

Zur Wirkung von Lehrerkommentaren bei der Bewertung schriftlicher Arbeiten

Günter Krampen

Universität Trier

Ausgehend von der Differenzierung verschiedener Kommentartypen wird die Wirkung von an einem sachlichen Gütemaßstab orientierten Kommentaren (Versuchsgruppe; $N = 48$) im Vergleich zu keinen Kommentaren (Kontrollgruppe; $N = 83$) analysiert. Es zeigte sich, daß bei durchschnittlich gleichen Testleistungen die Versuchsgruppe mit ihren Testergebnissen unzufriedener ist, sich mehr über die erhaltene Leistungsrückmeldung freut und die Note häufiger auf Anstrengung und Stoffschwierigkeit attribuiert als die Kontrollgruppe. Für die Leistungen in einem folgenden Wissenstest konnten keine signifikanten Unterschiede zwischen den Gruppen ermittelt werden. Die Diskussion bezieht sich auf die Auswirkungen verschiedener Typen von Lehrerkommentaren zu Noten in schriftlichen Arbeiten bei den Schülern.

Eine gängige Lehrmeinung in der Pädagogischen Psychologie, die in Einführungs- und Lehrbüchern häufig vertreten (vgl. etwa Campbell & Stanley, 1963; Zielinski, 1974; Gage & Berliner, 1977; Schmidt, 1978) und — vor allem an angehende Lehrer — weitergegeben wird, besagt, daß die motivationale Wirkung von Zensuren in schriftlichen Arbeiten oder Tests durch schriftliche Kommentare des Lehrers positiv beeinflusst werden kann. Hierin wird eine Möglichkeit gesehen, die Problematik von Noten zu entschärfen und ihre in Frage stehenden pädagogischen Funktionen zu erhöhen. Meist wird als Beleg dabei auf die klassische Untersuchung von Page (1958) zurückgegriffen, die durch ihre hohe externe Validität (Vermeiden reaktiver Versuchsbedingungen, sehr große Stichprobe) besticht (vgl. Campbell & Stanley, 1963).

Page (1958) ließ einen planmäßig durchgeführten Wissenstest zunächst von den Lehrern nach dem üblichen Vorgehen mit Noten bewerten. Im Anschluß ließ er die Tests der Schüler jeder Klasse in eine leistungsbezogene Rangreihe bringen. Mit einem Zufallswürfel wurde dann für die Schüler, deren Tests in der Leistungsangereihe die ersten drei Rangplätze einnahmen, bestimmt, welcher Schüler (a) keinen Kommentar (Kontrollgruppe), (b) einen spezifizierten Kommentar (vorher für die Notestufen festgelegte, kurze Anmerkungen wie „Ausgezeichnet! Weiter so“ bei der besten oder „Runter von dieser Note!“ bei der schlechtesten Note) und (c) einen freien Kommentar (vorher nicht festgelegte, in das Ermessen des Lehrers gestellte Anmerkungen) zur Note erhielten. Die Zufallsordnung der jeweils drei in der Rangordnung hintereinanderstehenden Schüler zu den drei Versuchsbedin-

gungen wiederholte Page für alle Schüler. Als Kriterienvariable diente die Leistung im nächsten Test. Es zeigte sich, daß die Schüler, die einen Kommentar zur Note erhalten hatten, im folgenden Test signifikant bessere Leistungen erbrachten, wobei die freien (also vorher nicht festgelegten) Kommentare der Lehrer besonders wirksam waren.

In Folgestudien konnte dieser Befund nicht durchgehend repliziert werden. Stewart & White (1976) stellten etwa für notenbezogene Kommentare bei mathematischen und sprachlichen Leistungen (415 Schüler der 5. und 7. Klasse; sechswöchiges Intervall zwischen den Arbeiten) keine positiven Effekte fest. Auch in ihrer Übersicht zu 13 Replikationsstudien kommen sie zu dem Ergebnis, daß die Beziehungen zwischen Lehrerkommentaren bei Zensuren und Leistungssteigerungen nicht eindeutig gesichert sind. Durch die Einführung von Randbedingungen soll hier nun versucht werden, diese Widersprüche aufzulösen und spezifische Hypothesen über die Wirkung von Lehrerkommentaren aufzustellen und eine dieser Hypothesen empirisch zu überprüfen.

