

## 7.3 Beurteilungsfehler

Die Europäische Kommission stellt fest:

„Die Beurteilung der Schüler ist eine problematische Angelegenheit; die Vorurteile, von denen Lehrer bestimmt sein können, sind heute eindeutig identifiziert. Das gravierendste Problem ist seit 1947 unter der Bezeichnung „Posthumuseffekt“ bekannt: Unabhängig davon, wie zu Beginn des Schuljahres die Fähigkeiten der Schüler einer Klasse verteilt sind, entspricht die Notengebung am Ende im Großen und Ganzen einer Gaußschen Verteilung.“ (Europäische Gemeinschaften - Kommission, 1994).

Kalthoff drückt diesen Sachverhalt wie folgt aus:

„Für Lehrpersonen ist es undenkbar, alle Schüler auf nur ein oder zwei Rangplätze anzusiedeln; es gibt in ihrer Vorstellung einfach gute und weniger gute Schüler“ (Kalthoff, 1996, S. 114). Nicht Gleichheit sei das Ziel, sondern Differenz. „Für die Unterrichtspraxis heißt dies, dass Lehrpersonen nicht nur einzelne Schülerleistungen bewerten, sondern eine distinkte Verteilung der Schüler über eine Skala organisieren.“ (S. 121).

Nach dieser Regel neigen die Lehrer dazu, das Unterrichtsniveau und die Beurteilung der Schülerleistungen so zu gestalten, dass die Noten jedes Jahr ungefähr die gleiche Gaußsche Verteilung aufweisen. Folglich kann ein Schüler X mit mittlerer Begabung (festgestellt z. B. auf Grund eines Tests, an dem alle Schüler eines Jahrgangs teilnehmen) in Klassenarbeiten unterschiedliche

Resultate erzielen, je nachdem, ob er Klasse A oder Klasse B angehört (siehe [Abbildung 13](#)). In Klasse A würde der Schüler X eine weniger gute Beurteilung bekommen, in Klasse B dagegen eine bessere – bei gleicher Leistung. Derjenige Schüler, der eine Klasse mit durchwegs schwachen Mitschülern besucht, hat wahrscheinlich Chancen, überbewertet zu werden, ja, er wird möglicherweise sogar zu den besten Schülern zählen. Der andere hingegen, der eine Klasse mit mehrheitlich guten Schülern besucht, läuft Gefahr, unterbewertet oder sogar vom Lehrer für nicht geeignet befunden zu werden, in die nächste Klasse vorzurücken.

Grundsätzlich muss man befürchten, dass die Lehrer ihre Beurteilung an Klassennormen ausrichten statt an Lernzielen. Tatsache ist, dass man sich auch im Schulwesen durch die Methoden der Psychometrie daran gewöhnt hat, die von Schülern in Wissens- oder Intelligenztests erzielten Leistungen im Hinblick auf ihre statistische Verteilung zu interpretieren: Die Leistung eines Individuums wird in Bezug auf andere beurteilt. Diese Methode ist üblich geworden. Heute geht die Mehrzahl der Lehrkräfte davon aus, dass die Bewertung von Schülerarbeiten zu einer Rangliste der Schüler und Schülerinnen führen soll.

Diese Bewertungsmethoden können nach Perrenoud als Ausdruck einer Kultur der Konkurrenz gelten. Ihm zufolge bildet die Schule eine Hierarchie des Erfolges heraus. In unserer Vorstellung aber „kann der Erfolg nur gesellschaftlichen Wert haben, wenn er nicht allen offen steht“ (Perrenoud, 1984, S. 70). Mehr noch,

„in jeder Schule hat sich ein ‘meritokratisches’ Modell herausgebildet [...], wonach denjenigen, die den meisten Erfolg errungen haben, die höchste Anerkennung gebührt, nachdem alle dieselben Bildungschancen hatten. Die Hierarchie des Erfolgs erscheint legitim und unantastbar und kann sich sogar in eine moralische Hierarchie verwandeln, vor allem dann, wenn der Erfolg in erster Linie vom Fleiß der Schüler, ihrer Bereitschaft zur Disziplin und ihrer Ausdauer abhängt“ (S. 81).

