

Ist „gut“ noch „ausreichend“?

Warum gleiche Mathematik-Ergebnisse ungleich zensiert werden:
Auszug aus dem Schröter-Buch „Zensuren? Zensuren!“

Die Aufsatzbeurteilung kann als eine Art Extremfall angesehen werden; denn hierbei trifft Subjektivstes beim Schüler mit Subjektivstem beim Lehrer zusammen. Als entgegengesetztes Extrem wird man vielleicht das Fach Mathematik ansehen.

Hier gibt es, so könnte man vermuten, keinerlei oder zumindest doch keine gravierenden Unterschiede in der Einschätzung von Schülerleistungen, weil es nur um objektive Dinge geht. Wer $7 \times 8 = 56$ ausrechnet, hat überall in der Welt null Fehler, und Fragen des Stils, der Weltsicht, des persönlichen Geschmacks und der Mentalität spielen hierbei, so sollte man doch meinen, keine Rolle, wie es beim Aufsatz der Fall sein kann.

Schon beim ersten Versuch innerhalb meines Hochschulseminars, das sich mit Fragen der Zensurengebung befaßte, wurde deutlich, daß auch die Mathematikzensur problematisch ist. Eine meiner Mitarbeiterinnen nahm die Mathematikarbeit eines Schülers des 3. Schuljahres der Grundschule, in der es vor allem um Mengenlehreaufgaben ging...

Die Arbeit des Schülers Ralf wurde so, wie er sie dem Lehrer abzugeben hatte, vervielfältigt und mit der Bitte, sie zu zensieren, an 20 Studenten der letzten Semester an der Pädagogischen Hochschule Kiel, die alle das Wahlfach Mathematik studiert hatten, weitergegeben. Da die Kenntnis mancher Lehrer in dem Gebiet „Mengenlehre“ als Teil der neuen Mathematik doch noch ziemlich unvollkommen war, mußten die Studenten des Wahlfaches Mathematik, die bereits mehrere Praktika hinter sich hatten und inzwischen alle Mathematiklehrer sind, als besonders kompetent angesehen werden.

Wir waren überrascht, daß die Arbeit eine Bewertungsspanne von 3 Zensuren erhielt! Von „gut“ bis „ausreichend“ differierte die Einschätzung der Leistung!

Bei einer Feinuntersuchung der Bewertungsweisen gelangten wir zu der Feststellung, daß die Zensur eng verknüpft war mit der Punktzahl, die von vornherein jeder Aufgabe zugeteilt worden war. Es darf als bekannt vorausgesetzt werden, daß Lehrer in der Regel Mathematikarbeiten danach zensieren, wieviel der erreichbaren Punkte ein Schüler erlangt.

Man könnte nun annehmen, daß es für das Endresultat gleichgültig ist, ob für jede der Aufgaben zwei, fünf oder zehn Punkte von vornherein in Anschlag gebracht werden, da es ja letztlich nur auf den Anteil der Punkte ankommt, die man im Verhältnis zur erreichbaren Höchstpunktzahl erzielte.

Unsere Untersuchung zeigte aber, daß diese Vereinfachung falsch ist: Je weniger Punkte für eine Aufgabe vergeben wurden, um so schlechter war die Zensur, je mehr Punkte vorgegeben waren, um so günstiger war die Bewertung. An dem Beispiel der beiden Extrembeurteiler wollen wir diesen Einfluß verdeutlichen:

Zensor A hatte für jede der 6 Einzelaufgaben der Klassenarbeit jeweils zwei Punkte in Ansatz gebracht. War aber in der Aufgabe ein wenn auch noch so kleiner Fehler vorhanden, konnte nur noch einer von den beiden Punkten gegeben werden, das heißt, daß 50 Prozent der Aufgabe als falsch angesehen wurden.

Beurteiler A entdeckte 5 Fehler in den Aufgaben, die die Ergebnisse teilweise unrichtig werden ließen. Er schrieb unter die Arbeit, daß 7 von den erreichbaren 12 Punkten erzielt wurden. Und da $7/12$ etwa 55 Prozent sind, scheint seine Zensur „ausreichend“, die er gab, rein rechnerisch gerechtfertigt zu sein.