Zunächst ist festzustellen, daß das von Page (1958) gewählte statistische Vorgehen des Rangplatzvergleichs der Leistungen von *drei* Schülern zu t_1 (1. Test) und t_2 (2. Test) ein für Veränderungen sehr sensibles Verfahren ist; alternative Auswertungsmodi sind hier zu präferieren. Wichtiger ist jedoch, daß inhaltlich bei notenbezogenen Lehrerkommentaren in Anlehnung an verschiedene Beurteilungsmodelle und ihnen unterliegende Normorientierungen (siehe etwa Zielinski, 1974; Rheinberg, 1980) verschiedene Typen unterschieden werden sollten:

Typ 1: Am sozialen Bezugssystem orientierte Kommentare, in denen notenorientiert die Stellung des Schülers im klassenbezogenen Leistungsvergleich hervorgehoben wird. Bei der verbalen, mehr oder weniger festgelegten Umschreibung der Note (Versuchsbedingung b bei Page, 1958) steht also eine soziale Bezugsnorm-Orientierung des Lehrers im Vordergrund, von der — je nach Leistungsstand des Schülers — positive (bei guten Schülern) und negative (bei schlechten Schülern) motivationale Effekte zu erwarten sind (vgl. Rheinberg, 1980).

Typ 2: An einem sachlichen Gütemaßstab orientierte Kommentare, die den Leistungsstand des Schülers mit Bezug auf den Lernstoff fokussieren, also auf eventuelle Fehler, gute Lösungsansätze, gute Antworten etc. ausführlicher eingehen. Hier sind ebenfalls nur bedingt positive Effekte zu erwarten, da die spezielle Lern- und Leistungsgeschichte des Schülers nicht beachtet wird. Sie geben dem Schüler allerdings gezielte Hinweise auf Lerndefizite und die fachliche Qualität seiner Leistungen, was ihm seinen sachbezogenen Kenntnisstand verdeutlicht, woraus zumindest bei leistungsschwächeren Schülern eine Überlegenheit dieses Typs gegenüber dem Typ 1 abgeleitet werden kann.

Typ 3: An einem intraindividuellen Leistungsvergleich orientierte Kommentare, die auf der Leistungskurve des einzelnen Schülers basieren. Bei diesem Typ ist aufgrund der individuellen Bezugsnorm-Orientierung des Lehrers mit positiven motivationalen Folgen zu rechnen (vgl. auch Rheinberg, 1980). Im Gegensatz zu den Typen 1 und 2 setzt dieses Vorgehen allerdings eine vertiefte Kenntnis der einzelnen Schüler bei dem Lehrer voraus (über Lernkurve, frühere Leistungen, besondere Schwächen, Versäumnisse, Lernsituationen etc.). Bei diesem Kommentartyp können insbesondere auch sozial-emotionale Aspekte der spezifischen Lehrer-Schüler-Beziehung verstärkt miteingehen, was die persönliche Bedeutung der Leistungsbewertung für den Schüler erhöhen kann. Butterworth & Michael (1975) konnten zeigen, daß solche, an einem intraindividuellen Vergleich orientierte Kommentare sowohl in späteren Leistungstests als auch in Fragebogen zu Schuleinstellungen und zu

leistungsbezogenen Kontrollüberzeugungen positive Effekte hatten. Für am sozialen Bezugssystem orientierte Kommentare (Typ 1) konnten diese Effekte nicht nachgewiesen werden.

Mischtypen: Lehrerkommentare, die diesem Typ angehören, weisen mit unterschiedlicher Wichtung Aspekte aller drei Typen auf. Die von Page (1958) verwendeten „freien“ Kommentare (Versuchsbedingung c) können diesem Mischtypus zugeordnet werden. Je nach Bezugsnorm-Orientierung, pädagogischer Einstellung und Schülerkenntnis der Lehrer sind hier unterschiedlichste Gewichtungen der drei Kommentartypen möglich.

Ähnlich wie auch bei dem sogenannten „Pygmalion-Effekt“ muß bei der Analyse und dem Einsatz von Lehrerkommentaren auf eine weitere Randbedingung geachtet werden: Die in Frage stehenden Leistungen müssen von den Schülern überhaupt (evtl. sogar kurzfristig) zu erreichen sein. D. h., daß mögliche Leistungsrückstände oder -versäumnisse nicht zu groß und die Schüler auch nicht durch objektive Fähigkeitsmängel im Unterricht überfordert sein dürfen. Ferner sollten neben späteren Leistungen und Noten auch kognitiv-motivationale und emotionale Variablen auf der Schülerseite als Kriteriumsvariablen beachtet werden, da durch sie eventuell später auftretende Leistungsveränderungen (direkt oder auch erst mit der Zeit, also in kumulativer Form) bedingt werden können. Solche Variablen können auch Aussagen über die Realisierung genuin pädagogischer (also motivationaler) Funktionszuschreibungen zu Zensuren (hier ergänzt um schriftliche Kommentare) erlauben.

