# Schularbeiten aus Mathematik ein Dilemma zwischen Ist und Soll

Günter HANISCH, Wien

## 1. Einleitung

Im folgenden möchte ich drei Begriffe aus der klassischen (psychologischen) Testheorie, nämlich die Objektivität, die Validität und die Reliabilität erläutern und Überlegungen anstellen, ob diese Begriffe auf Mathematikschularbeiten anwendbar sind und ob es sinnvoll ist, sie auf Mathematikschularbeiten anzuwenden.

Vorerst jedoch möchte ich Sie, geschätzte Leserin, respektive Leser ersuchen zwei Experimente durchzuführen und zwar:

### 1. Aufgabe

Im folgenden finden Sie eine Mathematikschularbeit für eine erste Klasse AHS, wie Sie etwa Ende März, Anfang April gegeben werden kann. Da die Lehrpläne für AHS und Hauptschulen wortident sind, könnte man sie aus diesem Grund auch in einer Hauptschule in einer ersten Leistungsgruppe stellen. Die Mathematikschularbeit wurde schon von einem Schüler durchgerechnet und ich möchte Sie bitten, diese nun so zu korrigieren und zu benoten, wie Sie es gewohnt sind. Sollten Sie noch nie Mathematikschularbeiten korrigiert haben, dann erinnern Sie sich bitte an Ihre Schulzeit und versuchen Sie so vorzugehen, wie Sie es selbst erfahren haben. Sie können beruhigt sein, auch Lehrer machen es nicht viel anders. BLÜML und PORSCH schreiben (S.35f., 1982), indem sie IVO (1982) zitieren: „Unser Wissen bezüglich des Korrekturhandelns stammt zum Teil aus der Erinnerung an die eigene Schulzeit, zum Teil aus zufälligen Gesprächen mit Kollegen, eventuell aus einer Art Meisterlehre durch den einführenden Lehrer im Probejahr. In unserer eigentlichen Ausbildung an der Universität haben wir – wie über so viele andere wichtige Dinge unseres Berufsalltags – nichts erfahren oder gelernt." Damit letzteres anders wird, gibt es heuer am Mathematischen Institut der Universität Wien ein Seminar über „Beurteilen im Mathematikunterricht".

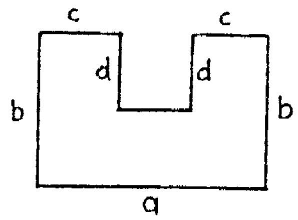
Nach dieser langen Vorrede nun die Schularbeit:

1A 5.Schularbeit 20.März 1986

1a) Berechne auf drei verschiedene Arten: $(9m^2 92cm^2 - 4m^2 91dm^2 87cm^2) : 1{,}01 =$

b) Gibt es noch mehr verschiedene Arten, nach denen man dieses Beispiel rechnen könnte?

2a) Gib eine Formel für den Umfang und eine für den Flächeninhalt nebenstehender Figur an!




b) Überprüfe beide Formeln für a = 3 cm, b = 2 cm und c = d = 1 cm!

Unser Klassenzimmer ist ungefähr 8 m lang, 6 m breit und 4 m hoch.

3a) Berechne den Rauminhalt des Klassenzimmers!

b) Wie schwer ist die im Klassenzimmer enthaltene Luft, wenn $1 m^3$ Luft 1,293 kg wiegt?

4) Wähle einen passenden Maßstab so, daß Du einen Plan des Klassenzimmers im Heft zeichnen kannst. Dabei soll die längste Seite der Zeichnung nicht größer als 10 cm, aber auch nicht kleiner als 5 cm sein. Gib den Maßstab an und zeichne den Plan!

Viel Erfolg!

1,01

1a) 90092 − 49187 = 40804 : 1,01 4 0500 $cm^2$

− 49200      00404

40804

900,92 − 491,87 = 40804 : 1,01 = 405 $dm^2$

9,0092 − 4,9187 = 40804 : 1,01 = 4,05 $m^2$

b) ja

2a) (c · d) 2 + (b : d) · a

(a 2) + (b · 2) (d 2)

b)

$2 \cdot 1 = 1 \cdot 3 = 3\,cm^2$

$2\,cm^2$

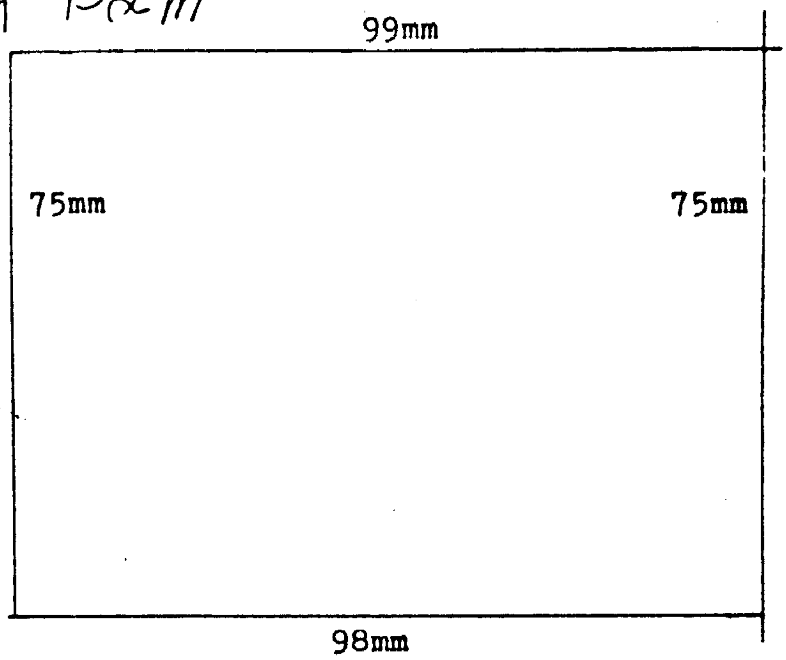
$5\,cm^2 = F$

$6 \cdot 4 + 2 = 12\,cm = U$

3 a) $8 \cdot 6 = 48 \cdot 4 = 192\,m^3$

b 1,239 · 192
11251
2478
238888

4) $8m = 10cm$
$4m = 5cm$




**2. Aufgabe**

Nachdem Sie die Schularbeit benotet haben, möchte ich Ihnen die zweite Aufgabe stellen. Die Idee dazu stammt von RHEINBERG (S.25f., 1980) und wurde von mir für Mathematikschularbeiten abgeändert, wobei ich auch die unterschiedliche Bewertungspraxis in Österreich im Gegensatz zur Bundesrepublik Deutschland berücksichtigt habe.