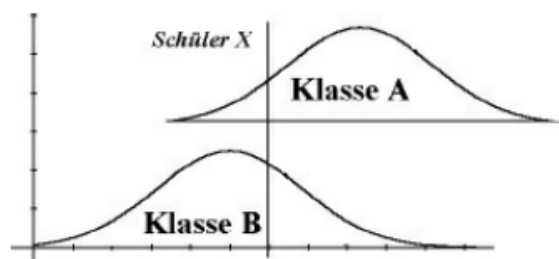


Abbildung 13: Leistungen des Schülers X und Klassenzugehörigkeit

Die Vorstellung, dazu verpflichtet zu sein, die guten und die schlechten Schüler zu unterscheiden und die Absicht, eine Rangliste der Schüler zu erstellen, führt die Lehrer zwangsläufig dazu, das Unterscheidende in den Vordergrund zu rücken und für jede Aufgabe eine oft künstliche

Werteskala aufzustellen, die eben im Idealfall zu einer Gaußschen Verteilung der Noten führt. Die Alternative liegt natürlich darin, den Klassenbezug (soziale Bezugsnorm) völlig zu ignorieren und Anforderungen von außen vorzugeben, z. B. durch Kompetenzraster.

Zahlreiche Untersuchungen weisen nach, dass die Entscheidung der Lehrer, einen Schüler die Klasse wiederholen zu lassen, in erheblichem Umfang von Willkür bestimmt ist. Beschränkt auf den Mikrokosmos ihrer Klasse und von einer Ideologie des Erfolgs geleitet, neigen die Lehrer dazu, den Unterricht an den besten Schülern auszurichten und damit Anforderungen aufzustellen, die die Lehrpläne übersteigen.

Im Folgenden seien solche und andere typische Fehler skizziert, die im schulischen Bewertungsprozess eine Rolle spielen (nach Weise, 1990).

Fall 1: Eine Lehrkraft nutzt nicht die ganze Notenspanne, sondern vermeidet die besseren und schlechteren Noten. Die Leistungsbeschreibung selbst kann dabei korrekt gewesen sein.

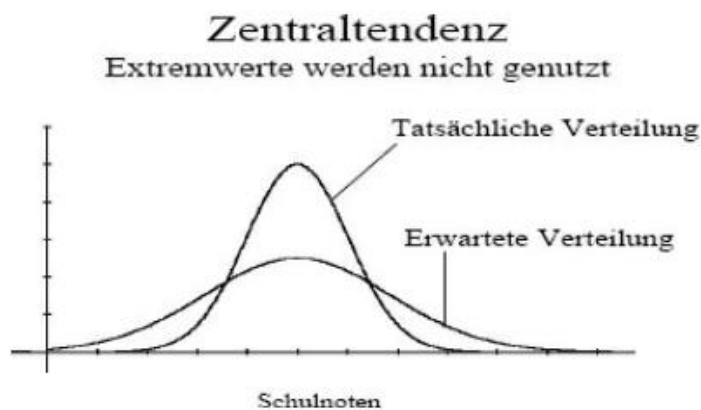


Abbildung 14: Zentraltendenz: Extremwerte werden nicht genutzt (nach Weise, 1990)

Fall 2: Eine Lehrkraft vergibt eine bessere Note, weil ihm/ihr der Schüler/die Schülerin sympathisch ist. Dabei kann die Leistungsbeschreibung durchaus richtig gewesen sein.

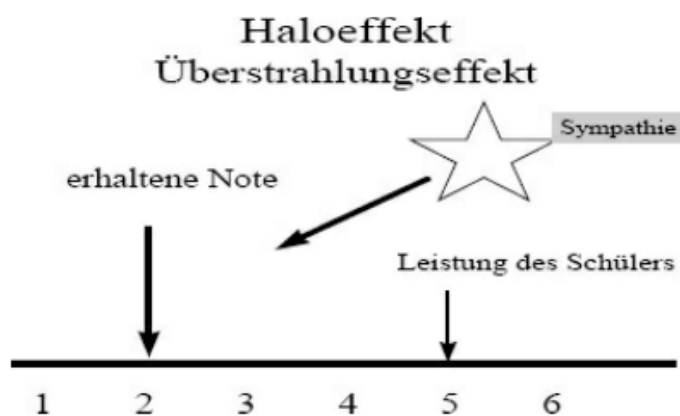


Abbildung 15: Haloeffekt: Sympathie führt zur besseren Note

Fall 3: Eine Lehrkraft vergibt im Vergleich zu anderen Lehrkräften bei gleicher Leistungsbeschreibung bessere Noten.

