

# Bedingungsanalysen von (erwartungswidrigen) Schulleistungen

Franz E. Weinert, Wolfgang Schneider und Bernhard Treiber

Kaum jemand erwartet, daß ein guter Schreiner, Schlosser oder Kraftfahrzeug-Mechaniker auch notwendigerweise ein besonders intelligenter Mensch ist. Demgegenüber wird der Zusammenhang zwischen Intelligenztestleistung und Schulerfolg als beinahe naturnotwendig angesehen. Und die empirischen Ergebnisse scheinen diese Erwartung zu bestätigen, erweist sich doch die Intelligenz als bester einzelner Prädiktor für Schulleistungen. Dies ist allerdings nicht erstaunlich, denn die Mehrzahl der Intelligenztests für Kinder und Jugendliche wurde zu dem Zweck entwickelt, Schulleistungen zu prognostizieren. Dementsprechend erfolgte ihre Validierung auch an diesem Kriterium. Inwieweit damit lediglich eine zirkuläre Beziehung zwischen ‚academic intelligence‘ und ‚academic achievement‘ gestiftet wurde, ist bis heute umstritten; d.h. es ist ungeklärt, inwieweit Intelligenz- und Schulleistungswerte auch prognostische Validität für die Lösungswahrscheinlichkeit und -qualität von Alltagsproblemen und für die Bewältigung der Aufgaben nicht-akademischer Berufe aufweisen (Lave, 1977; Ginsburg, 1977; Neisser, 1976).

Für den schulischen Bereich aber gilt uneingeschränkt der postulierte Zusammenhang zwischen Intelligenz und Schulleistung. Obwohl in der Regel nur korrelative Beziehungen ermittelt werden, wird der Intelligenz häufig die Funktion einer erklärungs-mächtigen und prognostisch validen, unabhängigen Variable zugeschrieben. Dabei gehen viele theoretische Modelle explizit oder häufiger implizit davon aus, daß intellektuelle Fähigkeiten eine zwar notwendige, aber nicht in allen Fällen hinreichende Bedingung für erfolgreiches Lernen akademischer Inhalte und damit für Schulleistungen darstellen.

Aufgrund dieser Annahme erscheint es zulässig, mit Hilfe interindividueller Intelligenztest-Unterschiede korrespondierende Schulleistungsdifferenzen vorherzusagen und Diskrepanzen zwischen diesen erwarteten und den tatsächlich erzielten Schulleistungen zum Anlaß zu nehmen, im allgemeinen wie im individuellen Fall nach zusätzlichen Bedingungen von Schulerfolg und -mißerfolg zu suchen. Nicht selten wird dabei von vornherein unterstellt, daß bei den meisten Schülern

die Intelligenz zugleich notwendige und hinreichende Bedingung für akademische Lernleistungen darstellt; – dies auch deshalb, weil für die Entwicklung intellektueller Fähigkeiten und für die Bearbeitung von Intelligenztest-Aufgaben ebenso (minimale) motivationale Voraussetzungen erforderlich sind wie für die erfolgreiche Bewältigung schulischer Leistungsanforderungen.

Das besondere Interesse der Forscher und Praktiker galt dementsprechend in den vergangenen Jahren Schülern mit erwartungswidrigen Schulleistungen, insbesondere jenen, die weniger leisten, als aufgrund ihrer gemessenen Intelligenz erwartet wird (*underachiever*).

Mit der Unterscheidung zwischen erwartungsgemäß und erwartungswidrig schlechten Schulleistungen scheint zugleich eine theoretisch begründete und praktisch brauchbare pädagogische Indikations-Kontraindikationsstellung möglich zu sein: Danach wären die Lehr-Lern-Anforderungen zu reduzieren, wenn erheblich unter dem Durchschnitt liegende Intelligenz- und Schulleistungen übereinstimmen; pädagogisch-psychologische Interventionsprogramme zur Verbesserung der Lernmotivation und Arbeitseffektivität sollten demgegenüber immer dann indiziert sein, wenn die Schulleistungen im Vergleich zu den Intelligenztest-Ergebnissen erwartungswidrig schlecht sind.