Eine durchschnittliche Schulklasse hatte im vergangenen Schuljahr 6 Mathematikschularbeiten, denen i.a. der jeweils letzte Unterrichtsstoff zugrundegelegt wurde. Die Maximalzahl der zu erreichenden Punkte betrug 40, der Klassendurchschnitt 25 Punkte. Neun Schüler erreichten bei den letzten drei Schularbeiten die unten angeführten Punkte.

Ihre Aufgabe besteht darin, *bei jedem der neun Schüler das letzte Ergebnis zu beurteilen.* Wenn Sie das Ergebnis eines Schülers für eine *gute Leistung* halten, so können Sie *einen bis fünf Pluspunkte* (++ ...) geben. Halten Sie dieses Ergebnis für eine *schlechte Leistung*, so können Sie *einen bis fünf Minuspunkte* (- ...) geben. Bitte geben Sie pro Ergebnis entweder *nur* Plus- oder *nur* Minuspunkte, also nicht beides gleichzeitig! Wenn Sie in eine Zeile weder Plus- noch Minuspunkte schreiben, so bedeutet das, daß Sie das Ergebnis weder für eine gute noch für eine schlechte Leistung halten.

| | erreichte 1.Schul-arbeit | Punkte 2.Schul-arbeit | bei der 3.Schul-arbeit | Beurteilung des letzten Schularbeitsergebnisses (bitte Plus- bzw. Minuszeichen in die Kästchen schreiben) | | | | |
|---|---|---|---|---|---|---|---|---|
| 1: | 35 | 33 | 31 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2: | 12 | 12 | 12 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3: | 20 | 22 | 25 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4: | 29 | 30 | 31 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5: | 25 | 25 | 25 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6: | 08 | 10 | 12 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7: | 29 | 27 | 25 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8: | 31 | 31 | 31 | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9: | 32 | 22 | 12 | ☐ | ☐ | ☐ | ☐ | ☐ |

Es kann sein, daß Sie bei einigen Schülern unsicher über die „richtige" Beurteilungsweise sind. Entscheiden Sie dann bitte so, wie Sie persönlich es für angemessen halten.

Wie man obige Aufgabe auswertet, darauf möchte ich später eingehen.

## 2. Objektivität

Nun, wie haben Sie die vorige Schularbeit benotet?
Vergleichen Sie Ihr Ergebnis mit den Noten, die beim Vortrag von den anwesenden Mathematiklehrer(inne)n gegeben wurden. (Die Beurteilungen, die von Nicht-Mathematiklehrern z.B. Studenten stammten, sowie die von drei Kolleg(inn)en, die mich darum ersuchten, wurden nicht berücksichtigt. Allerdings habe ich die Schularbeit noch anderen Kolleg(inn)en vorgelegt: diese Ergebnisse wurden mit aufgenommen.)

| Note | Anzahl | % |
|---|---|---|
| Sehr gut | 0 | 0 |
| Gut | 0 | 0 |
| Befriedigend | 2 | 10 |
| Genügend | 14 | 70 |
| Nicht genügend | 4 | 20 |
| $\sum$ | 20 | 100 |

Dieselbe Schularbeit, aber von einem anderen Schüler gerechnet, ließ ich ebenfalls korrigieren. Dabei gab es folgende Resultate:

| Note | Anzahl | % |
|---|---|---|
| Sehr gut | 0 | 0 |
| Gut | 0 | 0 |
| Befriedigend | 2 | 9 |
| Genügend | 11 | 50 |
| Nicht genügend | 9 | 41 |
| $\sum$ | 22 | 100 |

Sind Sie darüber erstaunt, daß dieselbe Schularbeit von verschiedenen Kolleg(inn)en verschieden beurteilt wurde?

In der Literatur wird des öfteren darüber berichtet:

So ließen STARCH und ELLIOT (S.69f., 1971) die Abschlußarbeit eines Schülers in Wisconsin von 128 Mathematiklehrern beurteilen. Es waren 10 Aufgaben, bei denen vor allem Beweise zu finden waren wie etwa „Beweise, daß jeder Punkt auf der Winkelhalbierenden gleich weit von den Winkelschenkeln entfernt ist!", aber auch Berechnungen am rechtwinkeligen Dreieck. Die maximale Punktezahl war 100, tatsächlich wurden hingegen 28 bis

92 Punkte für diese Arbeit vergeben. Diese Diskrepanz rührte vor allem daher, wie die schlechte äußere Form der Darstellung in die Beurteilung mit einbezogen wurde.

Berühmt wurde auch die Untersuchung von WEISS (S.62ff., 1976). Er ließ 153 Lehrer eine Mathematikarbeit der 4. und 119 Lehrer eine der 5.Schulstufe korrigieren, wobei er der Hälfte der Lehrer für die 1.Arbeit die Information gab:„Sie stammt von Fritz H., einem mathematisch begabten Buben mit einer Neigung zu originellen Lösungen." und für die 2.Arbeit:„Diese Arbeit stammt von Hans L., einem durchschnittlich begabten Buben. Die Original-Arbeit fiel durch unsaubere Form und schlampige Schrift auf." Der zweiten Hälfte der Lehrer wurden dieselben Arbeiten vorgelegt, nur wurden die beiden Informationen vertauscht. Eigentlich sollte das für die Korrektur der Arbeiten nichts ausmachen, wie aber waren die Ergebnisse?:

| | 4.Schulstufe | | 5.Schulstufe | |
|---|---|---|---|---|
| Sehr gut | 11% | 0% | 0% | 0% |
| Gut | 44% | 37% | 16% | 5% |
| Befriedigend | 40% | 46% | 37% | 36% |
| Genügend | 5% | 15% | 37% | 49% |
| Nicht genügend | 0% | 2% | 10% | 10% |

Wie sind solche Ergebnisse zu erklären?