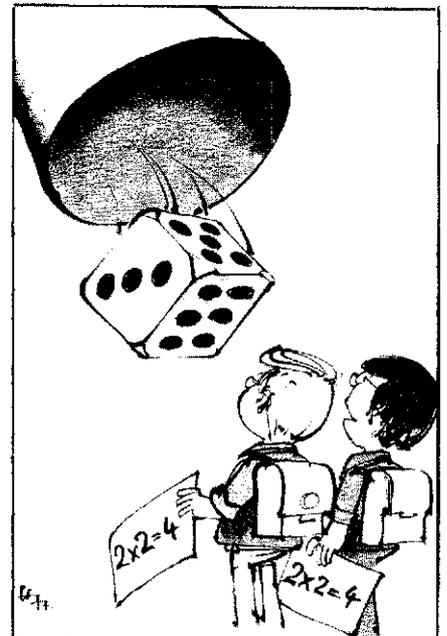
Der andere Extrembeurteiler (B) brachte für vier der 6 Aufgaben von vornherein je 5 Punkte in Ansatz. Für zwei weitere Aufgaben, die ihm etwas schwieriger zu sein schienen, gab er je 6 Punkte, so daß es insgesamt eine erreichbare Höchstpunktzahl von 32 Punkten gab. Auch er entdeckte die fünf kleinen Fehler des Drittkläblers, hielt sie aber für nur so schwerwiegend, daß lediglich jeweils ein oder zwei Punkte abgezogen wurden. Sein Endergebnis, das er unter die Arbeit schrieb, hieß: Es wurden 25 von 32 erreichbaren Punkten erzielt. Er kam daher auf die Zensur „gut“; denn nach seiner Aufrechnung wurden 78 Prozent der Aufgaben gelöst.

Wir sehen: Schon eine kleine, bei oberflächlichem Hinsehen als unbedeutend erscheinende Verfahrensweise der Punktevorgabe kann die Ursache dafür sein, daß ein Lehrer „gut“, der andere zur gleichen Leistung „ausreichend“ sagt.

entgegengesetzter Hinsicht aufschlußreich:

Es zeigte „eindeutig, daß bei vermeintlich positiver Vorinformation wesentlich freundlicher zensiert wurde als bei negativer“. So bewerteten elf Prozent der positiv beeinflussten Lehrer die Rechenarbeit des Viertkläblers mit „sehr gut“, aber keiner der negativ Vorinformierten. Nur fünf Prozent der positiv Beeinflussten sagten „ausreichend“ und niemand „mangelhaft“.

Bei den negativ Voreingenommenen dagegen urteilten 15 Prozent „ausreichend“ und zwei Prozent gar „mangelhaft“. Auch bei der Arbeit aus dem 5. Schuljahr ließ sich ein ähnliches Bild ablesen. „Zudem“, so Schröter, „macht die große Zensurenspanne für die gleiche Arbeit in einem als objektiv bewertbar bezeichneten Fach wie Mathematik doch betroffen. Sie beträgt genau



„Bin gespannt, wieviel Punkte ich wohl bekomme.“

vier verschiedene Zensuren bei jeder der eben genannten Beurteilergruppen und bei jeder der beiden Klassenarbeiten.“

Bei einem anderen Versuch wurden 37 sechste Klassen eines Berliner Bezirks mit Hilfe eines geeichten Rechenleistungstests überprüft. Innerhalb der Klassen stimmten die Bewertungsmaßstäbe dabei nahezu überall. Vergleiche zwischen Klassen hingegen ergaben: „Die Zensuren spiegelten nicht den Leistungsstand einer Klasse im Gegensatz zu einer anderen wider, sondern nur den Leistungsstand eines Schülers im Verhältnis zum Leistungsstand eines anderen Schülers seiner eigenen Klasse.“

Der Leistungsunterschied zwischen besten und schlechtesten Klassen hingegen war „gravierend“: Wer in der einen Klasse eine glatte 2 erhalten hatte, bekäme für die gleiche Leistung in seiner Parallelklasse ein „Mangelhaft“.