In der vorliegenden Studie wird exemplarisch die Wirkung von an einem sachlichen Gütemaßstab orientierten Kommentaren (Typ 2) auf spätere Leistungen und — ergänzend — auf kognitiv-motivationale Variablen (Leistungsvorsätze, Zufriedenheit mit der Note und Kausalattributionen für die Note) sowie Freude über die erhaltene Leistungsrückmeldung im Vergleich zu Noten allein (ohne Kommentar) untersucht. Da keine Daten zur individuellen Leistungskurve vorlagen (Psychologie als neues Fach bei Fachschülerinnen) war der Einbezug von Typ 3-Komentaren nicht möglich. Auf eine Fokussierung des sozialen Vergleichs als eine weitere Versuchsbedingung (Typ 1) wurde verzichtet, da der soziale Vergleich durch die Noten unter beiden Versuchsbedingungen für die Schülerinnen möglich war.

Methode

An der Untersuchung waren 131 Fachschülerinnen (angehende Fachlehrerinnen) im Alter von 18 bis 23 Jahren beteiligt. Im Rahmen ihrer regulären Ausbildung erhielten sie Unterricht in Pädagogischer Psychologie (eine zweistündige Vorlesung pro Woche für alle Fachschülerinnen), der in neun nach dem Zufall von der Schulverwaltung zusammengestellten Seminaren (einstündig) vertieft wurde. Im Abstand von acht Wochen wurden zwei informelle Wissenstests durchgeführt (1. Test zum Themenkomplex „Einführung in die Pädagogische Psychologie“; 2. Test zum Themenkomplex „Gütekriterien und statistische Probleme der Leistungsbeurteilung“), die in der Aufgabenstruktur und dem Auswertungsvorgehen identisch waren (jeweils zwei Mehrfachwahl-Aufgaben und sieben kurze offene Fragen, die objektiv auswertbar waren). Die Leistungen wurden nicht an Hand einer vorgegebenen Notenverteilung (etwa Normalverteilung), sondern nach vorher festgelegten Zuordnungen von Punktwerten und Noten bewertet. Die Tests wurden jeweils eine Woche nach ihrer

Abnahme in den Seminaren zurückgegeben. In drei nach Zufall ausgewählten Seminaren ($N_1 = 48$) wurde zusätzlich zur erreichten Punktzahl und zur Note ein kurzer, am sachlichen Gütemaßstab orientierter Kommentar, der Fehler verbesserte und die Berichtigung erläuterte, gute Antwortansätze und gute Antworten hervorhob, auf dem Testblatt aufgeführt (Versuchsgruppe). Die Kontrollgruppe bestand aus den Teilnehmerinnen der sechs restlichen Seminare ($N_2 = 83$). Den Teilnehmerinnen war zu den Test- und Rückgabeterminen nicht bewußt, daß sie an einem Experiment teilnahmen; die Befunde wurden ihnen später im Rahmen des normalen Unterrichts mitgeteilt.

Unmittelbar nach der Rückgabe des ersten Tests wurden die Fachschülerinnen gebeten, einen einseitigen Fragebogen auszufüllen. Dies wurde mit dem Wunsch des Dozenten begründet, ebenfalls eine Rückmeldung (über seine bisherige Unterrichtsleistung etc.) zu erhalten. Neben anderen Fragen (Einschätzung der Verständlichkeit, der Praxisrelevanz etc. des Lehrstoffs; Beurteilung der Unterrichtsgüte in Vorlesung und Seminaren) beantworteten die Fachschülerinnen auch die folgenden: (1) Sind Sie mit Ihrem Ergebnis in dem Test zufrieden? (Antwort auf 7-stufiger graphischer Schätzskala); (2) Freuen Sie sich, daß Sie eine Rückmeldung zu Ihrem Wissensstand in Psychologie erhalten haben? (Antwortmodus wie bei 1); (3) Auf welche Gründe bzw. Ursachen führen Sie Ihr Testergebnis zurück? Wie erklären Sie sich Ihr Testergebnis? (freie Antwort); (4) Welche Konsequenzen ziehen Sie aus Ihrem Testergebnis? (Antwortmöglichkeiten: (a) ich will für Psychologie weniger tun als bisher, (b) ich arbeite . . . weiter wie bisher, (c) . . . etwas mehr, (d) . . . mehr, (e) . . . viel mehr als bisher).