Zur Erklärung bieten sich fünf Möglichkeiten an und zwar:

1. Unklare Angaben
2. Mangelnder Konsens über Gewicht der Beispiele und einzelner Teile von ihnen
3. Mangelnder Konsens darüber, wie auftretende Fehler zu bewerten sind
4. Zuordnung zu den Beurteilungsstufen nicht normiert
5. Verschiedene Voreinstellungen zum betreffenden Schüler auf Grund divergierender Informationen

Im folgenden möchte ich obige Punkte der Reihe nach besprechen, obwohl mir klar ist, daß sie nicht unabhängig voneinander sind.

## 1. Unklare Angaben

Betrachten wir die Angaben der Schularbeit, die Sie vorhin korrigiert haben, näher, so erkennen wir, daß sie schlecht formuliert sind, da vieles erst aus dem Kontext, wie es unterrichtet wurde, verständlich wird, wie z.B. Beispiel 1a). Was sind drei verschiedene Arten? Beispiel 1b) wieder sollte etwa so formuliert werden:„Erläutere noch mindestens eine von den bisherigen verschiedene Methode, nach der man dieses Beispiel rechnen könnte!" So sind i.a. Auftragsaufgaben besser als Fragen, daher ist 3a) besser als 3b) formuliert. Auch 2b) läßt sich nur auf Grund der Kenntnis des vorangegangenen Unterrichts korrigieren. Wann gilt eine Formel als überprüft?

Bei der gemeinsamen Angabe von Beispiel 3) und 4) steht das Wort „ungefähr". Soll sich der Schüler daher bei Beispiel 3) überlegen, wie genau er das Ergebnis hinschreiben soll? Und was ist, wenn er den Plan ungenau gezeichnet hat? Oder sollten etwa Streifen eingezeichnet werden, die die Ungenauigkeit der Angabe charakterisieren?

Sie sehen, ohne Kenntnis des vorangegangenen Unterrichts ist diese Schularbeit allein schon deswegen nicht eindeutig zu korrigieren, weil nicht exakt bekannt ist, was eigentlich verlangt wurde.

## 2. Mangelnder Konsens über das Gewicht der Aufgaben

Normalerweise steht bei den Angaben von Mathematikschularbeiten, wieviel Punkte bzw. bei Unterteilung der Beispiele in a) b) usw. wieviel die einzelnen Teile Punkte einbringen, soferne sie richtig bearbeitet wurden. Dies fehlt hier, man kann m.E. daher von der Annahme ausgehen, daß allen Beispielen dieselbe Zahl von Punkten zugeordnet wird. Dafür spräche auch, daß es nur vier Beispiele sind und eine Faustregel besagt, daß kein Beispiel mehr als $\frac{1}{4}$ der Maximalpunkteanzahl der Schularbeit haben soll (GIRG u. MAIER, S.387, 1981).

Nehmen wir noch weiters an, daß alle Teile eines Beispiels gleich zählen (zumindest bei Beispiel 1) ist das eine gewagte Annahme), so hängt das Ergebnis einer Schularbeit von der Maximalpunkteanzahl ab. So berichtet SCHRÖTER (S.80, 1977) von einer Mathematikarbeit eines Schülers des 3.Schuljahres der Grundschule, die er von 20 Studenten der Pädagogischen Hochschule Kiel, die bereits in den letzten Semestern waren, korrigieren ließ, daß, je wenige Punkte ein Zensor für die Schularbeit vorgesehen hatte, die Bewertung um so schlechter wurde. Was im nachhinein auch klar ist. Wollen Sie einen kleinen Fehler ahnden, dann wirkt das Abziehen eines Punktes bei einer kleinen Gesamtpunkteanzahl natürlich schwerwiegender als bei vielen Gesamtpunkten. Manche Lehrer behelfen sich dann, daß sie halbe Punkte oder gar Viertelpunkte abziehen. Was für umfangreiche Beispiele müssen das sein, daß ein Fehler nur $\frac{1}{24}$ eines Beispiels wiegt, soferne Sie 6 Punkte pro Beispiel vorsehen! Aber mehr darüber später.

## 3. Bewertung der Fehler

Beim Bewerten selbst gibt es zwei Methoden:

a) von oben herunter oder

b) von unten hinauf.

Darunter verstehe ich, daß man
bei a) von der Maximalzahl der Punkte ausgeht und für jeden Fehler 1 bis 2 Punkte abzieht.
Bei b) hingegen wertet man extra jeden Teil, etwa dieser Ansatz ist richtig: 2 Punkte (dabei gehe ich von einer Maximalzahl von 6 Punkten pro Beispiel aus), Berechnung ist richtig: 2 Punkte, Antwort ist vorhanden und richtig: 1 Punkt, Probe wurde richtig durchgeführt: 1 Punkt.

Beide Methoden beinhalten verschiedene Geisteshaltungen. Bei der Oben-herunter-Methode suchen Sie ständig nach Fehlern, damit Sie einen Punkteabzug vornehmen können, bei der Unten-hinauf-Methode freuen Sie sich über jeden richtigen Teil des Beispiels und erhöhen dabei jeweils die Punktezahl. Ich selbst ziehe daher die letztere der ersteren vor, bewerte aber grundsätzlich ein Beispiel nach beiden Methoden und nehme in dubio pro reo das für den Schüler bessere Ergebnis, denn die beiden Methoden können durchaus verschiedene Ergebnisse liefern. Betrachten wir etwa den Fall, daß ein Schüler $\frac{2}{3}$ eines Beispiels richtig gelöst hat und im letzten Drittel sechs schwere Fehler gemacht hat. Die Oben-herunter-Methode liefert dann 0 Punkte, die Unten-hinauf-Methode jedoch 4 Punkte.

Bei der Oben-herunter-Methode geht es also darum, zu bewerten, welche Abzüge man für welchen Fehler vornimmt, bei der Unten-hinauf-Methode jedoch darum, welche Teile eines Beispiels wieviel zählen.

Nun, wie bewerte ich Fehler im allgemeinen:

- Einen Punkt Abzug gibt es meist für Rechen- oder Flüchtigkeitsfehler; darunter verstehe ich Fehler wie Übertragungsfehler, Flüchtigkeitsfehler, Ungenauigkeit, Rechenfehler (beim Addieren oder Multiplizieren), Vorzeichenfehler etc.
- Zwei Punkte Abzug nehme ich i.a. bei Denkfehlern vor. Dazu zähle ich Wurzelziehen aus dem Pythagoras, durch eine Summe kürzen etc.
- Drei oder mehr Punkte bei noch schwereren Denkfehlern abzuziehen, halte ich für unnötig, da diese ja i.a. die Lösung unmöglich machen, so daß beim Bewerten die Unten-hinauf-Methode die zweckmäßigere ist.