Stillschweigend kommt in dieser Unterscheidung auch die Erwartung zum Ausdruck, daß es sich bei der Intelligenz um ein besonders änderungsresistentes psychologisches Merkmal handelt, während zusätzliche leistungshemmende Faktoren (z.B. motivationale Bedingungen) durch pädagogische Maßnahmen als leichter beeinflussbar angesehen werden.

Diese durch theoriegeleitete empirische Forschung nur wenig elaborierten und präzisierten Plausibilitätsannahmen bilden die schmale Basis für die auch heute noch dominierende psychometrische Konzeptualisierung erwartungswidriger Schulleistungen. Die Art der Berücksichtigung von Intelligenztest- und Schulleistungskennwerten und die darauf aufbauenden verschiedenen methodischen Varianten zur Klassifikation erwartungsgemäßer und erwartungswidriger Lernergebnisse sind an anderer Stelle ausführlich beschrieben worden (Wahl, 1975; Weinert & Petermann, in diesem Band). Das gilt auch für die Kritik an diesem Konzept, die sich auf die methodischen Probleme (Thorn-dike, 1963; Wahl, 1975), die theoretischen Defizite (Atkinson, 1974; Krug & Rheinberg, in diesem Band; Weinert & Petermann, in diesem Band) und die praktischen Folgerungen (Yule, Rutter, Berger & Thompson, 1974) konzentriert.

Der gesamte Forschungsansatz zur Definition, Analyse und Behandlung erwartungswidrig schlechter Schulleistungen wie die vehemente Kritik an den dabei benutzten Konzeptualisierungen, Operationalisierungen und Schlußfolgerungen weisen auf eine Reihe grundlegender Probleme bei der Bedingungsanalyse von Schulleistungen hin. Im einzelnen:

a) Schulleistungen sind zum größten Teil das Ergebnis kumulativer Lernprozesse, die sich über einen längeren Zeitraum erstrecken, so daß Quantität und Qualität des Lernens zu verschiedenen Zeitpunkten unterschiedliche Determinationsmuster aufweisen können.

Ein allgemeines, statisches Prognosemodell auf der Basis relativ stabiler Personmerkmale der Lernenden (wie es bei der *underachievement*-Forschung üblich ist) wird diesem Verlaufsaspekt nicht gerecht. Die Abhängigkeit der einzelnen Leistungen von ihrer Position innerhalb eines sachlogisch vorgegebenen, didaktisch gesteuerten oder idiosynkratisch sich entwickelnden Lernverlaufs erfordert wenigstens die Berücksichtigung spezifischer Vorkenntnisse und aktueller Motivationen neben relativ stabilen Fähigkeiten und Motiven (Bloom, 1976; Keeves, 1972).

b) Lernerfahrungen und Schulleistungen wirken auf die Bedingungen des Lernens und Leistens zurück und beeinflussen deshalb die künftige Leistungsentwicklung. Das gilt nicht nur für relevante Vorkenntnisse und aktuelle Motivationen, sondern auch für die vom einzelnen Schüler wahrgenommenen Instruktions- und Interaktionscharakteristika des Lehrers. Dies könnte im Zeitlängsschnitt für umschriebene Gruppen von Schülern zu einer Homogenisierung der Prädiktionsmuster und zugleich zu einer Vergrößerung der Differenzen zwischen den Lernbedingungen verschiedener Schülertypen führen.

c) Ergebnisse der Instruktions- und der Motivationsforschung sprechen dafür, daß unter schulischen Bedingungen kein direkter Zusammenhang zwischen den Fähigkeiten und Motiven der Lernenden auf der einen und ihrer Leistungen auf der anderen Seite erwartet werden darf, sondern daß diese Beziehung durch schulklassentypische Merkmale der Instruktionsqualität und der Motivierungsbedingungen (im Sinne situativ zugänglicher, motivspezifischer Anregungen und Anreize) beeinflusst wird. Eine ungeprüfte Aggregation von Schülerdaten über verschiedene Schulklassen hinweg zum Zwecke der Erklärung oder Prognose von Leistungen ist unter diesen Umständen äußerst problematisch.

d) Nach Auffassung von Krug & Rheinberg (in diesem Band) benötigen wir für die Analyse von Schulleistungen sowohl Koppelungs- wie