Ergebnisse

Für die Leistungen im ersten Wissenstest konnten zwischen der Versuchs- und der Kontrollgruppe keine statistisch signifikanten Unterschiede festgestellt werden, $t(129) = 0,48$, *n. s.*. Von 16 möglichen Testpunkten erreichten die Fachschülerinnen in der Versuchsgruppe im Durchschnitt 7,3, die in der Kontrollgruppe 7,2 Punkte, was einer (durchschnittlichen) Note von "4" (ausreichend) entsprach. Bei objektiver Leistungsgleichheit in den beiden Gruppen kann aufgrund des recht schlechten Testergebnisses also mit hinreichend großen Verbesserungsmöglichkeiten gerechnet werden, wobei Leistungsrückstände wegen der relativ kurzen Unterrichtszeit in Psychologie nicht groß sein können. Aufgrund der erhaltenen Rückmeldungen in den Seminarveranstaltungen und der Befunde zu den Kausalattributionen (siehe unten) kann auch ausgeschlossen werden, daß sich die Fachschülerinnen durch den Lehrstoff überfordert sahen.

Bei der Rückgabe des ersten Tests gaben die Teilnehmerinnen der Versuchsgruppe im Durchschnitt eine geringere Zufriedenheit mit ihrem Testergebnis ($\bar{x}_v = 2,9$; $s = 1,7$) an als die aus der Kontrollgruppe ($\bar{x}_k = 4,1$; $s = 1,9$; $t(129) = 3,47$; $p < .001$). Umgekehrt verhält es sich bei der Antwort auf die Frage, ob man sich darüber freut, die Leistungsrückmeldung erhalten zu haben. In der Versuchsgruppe beträgt hier die Angabe im Mittel $\bar{x}_v = 4,8$ ($s = 1,8$), in der Kontrollgruppe dagegen nur $\bar{x}_k = 3,9$ ($s = 2,1$); dieser Mittelwertsunterschied ist bei Varianzhomogenität, die für alle Gruppenvergleiche nachgewiesen ist (Bartlett-Box Test), auf dem 5 %-Niveau signifikant, $t(129) = 2,18$. Auch bei den erhobenen Kausalattributionen für das Testergebnis, die inhaltlich durch zwei Auswerter klassifiziert wurden (Übereinstimmung im ersten Durchgang 92 %) zeigen

sich einige Unterschiede. In der Versuchsgruppe attribuieren 87,9 % der Fachschülerinnen auf mangelnde Vorbereitung, mangelnde Anstrengung o. ä. (also internal-variabel), in der Kontrollgruppe nur 64,3 %. 20,8 % der Versuchsgruppe begründen ihr Testergebnis u. a. durch hohe Stoffschwierigkeit (Kontrollgruppe: 13,3 %) und 12,5 % durch geringe Stoffschwierigkeit (Kontrollgruppe: 1,2 %). Unter der Versuchsbedingung „sachbezogener Lehrerkommentar“ werden also häufiger internal-variable Attributionen vorgenommen, $Chi^2(2) = 26,69$, $p < .001$; Kausalattributionen auf Stoffschwierigkeit treten ebenfalls häufiger auf, wobei insbesondere geringe Stoffschwierigkeit häufiger genannt wird, $Chi^2(2) = 9,73$, $p < .01$.

Bei der quantitativen Auswertung der frei erhobenen Kausalattributionen für die Note im ersten Wissenstest ergab sich in der Gesamtstichprobe das folgende Ergebnis: 75,5 % der Fachschülerinnen attribuieren auf hohe oder geringe Anstrengung (gute/schlechte Vorbereitung o. ä.), 21,4 % auf hohe/geringe Stoffschwierigkeit, 8,4 % auf hohe/geringe Unterrichtsgüte und 6,9 % auf hohes/geringes Interesse am Fach. Weitere Kausalfaktoren, insbesondere die in der Literatur häufig anzutreffenden Attribuierungen auf Zufall (Glück/Pech) und Begabung (das Fach liegt mir/liegt mir nicht), traten bei den frei erhobenen Antworten der Fachschülerinnen nicht auf.