Schwerer tue ich mir allerdings mit formalen Fehlern, wie etwa

$$10 + 2 = 12 + 2 = 14 + 2 = 16,$$

wie man es – leider scherze ich nicht – den Kindern in der Volksschule beibringt. Wie schwer soll man so einen Fehler oder ähnliche werten?

Dieses Problem tritt in noch schärferer Form bei der Korrektur von Deutschschularbeiten auf: Die Note einer Deutschschularbeit setzt sich – grob gesprochen – aus dem Inhalt, der Ausdrucksweise und der Sprachrichtigkeit zusammen. Ein Dilemma tritt dann auf, wenn zwischen diesen drei Beurteilungskriterien große Unterschiede bestehen. Wie wird dann gewichtet?

Ähnlich ist auch das Problem bei Mathematikschularbeiten gelagert. Hier geht es um logische, um rechnerische und um formale Richtigkeit (wobei ich das Problem der Genauigkeit von Zeichnungen ausklammere). I.a. betrachten wir die logische Richtigkeit als conditio sine qua non, d.h. Rechnungen, mögen sie auch ohne Fehler sein, tragen zu keiner Punkteerhöhung bei, wenn sie mit der gestellten Aufgabe nichts zu tun haben. Rechnerische Fehler haben wir im i.a. im Griff, ich habe oben angeführt, wie ich den Punkteabzug handhabe. Bei formalen Fehlern fehlt hingegen der Konsens. Da sollten wir uns einmal zusammensetzen und darüber diskutieren.

## 4. Notengebung

Wir kommen nun zum Zuordnen der Punktezahl zu einer Note. Nach meinen Erfahrungen haben sich bei uns in Österreich an den Höheren Schulen folgende Usancen eingebürgert:

| Note bei | 40 Punkten | 24 Punkten | 16 Punkten |
|---|---|---|---|
| Sehr gut | 37 - 40 | 23, 24 | 15, 16 |
| Gut | 32 - 36 | 20 - 22 | 13, 14 |
| Befriedigend | 26 - 31 | 16 - 19 | 11, 12 |
| Genügend | 20 - 25 | 12 - 15 | 8 - 10 |
| Nicht genügend | 0 - 19 | 0 - 11 | 0 - 7 |

Allerdings handelt es sich bei obigen Zuordnungen nur um eine Usance. In der Bundesrepublik sind andere (mildere) Benotungssysteme bei Mathematikschularbeiten üblich.

## 5. Pädagogische Notengebung

Aus der vorhin erwähnten Untersuchung von WEISS kann man erkennen, daß auch Voreinstellungen die Notengebung beeinflussen. Ich möchte das aber keineswegs nur negativ sehen, denn eine Note muß auch pädagogisch gerechtfertigt sein. Und ist es gerecht, jeden Fehler gleich zu werten?

Ein Beispiel: Schüler A, der zumeist „Sehr gut" schreibt, macht einen Vorzeichenfehler. Dabei handelt es sich aller Wahrscheinlichkeit nach um ein Versehen. Schüler B, der pro Schularbeit etwa zwei bis drei Vorzeichenfehler macht, macht ebenfalls wieder einen. Hier handelt es sich vermutlich um ein Nichtkönnen. Soll man das gleich werten?

Oder ein anderes Beispiel: Ein Schüler, der sich allem Anschein nach schwer tut, aber wirklich bemüht, bislang „Nicht genügend" geschrieben hat, liefert eine Schularbeit mit – „objektiv" korrigiert – 11 Punkten (von maximal 24 Punkten). Ist es hier pädagogisch nicht nur nicht gerechtfertigt, sondern sogar erforderlich, doch noch einen Punkt herzuzaubern?

Würde so ein Vorgehen aber nicht die Objektivität verletzen? Was versteht man eigentlich darunter? Nach der klassischen Testtheorie ist ein Test dann objektiv, wenn das Testergebnis unabhängig von der Person des Testers ist, daß also verschiedene Tester zum selben Ergebnis kommen. Nun ist es aber bei psychologischen Tests so, daß mit einem Test i.a.

auch eine genaue Anleitung zum Auswerten dazu gegeben wird, so daß zumindest bei Intelligenztests die Objektivität daher gegeben ist. Auf die Schule übertragen würde das heißen, daß zu einer Mathematikschularbeit sowohl eine Musterschularbeit (wie bei der Reifeprüfung) und ein Fehlerkodex als auch eine Zuordnungsskala zu den einzelnen Noten mitgeliefert wird. Das würde ich mit *„schwacher Objektivität"* bezeichnen. In der Praxis ist diese aber bei der Beurteilung von Schülern irrelevant. Gefordert wird vielmehr etwas, das ich *„starke Objektivität"* nennen möchte, nämlich, daß das Prüfverfahren nicht vorgegeben ist, verschiedene Prüfer aber trotzdem zum gleichen Ergebnis kommen. Herrscht also Konsens darüber, daß eine Note auch pädagogisch gerechtfertigt sein sollte, denn Ungleiche gleich zu behandeln, halte ich für ungerecht, würde die starke Objektivität darunter nicht leiden.

Welche Wege gäbe es nun, die starke Objektivität zu verbessern? Zwei Möglichkeiten bieten sich dazu an und zwar:
a) normorientiert oder
b) kriteriumsorientiert zu beurteilen.

### a) Normorientierte Beurteilung

INGENKAMP (S.157f., 1971) gab 37 Klassen der 6.Schulstufe eines Berliner Bezirks einen Mathematiktest vor, der auf Grund des Lehrplans zusammengestellt und an einer repräsentativen Zahl von Schülern geeicht worden war. Dreierlei zeigte sich:

1. Schüler einer Klasse, die im Rechentest besser waren als ihre Mitschüler, erhielten auch bessere Mathematiknoten, d.h. das Bezugssystem war klassenintern in Ordnung.
2. Die Note aber, die ein Schüler in der einen Klasse hatte, entsprach i.a. nicht der Note, die ein Schüler mit derselben Testleistung in einer anderen Klasse hatte.
3. Gerade in den besten Klassen wurden überdurchschnittlich viele schlechte, in den schlechten Klassen fast keine schlechten Noten vergeben.