Kompensationsmodelle zur Beschreibung notwendiger und hinreichender Lernbedingungen. Während zur Erbringung der curricular geforderten Schulleistungen stets ein Minimum an Fähigkeiten und Anstrengung auf Seiten der Schüler notwendig ist (minimales Kopplungsmodell), gibt es im mittleren Schwierigkeitsbereich offenbar Möglichkeiten abgestufter Kompensationen zwischen Fähigkeiten, Vorkenntnissen, Anstrengungen und verfügbaren Lernhilfen. Bei sehr schwierigen Aufgaben nehmen diese Kompensationsmöglichkeiten ab, so daß ein maximales Koppelungsmodell angemessen erscheint. Es muß also davon ausgegangen werden, daß kompensierbare und nicht-kompensierbare Bedingungen von Schulleistungen zwischen verschiedenen Schulklassen, Schülern, Aufgaben und Zeitpunkten erheblich variieren können. Konsequenz daraus ist die programmatische Forderung, nicht nur – wie bisher üblich – nach notwendigen Bedingungen schulischer Leistungen zu suchen, sondern die Analyse populations- und aufgabenspezifischer hinreichender Bedingungssätze verstärkt in Angriff zu nehmen.

e) Bei dieser Suche nach vorfindbaren oder konstruierbaren hinreichenden Bedingungen des Schulerfolgs sollte die pädagogische Beeinflußbarkeit der berücksichtigten Variablen bedacht werden, will man auch praktischen Nutzen aus der Forschungsarbeit ziehen. So haben sich z.B. in vielen Studien kognitive Fähigkeiten als sehr änderungsresistent erwiesen; Vorkenntnislücken sind demgegenüber wesentlich leichter beeinflussbar. Das gilt auch für die Lernmotivation und in abgeschwächter Form für die Veränderbarkeit von (dispositional definierten) Motiven der Schüler. Die Unterrichtsqualität schließlich hat sich, in Abhängigkeit von Personmerkmalen der Lehrer und von der Art der Änderungskriterien, als unterschiedlich gut modifizierbar erwiesen.

Die skizzierten, für die Analyse von Schulleistungen typischen Probleme wurden bei der Planung einer empirischen Untersuchung, über die im folgenden berichtet wird, berücksichtigt. Ziel dieser Erkundungsstudie ist die schrittweise Analyse von Mathematikleistungen der Schüler fünfter (bzw. sechster) Hauptschulklassen. Dabei wird zwischen ‚leichten‘ und ‚schweren‘ Lernaufgaben unterschieden, und ein sparsamer Satz theoretisch gut begründeter, unterschiedlich änderungsresistenter Bedingungsvariablen findet Verwendung (Intelligenz, Vorkenntnisse, Lernmotivation der Schüler, von Schülern wahrgenommene Instruktionsqualität und Motivierungsqualität des Unterrichts). Die vorgesehene schrittweise Analyse der Schulleistungen bezieht folgende einzelne Fragestellungen ein:

(1) Ergeben sich bei schulklassen-unspezifischer Datenverknüpfung konsistente und bedeutsame Zusammenhangsmuster zwischen Schülermerkmalen und Leistungskriterium? Das wird aufgrund der fehlenden Berücksichtigung moderierender Unterrichtsvariablen nicht erwartet.

(2) Ergeben sich bei schulklassen-spezifischer Datenverknüpfung konsistente und bedeutsame Beziehungen zwischen Schülermerkmalen und dem Leistungskriterium, und wie different sind die Zusammenhangsmuster zwischen verschiedenen Schulklassen?

Erwartet werden erhebliche Inter-Schulklassen-Differenzen bei der Prädiktion von Schulleistungen, wobei die Stabilität der Schätzungen wegen der jeweils zu geringen Stichprobengröße unbefriedigend bleiben muß.

(3) Lassen sich Typen von Schulklassen identifizieren, für die konsistente und bedeutsame Zusammenhänge zwischen Schülermerkmalen und Leistungskriterium nachweisbar sind?