### Milde-Streng-Tendenz Gesamte Skala wird verschoben

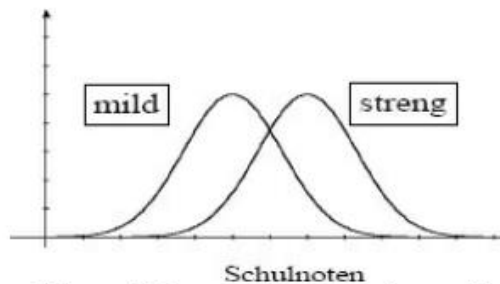


Abbildung 16: Fehler Milde-Streng-Tendenz: die gesamte Skala wird verschoben

Fall 4: Eine Lehrkraft mit den Fächern Latein und Mathematik vergibt wegen der guten Lateinleistungen ihres Schülers/ihrer Schülerin eine bessere Note in Mathematik, als es der Leistungsbeschreibung entspricht.

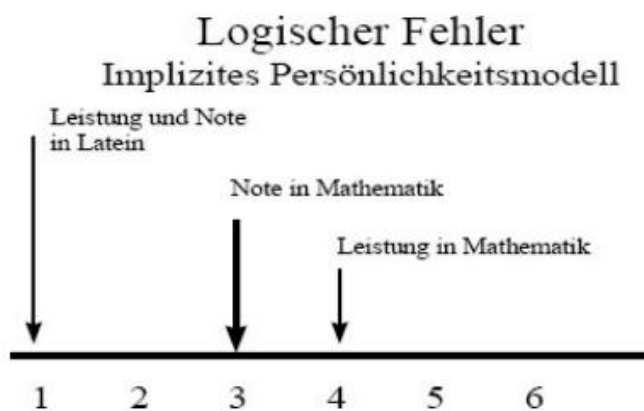


Abbildung 17: Logischer Fehler

Fall 5: Eine sprachbegabte Lehrkraft vergibt schlechtere Noten als eine sprachlich durchschnittlich begabte Lehrkraft, weil die eigene Sprachleistung als normal angesehen wird.

### Kontrast-/Ähnlichkeitsfehler schließen von sich auf andere

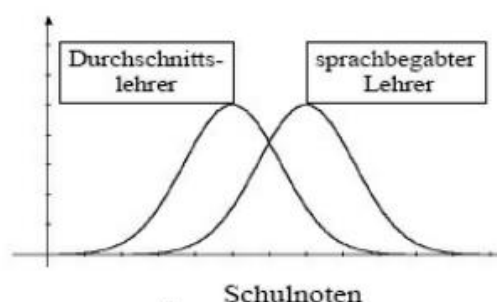


Abbildung 18: Kontrast-/Ähnlichkeitsfehler

Fall 6: Eine Lehrkraft für Mathematik und Biologie vergibt eine gute Note in der zweiten Stunde Mathematik, weil dieser Schüler/diese Schülerin in der ersten Stunde Biologie so geblüht hat.

### Nähe-Fehler zeitliche Nähe

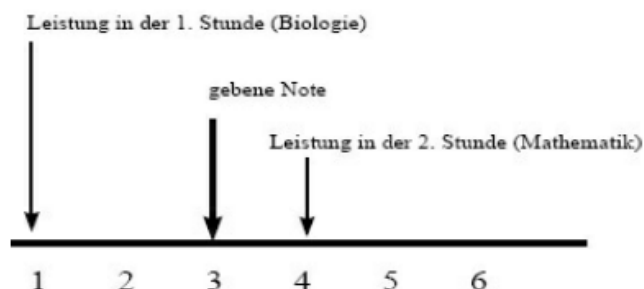


Abbildung 19: Nähe-Fehler

Fall 7: Eine Lehrkraft verändert seine/ihre Note zum Positiven, weil ansonsten der Schüler/die Schülerin sitzenbleiben würde. Sie tut dies, weil die Gesamtprognose positiv ausgefallen ist.



Abbildung 20: Wissen-um-die-Folgen-Fehler

An diesem Fehler wird die Kuriosität der Notengebung ganz offensichtlich: Die Note entspricht nicht mehr der Leistung, weil nämlich die Lehrkraft etwas pädagogisch Sinnvolles tut: Sie geht bei der Beurteilung von der Gesamtsituation des Schülers/der Schülerin aus (Beispiele bei Kalthoff, 1996). Dabei wird sehenden Auges gegen die korrekte Notengebung verstoßen.



Fall 8: Eine Lehrkraft vergibt weniger gute Noten, als es dem Leistungsstand ihrer Klasse entspricht, weil z. B. ihrer Ansicht nach nicht so viele Schüler auf das Gymnasium sollen. Hier wird die Lehrkraft zum Schicksalsbestimmer, die dadurch ihre eigenen Gesellschaftsbilder transportiert und damit ihre Unfähigkeit für diesen Beruf beweist. Umgekehrte Tendenz: Eine Lehrkraft vermeidet schlechte Noten, weil sie richtig erkannt hat, dass schlechte Noten keineswegs motivierend wirken

(man vergleiche die Notendurchschnitte verschiedener Fächer miteinander!).