Das gleiche Ergebnis wie INGENKAMP erzielte WECK, der Halbjahresnoten in Mathematik von 36 zweiten Klassen der Oberschule mit einem Schulleistungstest kontrollierte (S.51 ff., 1976). So hatte es Klassen gegeben, in denen, legte man die aus dieser Kontrolluntersuchung festgelegten in anderen Schulen üblichen Anforderungen zugrunde, mehr als die Hälfte „Sehr gut" hätte gegeben werden müssen, was aber nicht geschehen war. Umgekehrt gab es aber auch Klassen, die die gestellten Anforderungen wesentlich schlechter erfüllten; bei einer anderen Untersuchung nur zu 12%. „Die Bedingungsanalyse ergab, daß die Hauptursache in der ungenügenden Einstellung der Lehrerin zu ihrer Arbeit zu suchen war. Verstärkte Kontrolle führte zu besserem Unterricht, und schon nach vier Monaten erreichte die Klasse bei einer zweiten Kontrollarbeit vergleichbaren Schwierigkeitgrades mehr als 60%" der Anforderungen (S.66).

Auch DICKER (1979), der Lehrer ersuchte, 15 Arbeiten verschiedener Schüler in eine Rangordnung zu bringen, konnte feststellen, daß dabei keine signifikanten Unterschiede zwischen den Beurteilungen auftraten.

Zusammenfassend kann man also feststellen: Mathematiklehrer geben ihre Noten i.a. im Rahmen eines klasseninternen Bezugssystems, was ihnen auch recht gut gelingt. Vergleichbar ist diese Note aber mit der einer anderen Klasse nicht. Vielmehr ist es so, daß Schüler, die in einer besseren Klasse sitzen, häufiger schlechte Noten bekommen als umgekehrt.

Hat daher der Lehrer eine bestimmte Klasse zu beurteilen, so braucht nur Konsens über die Notenverteilung herrschen, um stark objektiv innerhalb der Klasse zu benoten. Oft wird etwa die Normalverteilung gefordert (in Anlehnung an das Bestimmen des Intelligenzquotienten, wo diese auch a priori festgelegt wird), wobei es Schüler gibt, die dem Vernehmen nach darüber froh wären, nämlich solche, in deren Klassen fast 50% der Mathematikschularbeiten regelmäßig mit „Nicht genügend" beurteilt werden.

Allerdings wird die Notenverteilung ja von einem erfahrenen Lehrer nicht erst beim Korrigieren, sondern schon beim Zusammenstellen der Schularbeit festgelegt. Sie protestieren? Sie meinen, wesentlich ist das, was *der* Schüler kann, nicht das, was die anderen können? Nun, das möchte ich im folgenden Abschnitt behandeln.

### b) Kriteriumsorientierte Beurteilung

Bei dieser geht es darum zu messen, ob und wie gut ein Schüler ein bestimmtes Kriterium (=Lehrziel) erreicht hat. Dies ist auch die Vorgangsweise, die – ausgenommen die 50%-Klausel bei Schularbeiten, die normorientiert ist – von der Obrigkeit gefordert wird, denn es heißt in §14 der Verordnung über die Leistungsbeurteilung:„...sind Leistungen zu beurteilen, mit denen der Schüler die nach Maßgabe des Lehrplanes gestellten Anforderungen in der Erfassung und in der Anwendung des Lehrstoffes sowie in der Durchführung der Aufgaben ... erfüllt ...". Das kann ich allerdings nur als Absichtserklärung auffassen. Was heißt schon „nach Maßgabe des Lehrplans"? Wenn Sie etwa die drei Oberstufenlehrbücher für Mathematik an AHS (alle sind approbiert!) durchschauen, sehen Sie, wie unterschiedlich die Auffassungen darüber sein können, was Mathematik ist und was ein Schüler können sollte.

### Resümee

Beide Wege gehen ins Leere. Eine Objektivität im schwachen Sinn ist realitätsfern, eine im starken Sinn bringt Dinge herein (siehe a)), die wir nicht wollen, oder läßt sich nicht verwirklichen (siehe b)). Ob da verbale Beurteilung und dadurch eine *Messung des individuellen Lernfortschritts* nicht pädagogisch sinnvoller wäre? Um dies zu erproben, führt das Ludwig Boltzmann Institut für Schulentwicklung und international vergleichende Schulforschung gemeinsam mit anderen Instituten im Schuljahr 1986/87 den großangelegten Versuch „Motivieren ohne Noten" durch. Warten wir die Ergebnisse ab.

Damit Sie selbst feststellen können, welche Art der Leistungsbeurteilung Sie vornehmen, habe ich Sie einige Seiten vorher ersucht, die Leistungen von 9 Schülern bei der letzten Mathematikschularbeit zu beurteilen.

Wie Sie leicht erkennen können, handelt es sich bei den Schülern 1, 4 und 8 um in Mathematik überdurchschnittliche, bei denen mit den Nummern 3, 5 und 7 um durchschnittliche und bei denen mit den Nummern 2, 6 und 9 um unterdurchschnittliche Schüler. Dabei gibt es aber in jeder der obigen drei Gruppen einen Schüler, dessen Leistungen sich verbessern (3, 4 und 6), einen, dessen Leistungen gleichbleiben (2, 5 und 8), und einen, dessen Leistungen abfallen (1, 7 und 9).

Sehen Sie die Leistung vor allem norm- oder kriterienorientiert, d.h., Sie nehmen einen Querschnittsvergleich vor, so haben Sie den Schülern der ersten Gruppe viele + und denen der letzten Gruppe viele – gegeben. Haben Sie aber vorwiegend den individuellen Lernfortschritt beachtet, also einen Längsschnittsvergleich vorgenommmen, so erhalten die in ihren Leistungen besser werdenden Schüler +, die in ihren Leistungen abfallenden hingegen die –. Wohin tendieren Sie mehr?

Falls Ihnen die verbale Beurteilung zu radikal sein sollte:„Was halten Sie von OLECHOWSKIs Vorschlag einer dreistufigen Notenskala verbunden mit einer Kommentierung der Note?" (1984).