Wir haben in der vorliegenden Studie als Gruppierungskriterium die Höhe des Determinationskoeffizienten der Kriteriumsleistung durch die Intelligenztestwerte verwendet. Die unterschiedliche Vorhersagbarkeit der Schulleistungen durch kognitive Merkmale der Schüler diskriminiert nämlich nach vorliegenden empirischen Befunden gut zwischen Schulklassen und ist als Indikator für die didaktische (insbesondere die remediale) Qualität des Unterrichts interpretiert worden (Simons, Weinert & Ahrens, 1975). Gefragt werden muß in diesem Zusammenhang natürlich, ob sich eine identische Gruppierung der Schulklassen ergibt, wenn man andere Klassifikationsmerkmale, z.B. die von Schülern wahrgenommene Instruktions- und Motivationsqualität des Unterrichts heranzieht. Schließlich interessieren jene Charakteristika des Unterrichts, die am stärksten zwischen den Klassen differenzieren, weil auf diese Weise überprüft werden kann, ob die empirisch ermittelten Typen theoretische Relevanz besitzen.

(4) Läßt sich durch die Kombination von Schulklassentypen, Aufgabenschwierigkeiten und Lernniveau der Schüler eine theoretisch befriedigende Differenzierung der Schulleistungsdeterminanten durch die verwendeten Variablengruppen (Intelligenz, Vorkenntnisse, Lernmotivation, Instruktions- und Motivationsqualität des Unterrichts) erreichen?

## Methoden

Die folgenden Variablen wurden herangezogen:

**Mathematiktest.** 31 Mathematikaufgaben eines eigens konstruierten Leistungstests, die als curriculumvalide für den Mathematikunterricht sechster Hauptschulklassen gelten konnten und sich als Rasch-skalierbar erwiesen, wurden in Mehrfach-Wahlantwort-Form präsentiert und waren nach zunehmender Schwierigkeit angeordnet. Ihre mittlere Schwierigkeit betrug .53, ihre mittlere Trennschärfe .19 und Cronbachs Alpha .73.

**Testintelligenz.** Kognitive Fähigkeitsunterschiede zwischen den Schülern wurden durch mehrere Intelligenzdimensionen beschrieben:

**Sprachverständnis:** Wortschatz und verbales Verständnis als Subtest VI des Kognitiven Fähigkeits-Tests (KFT 4–13) von Heller, Gaedike & Weinläder (1976);

**Sprachlogisches Verständnis (Analogie)** als Subtest 2 der ‚Aufgaben zum Nachdenken‘ (AzN4+) von Hylla & Kraak (1970).

**Rechenfertigkeit:** Subtest 9 (Einfache Additionsaufgaben) des ‚Prüf-systems für Schul- und Bildungsberatung‘ (PSB) von Horn (1969);

**Induktion:** Erkennen von Regeln und Gesetzen (Zahlenreihen) des Subtests Q3 des KFT 4–13;

**Veranschaulichung:** Subtest 7 (Falttest) des PSB;

**Schnelligkeit:** Subtest 10 (Zahlenreihenvergleich) des PSB.

**Vorkenntnisse.** Vier Grundmengen von Testaufgaben wurden hierarchisch geordnet und wie folgt definiert:

**Oberste Vorkenntnisebene:** Division 5- und 6stelliger Dividenden durch 2stellige Divisoren. Beispiel:  $623\ 193 : 31$

**Obere Vorkenntnisebene:** Division 3- und 4stelliger Dividenden durch 2stellige Divisoren. Beispiel:  $925 : 25$

**Mittlere Vorkenntnisebene:** Division 5- und 6stelliger Dividenden durch einstellige Divisoren. Beispiel:  $845\ 971 : 7$

**Untere Vorkenntnisebene:** Division 3- und 4stelliger Dividenden durch einstellige Divisoren. Beispiel:  $792 : 4$

Die Annahme einer hierarchischen Struktur dieser Aufgabenmengen ließ sich im wesentlichen bestätigen, und zwar (vgl. Klauer, 1976) – durch den Nachweis zunehmender Schwierigkeit, je nach Höhe der Hierarchieebene (.49, .64, .76 und .84); – durch den Nachweis der Vorhersagbarkeit der Kriteriumsleistung durch die untergeordneten Vorkenntnis-

leistungen ( $R_{1,234} = .534$ ); – durch den Nachweis, die Kriteriumsleistungen durch Leistungen auf der übernächsten und allen folgenden Hierarchieebenen nicht mehr besser als durch die kriteriumsnächsten Vorkenntnisleistungen vorhersagen zu können.

Die Überprüfung dieser Annahmen wird ausführlicher in Treiber (in Vorb.) beschrieben.