Mit einer dreifaktoriellen Varianzanalyse wurde dann die Abhängigkeit der Leistungen im zweiten Wissenstest (t_2) von (1) der Gruppenzugehörigkeit (Faktorstufen: sachbezogener Kommentar/kein Kommentar), (2) der Note im ersten Test (Faktorenstufen: Noten 1 bis 6) und (3) den nach der Rückgabe des ersten Tests erhobenen Leistungsvorsätzen (5 Faktorstufen; siehe Abschnitt 2) überprüft. Es ergaben sich keine statistisch bedeutsamen Haupteffekte und Interaktionen ($p > .05$). Die Versuchsgruppe (mit Lehrerkommentar) und die Kontrollgruppe (nur Punktwert und Note) unterscheiden sich also in der Testleistung zu t_2 nicht bedeutsam voneinander, $F(1/97) = 1,606$, *n. s.*. Auch die Note im ersten Wissenstest und die nach der Rückgabe des ersten Tests erhobenen Leistungsvorsätze stehen in keiner statistisch signifikanten Beziehung zu der späteren Leistung, $F(4/97) = 1,999$, *n. s.* und $F(3/97) = 0,621$, *n. s.*.

Diskussion

An einem sachlichen Gütemaßstab orientierte Lehrerkomentare zu Noten in schriftlichen Arbeiten führten in der vorliegenden Untersuchung zu Unterschieden in kognitiv-motivationalen und emotionalen Variablen unmittelbar nach der Rückgabe der Arbeit, jedoch nicht zu Leistungsverbesserungen in einem späteren Test. Die Fachschülerinnen der Versuchsgruppe waren bei durchschnittlich gleichen Leistungen im ersten Test mit ihrem Testergebnis weniger zufrieden als die aus der Kontrollgruppe, freuten sich jedoch gleichzeitig stärker darüber, eine Leistungsrückmeldung erhalten zu haben. Ebenso wie bei den z. T. unterschiedlichen Kausalattributionen für die erhaltene Note deuten sich hier motivationale und emotionale Effekte einer an einem sachlichen Gütemaßstab orientierten

Kommentierung von Zensuren an. Die Kommentierung von Fehlern, guten Antwortansätzen, Irrtümern, besonders guten Antworten etc. führt also augenscheinlich zu einer höheren Akzeptanz der Leistungsbeurteilung beim Schüler, zu differenzierteren naiven Begründungen der Leistung und zu einer erhöhten Unzufriedenheit mit der eigenen Leistung, was u. a. durch die Hervorhebung sachlicher Fehler etc. erklärbar ist. Direkte Effekte auf die Leistungen in einem folgenden Wissenstest konnten jedoch nicht festgestellt werden. Hypothetisch kann dies darauf zurückgeführt werden, daß die intraindividuelle Perspektive in der Kommentierung (Typ 3-Kommentar), die zu einer stärkeren persönlichen Involvierung des Schülers führt, fehlt. Ein weiterer Grund kann darin gesehen werden, daß die beachtete Zeitspanne zwischen dem ersten und dem zweiten Wissenstest mit acht Wochen recht niedrig liegt. Eventuell ergeben sich positive Effekte erst (in Form eines kumulativen Effekts) nach einer längeren Zeitspanne, in der die Kommentierung kontinuierlich eingesetzt wird.

Die Ergebnisse stehen in Einklang mit einer der im ersten Abschnitt dargestellten differentiellen Hypothesen zur Wirkung von unterschiedlichen Typen des Lehrerkommentars. Die von Stewart & White (1976) zusammengestellten widersprüchlichen Befunde zum Zusammenhang von Lehrerkommentaren und Leistungssteigerungen können insofern aufgelöst werden, als daß eine Spezifikation des Kommentartypes nötig ist. Überläßt man dem Lehrer ganz die Art des Kommentares („freie Kommentare“ als häufigste Versuchsbedingung), so kann dies je nach Bezugsnorm-Orientierung, Schülerkenntnis etc. des Lehrers zu verschiedenen Wichtungen in den Kommentaren führen, die klare Befunde erschweren. So kann etwa vermutet werden, daß Page (1958) vor allem mit Lehrern zusammengearbeitet hat, die ihre Schüler längere Zeit kannten und evtl. auch eher eine individuelle Bezugsnorm-Orientierung vertraten, was Kommentare des Typus 3 gestattet und wahrscheinlich macht. In Replikationsstudien mit widersprüchlichen Befunden könnten dagegen etwa vermehrt Lehrer beteiligt gewesen sein, die die zu beurteilenden Schüler weniger gut kannten, die über soziale Bezugsnorm-Orientierungen verfügen etc., was zu Kommentaren des Typ 1 oder 2 führen würde.