## 2. Validität

Auch dieser Ausdruck stammt aus der Testpsychologie. Dort unterscheidet man mehrere Arten von Validität. Für unser Problem ist der Begriff der Konstruktvalidität zweckmäßig. „Damit ist das Ausmaß gemeint, in welchem der Test diejenige Eigenschaft mißt, die er messen soll" (FISCHER, S.76, 1972).

Auf Mathematikschularbeiten angewandt, heißt das: „Werden bei der Schularbeit jene Kenntnisse und Fertigkeiten abgeprüft, die für das Fach relevant sind?" Um darauf näher einzugehen, müßte man sich vorher überlegen, warum und weshalb Mathematik unterrichtet wird bzw. sollte. Dies würde jedoch den Rahmen dieser Arbeit sprengen (siehe auch HANISCH und SCHWENDENWEIN, 1986). Daher möchte ich hier nur einige Negativbeispiele anführen:

- Kommen zur Schularbeit dieselben Aufgaben (eventuell mit anderen Zahlen) wie zu einer Probeschularbeit, so wird m.E. insbesondere in höheren Klassen nicht das abgeprüft, was Mathematik ist.

- Behandelt keines der Beispiele der Schularbeiten der 7. und 8.Klasse AHS Stochastik, so geht der Unterricht und daher auch die Schularbeit m.E. an den Lehrzielen des Mathematikunterrichts vorbei.

- Werden hingegen andererseits Beispiele verlangt, die nicht ausreichend geübt worden sind, die Schularbeit daher mehr einem Intelligenztest gleicht als das sie mathematisches Wissen und Können überprüft, geht sie auch m.E. an den Zielen des Mathematikunterrichts vorbei.
- Eine Schularbeit wiederum, deren Aufgaben wegen ihres Umfangs vor allem Schnellschreiben und Fingerfertigkeit beim Eintippen in den Taschenrechner verlangen, prüft auch nicht das, was ich unter Mathematik verstehe.
- Insbesondere sollte eine Schularbeit auch Aufgaben zum Begründen, Argumentieren und dergleichen mehr enthalten.

Auch hier scheint es – um die Schwierigkeiten, die sich beim Erstellen von Mathematikschularbeiten ergeben, zu verdeutlichen – zweckmäßig zu sein, zwischen *schwacher und starker Validität* zu unterscheiden. Erstere liegt dann vor, wenn die Fähigkeit, die gemessen wird, durch den Test definiert wird, d.h. etwa:„Intelligenz ist das, was ein bestimmter Intelligenztest mißt."

Beim Mathematikunterricht liegen hingegen die Verhältnisse ganz anders. Vor allem die allgemeinen Lehrziele sind sehr vage definiert, obwohl die neuen Lehrpläne bereits eine Verbesserung bringen; insbesondere ist die Gewichtung der/dem Lehrer(in) überlassen, was pädagogisch höchst sinnvoll ist, da sie/er sich dadurch sowohl nach den Interessen der Schüler(innen) als auch nach den eigenen richten kann. Trotzdem sollen alle Schüler Mathematik lernen und die Schularbeit überprüfen, inwieweit dies geschehen ist.

Da sich aber die allgemeinen Lehrziele schwer operationalisieren lassen, leidet naturgemäß die Objektivität darunter, wenn man diese bei einer Schularbeit abprüft. Dies haben FISCHER, MALLE und BÜRGER (S.320, 1985) mit der Unschärferelation verglichen: Je valider, desto weniger objektiv.

Verzichtet man aber weitgehend diese Lehrziele in Schularbeitsaufgaben zu verpacken, überprüft man vorwiegend nur Rechentechniken, die sich ja leichter beurteilen lassen, aber in erster Linie – vor allem in der Oberstufe – Mittel zum Zweck sein sollten, deren Aufgabe also ist, das Vehikel zu sein, mit dem die allgemeinen Lehrziele transportiert werden sollen, so beschränkt sich auch der Unterricht zumeist nur auf die Vermittlung bloßer Rechentechniken. Der Lehrer sollte ja beim Unterrichten ein bestimmtes Lehrziel quasi „im Hinterhaupt" haben; so könnte er etwa, wenn er Stereometrie unterrichtet, seinen Unterricht so auslegen, daß er die Raumvorstellung (genaueres über diesen Begriff in HANISCH, 1985a) seiner Schüler(innen) optimal fördert. Es sind ja die Ergebnisse, die man erhält, wenn man Maturanten einige Zeit nach der Matura über ihre Kenntnisse, die sie noch von Mathematik haben, äußerst enttäuschend (siehe NÄGERL u.a., 1975, ROSNICK u. CLEMENT, 1980 oder HANISCH, 1985b). So war etwa eine der Antworten auf die Frage, was das „Differenzieren" sei:„Das ist das mit dem Strich" oder auf die Frage nach dem „Sinus":„Alpha, Beta, Gamma. – Nein, das ist etwas anderes, das sind ja Winkel." Die Hoffnung, die wir haben, ist, daß im Mathematikunterricht eben mehr als „nur" Trigonometrie, Analysis etc. unterrichtet wurde. Obwohl man m.E. auch noch Jahre nach der Matura wissen sollte, was „Differenzieren" und was der „Sinus" ist (Ich weiß z.B. auch noch, was der AcI ist und wo der Golfstrom fließt). Aber wird bei Mathematikschularbeiten wirklich überprüft, was „Differenzieren" etc. ist?

## 3. Reliabilität

Die Reliabilität gibt an, „wie präzise, konsistent oder stabil die Testleistung eines Schülers über verschiedene Testdurchgänge und Zeitspannen hinweg ist" (GAGE und BERLINER, S.723, 1979). Dies wird meist dadurch gemessen, daß man denselben Test einige Zeit später noch einmal vorgibt (Wiederholungsreliabilität; sollte dies nicht möglich sein, verwendet man andere Methoden). Schneidet die Versuchsperson dabei etwa gleich ab, heißt man den Test reliabel. Ich würde dem dazusetzen *„schwach reliabel"*, denn obiger Sachverhalt ist für Mathematikschularbeiten irrelevant – abgesehen von den an sich schon sehr wenigen Fällen einer Berufung, bei denen noch dazu die Unterlagen nicht ausreichend sind, so daß es zu einer kommissionellen Prüfung kommt.