*Schülermotivation.* Ein Schülerfragebogen enthielt u.a. 17 Items zur Erfassung motivationaler Merkmale, die sich in ‚Stimmt – Stimmt nicht‘-Form auf die Schulunlust („In der Schule machen mir eigentlich nur wenige Dinge Spaß“), Selbstwertgefühl („Die anderen Schüler in der Klasse können oft vieles besser als ich“.), Prüfungsangst („In Mathematikaufgaben mache ich oft viele Fehler, weil ich zu aufgeregt bin“.) und Erfolgszuversicht („Vor neuen Aufgaben bin ich sicher, daß ich sie schaffen werden“.) der Schüler bezogen.

*Unterrichtsqualität.* 17 Items des bereits erwähnten Schülerfragebogens bezogen sich auf Aspekte der Unterrichtsqualität aus der Sicht der Schüler und zwar – auf die Instruktionsqualität des unterrichtenden Lehrers (7 Items; Beispielitems: „Verstehst Du Deinen Mathematiklehrer, wenn er etwas Neues erklärt?“ „Hilft Dir Dein Lehrer, wenn Du eine Mathematikaufgabe nicht verstanden hast?“); – auf die Motivierungsqualität des unterrichtenden Lehrers (10 Items; Beispielitems: „Der Unterricht in Mathematik ist oft langweilig“; „Der Lehrer lobt uns, wenn wir gut mitarbeiten“.).

*Personstichprobe.* Die Daten entstammen einer größeren, von der Deutschen Forschungsgemeinschaft unterstützten, einjährigen Längsschnitterhebung der kognitiven Leistungsentwicklung von 2 200 Schülern im Mathematikunterricht. Fünfzehn fünfte (und im nachfolgenden Schuljahr sechste) Hauptschulklassen ( $N = 402$ ) mit einem für die Zwecke dieser Untersuchung vollständigen Datensatz wurden davon ausgewählt.

*Untersuchungsablauf.* Die Variablenenerhebung verteilte sich über die Zeit der Längsschnittstudie: Im Dezember 1976 wurden die Testintelligenz-, Schülermotivations- sowie Unterrichtsqualitäts-Variablen in insgesamt zwei Schulstunden erhoben; im Juni 1977 folgte die Erhebung der Mathematik-Vorkenntnisse in einer weiteren Schulstunde; und Ende August 1977 wurde schließlich der Mathematiktest bearbeitet. Banknachbarn bearbeiteten jeweils verschiedene Testformen (A, B). Die Aufsicht führte ein Testleiter; der Fachlehrer war während der Erhebung nicht anwesend.

*Ergebnisse*  
*Klassenunspezifische Determination von Schulleistungen*

Im ersten Schritt wurden dafür die Mathematikleistungen sämtlicher 402 Schüler in den aufgrund der Schwierigkeitsindizes als ‚leicht‘ bzw. ‚schwer‘ eingestuften Aufgaben mit den Merkmalsgruppen – Testintelligenz (5 Subtests), – Vorkenntnisse (4 Kenntnistests) und – Schülermotivation (17 Fragebogenitems) verknüpft. Die Ergebnisse multipler Korrelationsanalysen, die für diese Merkmalsgruppen jeweils getrennt gerechnet wurden, zeigt Tab. 1. Danach bleibt der durch die Linearkombination von Prädiktoren innerhalb dieser Merkmalsgruppen jeweils maximal erklärbarer Anteil an den Schülerleistungsunterschieden bei schweren und leichten Mathematikaufgaben vergleichsweise gering; insbesondere erscheinen varianzpartitive Folgeanalysen zur weiteren Aufhellung genereller bzw. differentieller Variablenzusammenhänge danach aussichtslos. Andererseits läßt dieses Ergebnismuster aber auch die Möglichkeit offen, daß differentielle Merkmalsbeziehungen unterhalb der ‚über alle Schüler‘ gerechneten Korrelations- und Regressionsanalysen durch diese Globalauswertung (vgl. Cronbach & Snow, 1977) maskiert werden und z.B. erst bei schulklassen-spezifischer Analyse nachweisbar sind.

Tab. 1. Determinationskoeffizienten ( $R^2$ ) für die zusammengefaßte Vorhersage der individuellen Mathematikleistung von 402 Hauptschülern aus 15 sechsten Klassen bei leichten und schweren Aufgaben durch verschiedene Schülermerkmale.