Durch Butterworth & Michael (1975) sind die positiven Effekte von Lehrerkommentaren des Typus 3 (intraindividuell orientiert) auf Leistungs- und Motivationsvariable belegt. Gleichzeitig konnten sie für den Typ 1-Kommentar (am sozialen Vergleich orientiert) keine positiven Effekte feststellen, was mit den Befunden von Page (1958) und denen von einigen Replikationsstudien (im Überblick siehe Stewart & White, 1976) insofern übereinstimmt, als daß für die Versuchsbedingung „spezifizierte Kommentare“ (vorher festgelegte, an den Notestufen orientierten Anmerkungen) keine oder nur geringe Effekte ermittelt werden konnten. In der vorliegenden Untersuchung konnten nun für Kommentare des Typus 2 (an einem sachlichen Gütemaßstab orientierte Anmerkungen) zwar kurzfristig motivationale und emotionale Effekte festgestellt werden, die

jedoch — ebenfalls relativ kurzfristig — nicht zu einer bedeutsamen Leistungssteigerung führten. Obwohl Lehrerkommentare in der Praxis wohl meist dem Mischtypus angehören, ist eine weitere Abklärung der Wirkungen der verschiedenen Kommentartypen notwendig, um zu motivational positiven Kombinationen der verschiedenen Aspekte in der Kommentierung zu gelangen (evtl. auch je nach Leistungsstand des Schülers unterschiedlich!) und somit die Annahme, irgendetwelche schriftlichen Anmerkungen zu Zensuren hätten beim Schüler schon positive Effekte, durch Präzisierungen zu korrigieren.

Summary

Starting from a differentiation of types of teacher comments, the effects of comments which are oriented at the subject matter (treatment group, $N = 48$) are analysed in comparison to no comments (control group, $N = 83$). Results show that — although the grades are not significantly different — (1) the treatment group is less satisfied with their performance than the control group, (2) the treatment group is more pleased about having a feedback, and (3) the treatment group attributes their performance more often to effort and to difficulty of the subject matter. In a following (second) informal test there was no significant difference in the performance of the two groups. The discussion refers to the effects of the various types of teacher comments on performance and motivational variables.

Literatur

- Butterworth, T. W. & Michael, W. B., The relationship of reading achievement, school attitude, and self-responsibility behaviors of 6-grade pupils to comparative and individualized reporting systems. *Educational & Psychological Measurement*, 1975, **35**, 987—991.
- Campbell, D. T. & Stanley, J. C., Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Hg.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963 (Deutsche Bearbeitung in K. Ingenkamp (Hg.), *Strategien der Unterrichtsforschung*. Weinheim: Beltz, 1973. S. 99—194).
- Gage, N. L. & Berliner, D. C., *Pädagogische Psychologie*. München: Urban & Schwarzenberg, 1977.
- Page, E. B., Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology*, 1958, **49**, 173—181.
- Rheinberg, F., *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe, 1980.
- Schmidt, W., Darstellung und Verwendung der Leistungsmessung und Leistungsbeurteilung in der Schule. In E. Stephan & W. Schmidt (Hg.), *Messen und Beurteilen von Schülerleistungen*. München: Urban & Schwarzenberg, 1978. S. 153—171.
- Stewart, L. G. & White, M. A., Teacher comments, letter grades, and student performance: What do we really know? *Journal of Educational Psychology*, 1976, **68**, 488—500.
- Zielinski, W., Verfahren zur Beurteilung des Unterrichts. In F. E. Weinert, C. F. Graumann & M. Hofer (Hg.), *Funk-Kolleg Pädagogische Psychologie*, Band 2. Frankfurt/Main: Fischer, 1974. S. 901—923.

Anschrift des Verfassers:

Dr. Günter Krampen, Dipl.-Psych., Universität Trier, FB I — Psychologie
Schneidershof, 5500 Trier