Was wir Mathematiklehrer(innen) erhoffen, ist, daß ein(e) Schüler(in) Mathematik nicht nur zur Schularbeit, sondern auch unabhängig davon kann. Daraus folgt aber, daß Mathematikschularbeiten nicht zu sehr vorbereitet werden sollten, weder vom Lehrer noch von den Schüler(inne)n. Die Vorbereitung vom Lehrer erfolgt oft durch das abhalten von Probeschularbeiten. Einerseits wirken diese angstabbauend, andererseits werden die Schüler(innen) aber gerade dadurch auf *diese* Schularbeit zu *diesem* Zeitpunkt vorbereitet. Ein Ausweg würde m.E. folgender sein: Probeschularbeiten nicht einige Tage vor der Schularbeit abzuhalten, sondern regelmäßig, insbesondere zum Abschluß eines Kapitels. Wenn Ihre Schüler(innen) dabei die Aufgaben selbst zusammenstellen, glaube ich, dürfte auch bei einer Schularbeit nichts mehr passieren. Das heißt also: Für die Schularbeit selbst sollten sich die Schüler(innen) nicht extra vorbereiten.

Wie aber schaut dies in der Realität aus? Um dies festzustellen, befragte ich zu Beginn des Schuljahres 1985/86 über 170 Maturanten aus dem Osten Österreichs über ihre Erfahrungen mit Tests, Schularbeiten und der Reifeprüfung. Die Umfrage war allerdings deswegen nicht ganz repräsentativ, weil 29 der Befragten (d.h. 17%) Mathematik studierten. (Der Grund dafür lag darin, daß ich untersuchen wollte, ob Mathematikstudenten bei der Reifeprüfung die mündlichen Fächer aus anderen Gründen als die anderen Maturanten gewählt haben; siehe HANISCH 1986.)

In der folgenden Tabelle finden Sie, wieviele Tage bzw. Stunden man sich durchschnittlich (= arithmetisches Mittel) für eine Schularbeit bzw. einen Test vorbereitet hat:

| Schularbeiten und Tests | Tage | Vorbereitung Stunden (insg.) |
|---|---|---|
| Mathematik | 4,8 | 13,0 |
| Deutsch | 0,3 | 0,8 |
| Englisch | 2,3 | 6,2 |
| Latein | 4,6 | 12,0 |
| Französisch | 4,3 | 10,8 |
| Darst. Geometrie | 3,9 | 13,9 |
| Religion | 0,7 | 1,8 |
| Musikerziehung | 1,2 | 2,3 |
| Biologie u. Umweltkunde | 2,3 | 6,2 |
| Geschichte u. Sozialkunde | 2,3 | 6,6 |
| Philosoph. Einf.unterr. | 2,5 | 6,3 |
| Chemie | 2,9 | 7,3 |
| Geographie u. Wirtschaftsk. | 3,2 | 7,0 |
| Physik | 3,5 | 7,6 |

Allerdings kann mangelnde Reliabilität auch durch die Beurteiler verursacht werden, wie DICKER (S.123, 1979) nachwies. Er gab 24 Lehrern eine Mathematikschularbeit zu beurteilen und legte dieselbe Schularbeit drei Monate später ihnen noch einmal zum Benoten vor. Das Ergebnis war folgendes:

| Note | 3 Monate später | | | $\Sigma$ |
|---|---|---|---|---|
| | 3 | 4 | 5 | |
| 3 | 2 | 1 | 0 | 3 |
| 4 | 4 | 10 | 4 | 18 |
| 5 | 0 | 2 | 1 | 3 |
| $\Sigma$ | 6 | 13 | 5 | 24 |

Daher gaben nur 13 (Hauptdiagonale) von 24 Lehrern dieselbe Note! (DICKERs Ergebnis waren nur 8 Lehrer, aber er betrachtete dabei auch Zwischennoten.)

## 4. Zusammenfassung

Die Forderung, daß Mathematikschularbeiten sowohl stark objektiv, valide und reliabel sein sollen, verlangt m.E. einander teilweise ausschließende Eigenschaften. Denn werden Schularbeiten so gestellt, daß sie objektiv sind, messen sie zumeist keine höheren Lehrziele, gehen somit am Wesen der Mathematik vorbei und sind daher nicht valide. Legt man hingegen großen Wert auf starke Validität, so ist die Schularbeit schwer zu stellen und zu beurteilen. Insbesondere müßte näher überprüft werden, ob dann das „Überprüfen" höherer Lehrziele, wenn die Latte zu hoch gelegt wird, nicht im Auswendiglernen von Beispielen - eben neuerer Art - mündet, denken Sie nur an das Scheitern der New-Maths im Schulunterricht.

Die Reliabilität ließe sich m.E. dadurch verbessern, daß die Schüler(innen) nicht gezielt auf Mathematikschularbeiten, sondern auf Mathematik vorbereitet werden. Daher sollten Probeschularbeiten nicht unmittelbar vor der Schularbeit, sondern jeweils nach Ende eines Unterrichtsabschnitts durchgeführt werden.

## 5. Literatur

Hier führe ich nur die in dieser Arbeit zitierte Literatur an und nicht die von mir zusätzlich gelesene. Ein Literaturverzeichnis mit über 100 Titel finden Sie etwa in HANISCH (1987).


1. BLÜML, K. u. PORSCH, M.: Schularbeiten in Deutsch. Neue Formen schriftlichen Arbeitens - Vorbereitung und Gestaltung. Allgemeinbildende Pflichtschulen - Analysen, Berichte und Dokumentationen, Nr. 8. Wien, 1985.
2. DICKER, H.: Untersuchung zur Beurteilung von Mathematikaufgaben. Aus: INGENKAMP, K. (Hrsg.): Schüler- und Lehrerbeurteilung. S. 171-193, Weinheim, 1977.
3. DICKER, H.: Leistungsbeurteilung in Mathematik. Frankfurt am Main, 1979.
4. FISCHER, G.: Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen. Bern, 1972.
5. FISCHER, R., MALLE, G. u. BÜRGER, H.: Mensch und Mathematik. Eine Einführung in didaktisches Denken und Handeln. Lehrbücher und Monographien zur Didaktik der Mathematik. Band 1. Mannheim, 1985.
6. GAGE, N. L. u. BERLINER, D. C.: Pädagogische Psychologie. München, 1979².
7. GIRG, R. u. MAIER, H.: Die Leistungsfestellung im Fach Mathematik. In: Blätter für Lehrerfortbildung, S. 384-389, Oktober 1981.