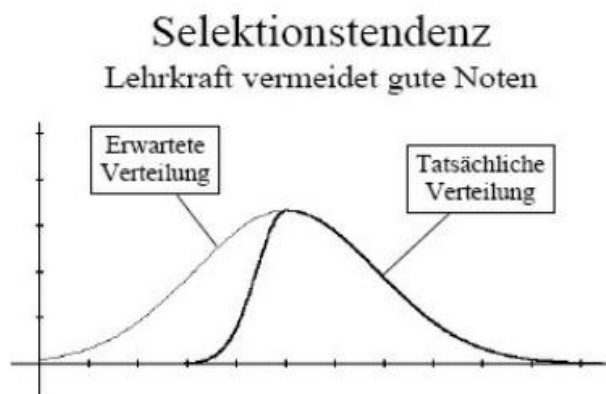


Abbildung 21: Fehler Selektionstendenz

Bei den meisten Fehlern kann es durchaus sein, dass die Leistungsfeststellung korrekt war, die -beurteilung aber fehlerhaft ist, also den bereits diskutierten Gütekriterien widersprechen. Die Ursachen hierfür sind vielfältig. Ein Ursachenbündel liegt allerdings in dem Rollenkonflikt, den die Lehrkräfte zu ertragen haben: Gleichzeitig Helfer und Beurteiler sein zu müssen. Dies führt zu teilweise verständlichem, da systembedingtem, Ausweichverhalten.

Die genannten Fehler wurden alle am Beispiel der Notengebung erläutert, gelten aber auch für notenfreie Verfahren. Dies bedeutet, dass man durch Abschaffung der Note nicht zwingend Beurteilungsfehler vermeidet und dies deshalb auch nicht vernachlässigen darf. Dies muss

Aufforderung sein, derartige Fehler auf unterschiedlichen Ebenen der Leistungsfeststellung und -beurteilung anzugehen. Angezielt werden muss eine „kontrollierte Subjektivität“ wie es Kleber (1992) nannte.

Es bleibt allerdings die Frage, ob derartige Reparaturversuche erfolgreich sein können. Es hat sich nämlich herausgestellt, dass Lehrkräfte die Leistungen ihrer Schüler fast durchgängig überbewerten.

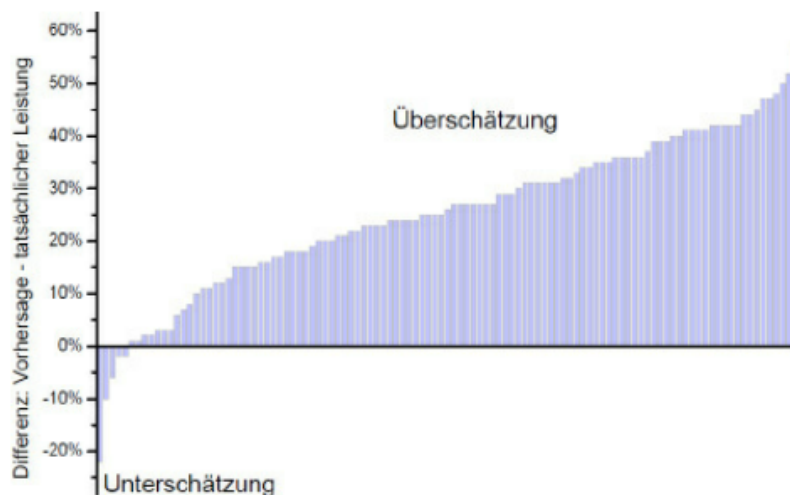


Abbildung 22: Tendenzen der Über- und Unterschätzung im Lehrerurteil (Helmke, Hosenfeld, Scherthan & Wagner, 2004)

Man muss sich also die Frage stellen, ob alle Optimierungsversuche, das Lehrerurteil zu verbessern, überhaupt zum Erfolg führen können. Es ist daher zu prüfen, ob die völlige Auslagerung der Beurteilung der Schülerinnen und Schülern aus dem Bereich der Lehrtätigkeit zumindest im Fachleistungsbereich erfolgen müsste. Objektive, von außen gegebene und transparente Kompetenzanforderungen könnten Grundlage von derartigen Schülerbeurteilungen (Mini-Tests) sein.