5=immer; **= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,01$ ).).....	16
Tabelle 5 (Nutzung der Leistungsmessungsergebnisse in den 5. und 9. Klassen. Gültige Fälle, Mittelwerte und Standardabweichungen. 1=gar nicht wichtig, 5=sehr wichtig, **/= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,01/ \alpha < 0,05$ ). ).....	25
Tabelle 6 (Korrelationen der Einschätzungen zu den Formen des Umgangs mit der der Leistungsmessung.) .....	26
Tabelle 7 (Rückmeldungen der Leistungsmessungsergebnisse in den 5. und 9. Klassen. Gültige Fälle, Mittelwerte und Standardabweichungen. 1=nie, 5=immer, *= signifikante Unterschiede zwischen 5. und 9. Klassen ( $\alpha < 0,05$ ).).....	26