8. HANISCH, G.: Mathematik und Raumvorstellung. Aus: DÖRFLER, W. u. FISCHER, R. (Hrsg.): Empirische Untersuchungen zum Lehren und Lernen von Mathematik. S. 83-90, Wien, 1985a.

9. HANISCH, G.: Was bleibt vom Mathematikunterricht hängen? Ergebnisse einer Befragung von Personen, deren Reifeprüfung länger zurückliegt. Aus: DÖRFLER, W. u. FISCHER, R. (Hrsg.): Empirische Untersuchungen zum Lehren und Lernen von Mathematik. S. 75-82, Wien, 1985b.

10. HANISCH, G.: Reifeprüfung - Illusion und Wirklichkeit (Ergebnisse einer empirischen Untersuchung). Vortrag am 20.Bundeskongreß der Gesellschaft für Didaktik der Mathematik in Bielefeld, März 1986 (im Druck).

11. HANISCH, G.: Mathematikschularbeiten - ein notwendiges und hinreichendes Übel? Wien, 1987 (in Vorbereitung).

12. HANISCH, G. u. SCHWENDENWEIN, W.: Zum Problem der Leistungsbeurteilung im Sekundarbereich. In: Erziehung und Unterricht, 1986 (im Druck).

13. IVO, H.: Lehrer korrigieren Aufsätze. Frankfurt am Main, 1982.

14. INGENKAMP, K.: Die Fragwürdigkeit der Zensurengebung. Weinheim, 1971.

15. NÄGERL, H., BECKER, H. u. ZERBST, J.: Über die Schwierigkeiten der Studienanfänger in Medizin im Umgang mit Mathematik. In: Didaktik der Mathematik, S. 150-158, Heft 2, 1975.

16. OLECHOWSKI, R.: Reform der Leistungsbeurteilung im Prozeß zur Humanisierung der Schule. In: BMUKS (Hrsg.): Schulische Leistung und ihre Bewertung. Wien, 1984.

17. RHEINBERG, F.: Leistungsbewertung und Lernmotivation. Göttingen, 1980.

18. ROSNICK, P. u. CLEMENT, J.: Learning without Understanding: The effect of Tutoring Strategies on Algebra Misconceptions. Journal of Mathematical Behavior, S. 3-27, Nr. 1, Vol. 3, 1980.

19. SCHRÖTER, G.: Zensurengebung. Allgemeine und fachspezifische Probleme. Kastellaun, 1977.

20. SCHRÖTER, G.: Zensuren? Zensuren. Allgemeine und fachspezifische Probleme. Grunderkenntnisse und neue Forschungsergebnisse für Lehrer, Eltern und interessierte Schüler. Kastellaun, 1981[3].

21. STARCH, D. u. ELLIOT, E. C.: Die Verläßlichkeit der Zensuren von Mathematikarbeiten. Aus: INGENKAMP, K.: Die Fragwürdigkeit der Zensurengebung. Weinheim, 1971.

22. WECK, H.: Leistungsermittlung und Leistungsbewertung im Unterricht. Berlin, 1976.

23. WEISS, R.: Leistungsbeurteilung durch Ziffernnoten. Linz, 1968.

2.Schularbeit 24.4.1986 9.B

1)Ein Dreieck ABC ( A(-4/-2);B(6/2);C(4/6) ) ist gegeben. Berechnen Sie den Umkreis und stellen Sie die Eulersche Gerade auf.

2)Über der Ebene $2x - y + 2z = 8$ wird eine dreiseitige Pyramide ABCD erreichtet. Der Punkt $A(3/y_1/0)$ und der Punkt $C(-1/y_2/7)$ liegen dementsprechend in dieser Ebene. Der Punkt B liegt auf der Geraden

$$x = \begin{pmatrix} 2 \\ 16 \\ -17 \end{pmatrix} + t \begin{pmatrix} 1 \\ -5 \\ 10 \end{pmatrix}$$ . Die Spitze D liegt ebenfalls auf der Geraden g

und befindet sich senkrecht über dem Punkt C.Berechnen Sie das Volumen der Pyramide.

3)Die Gerade $7x-y=-6$ ist eine Polare bezüglich dem Kreis:$(x+4)^2+(y-3)^2=25$ Berechnen Sie die Fläche mit der Formel $A= \frac{g.h}{2}$ des entsprechenden Poldreiecks.

4)Die Punkte P(7/17) und Q(2/16) liegen auf einem Kreis, dessen Mittelpunkt auf der Geraden $5x-2y=27$ liegt. Berechnen Sie die Winkel des Dreiecks PMQ.

Viel Erfolg!!!

2.Schularbeit 24.4.1986 9.B

1)Ein Dreieck ABC ( A(-4/-2);B(6/2);C(4/6) ) ist gegeben. Berechnen Sie den Umkreis und stellen Sie die Eulersche Gerade auf.

2)Über der Ebene $2x - y + 2z = 8$ wird eine dreiseitige Pyramide ABCD erreichtet. Der Punkt $A(3/y_1/0)$ und der Punkt $C(-1/y_2/7)$ liegen dementsprechend in dieser Ebene. Der Punkt B liegt auf der Geraden

$$x = \begin{pmatrix} 2 \\ 16 \\ -17 \end{pmatrix} + t \begin{pmatrix} 1 \\ -5 \\ 10 \end{pmatrix}$$ . Die Spitze D liegt ebenfalls auf der Geraden g

und befindet sich senkrecht über dem Punkt C.Berechnen Sie das Volumen der Pyramide.

3)Die Gerade $7x-y=-6$ ist eine Polare bezüglich dem Kreis:$(x+4)^2+(y-3)^2=25$ Berechnen Sie die Fläche mit der Formel $A= \frac{g.h}{2}$ des entsprechenden Poldreiecks.

4)Die Punkte P(7/17) und Q(2/16) liegen auf einem Kreis, dessen Mittelpunkt auf der Geraden $5x-2y=27$ liegt. Berechnen Sie die Winkel des Dreiecks PMQ.

Viel Erfolg!!!