Variablengruppe	Variablenanzahl	Aufgaben	
		leichte $R^2$	schwere $R^2$
Testintelligenz	6	.03	.027
Vorkenntnisse	4	.00	.001
Schülermotivation	13	.13	.044

*Klassenspezifische Determination von Schulleistungen*

Zur Überprüfung dieser Frage wurde der beschriebene Variablensatz mit Hilfe multipler Korrelationsverfahren getrennt für jede Klasse analysiert. Dabei zeigten sich – wie erwartet – zwischen den Klassen erheblich variierende Variablenzusammenhänge. Die vergleichsweise geringe Stichprobengröße bei schulklassen-spezifischer Auswertung

Tab. 2. Unkorrigierte und (in Klammern) adjustierte Determinationskoeffizienten bei schulklassenspezifischer Vorhersage der individuellen Mathematikleistung für leichte und schwere Aufgaben durch Variablen der Merkmalsgruppen ‚Intelligenz‘, ‚Vorkenntnisse‘ und ‚Schülermotivation‘ in 15 sechsten Hauptschulklassen.

Schulklasse Nr.	N	leichte Aufgaben			schwere Aufgaben		
		Intelligenz	Vorkenntnisse	Motivation	Intelligenz	Vorkenntnisse	Motivation
0102	23	.35 ( .09)	.04 (−.17)	.60 (−1.09)	.31 ( .30)	.43 ( .30)	.57 (−1.23)
0201	31	.13 (−.08)	.04 (−.10)	.64 ( .17)	.28 ( .10)	.04 (−.10)	.74 ( .40)
0301	29	.39 ( .22)	.15 ( .01)	.75 ( .36)	.19 (−.02)	.24 ( .12)	.84 ( .60)
0802	33	.15 (−.04)	.09 (−.03)	.55 ( .04)	.21 ( .03)	.21 ( .10)	.62 ( .21)
1205	26	.11 (−.17)	.10 (−.07)	.57 (−.33)	.06 (−.226)	.23 ( .09)	.82 ( .45)
2201	24	.09 (−.23)	.21 ( .03)	.57 (−.64)	.30 ( .06)	.29 ( .15)	.69 (−.16)
2401	28	.28 ( .08)	.18 ( .04)	.64 ( .04)	.19 (−.03)	.13 (−.01)	.61 (−.05)
2402	26	.11 (−.16)	.09 (−.08)	.66 (−.06)	.49 ( .32)	.10 (−.06)	.82 ( .45)
3202	32	.13 (−.08)	.01 (−.14)	.58 ( .06)	.29 ( .12)	.15 ( .02)	.56 ( .02)
3301	28	.15 (−.09)	.14 (−.00)	.61 ( .04)	.34 ( .16)	.05 (−.11)	.66 ( .08)
3302	26	.12 (−.16)	.15 (−.00)	.68 ( .02)	.30 ( .08)	.39 ( .28)	.70 ( .07)
3401	24	.23 (−.03)	.06 (−.14)	.64 (−.38)	.14 (−.16)	.14 (−.04)	.73 (−.00)
2601	26	.19 (−.05)	.05 (−.14)	.76 ( .36)	.03 (−.27)	.12 (−.04)	.72 ( .13)
2701	24	.42 ( .22)	.18 ( .01)	.86 ( .49)	.47 ( .28)	.10 (−.08)	.76 ( .11)
4201	21	.34 ( .07)	.38 ( .22)	.88 ( .21)	.35 ( .08)	.38 ( .23)	.63 (−1.44)
	402	.2127	.1247	.666	.261	.200	.698
min.		.09	.01	.55	.03	.04	.56
max.		.42	.38	.88	.47	.43	.84

führte jedoch zu deutlich instabilen Schätzungen der Determinationskoeffizienten bei multipler Vorhersage der Schülerleistungen. Dies verdeutlicht der Vergleich der unkorrigierten mit den adjustierten Determinationskoeffizienten. Eine verlässliche multivariate Koeffizientenschätzung und Annahmenüberprüfung in separaten Schulklassenanalysen erwies sich damit als nicht möglich (vgl. Cronbach, 1976).

### *Klassentypspezifische Determination von Schulleistungen*

Im dritten Schritt ging es darum, eine theoretisch begründete und prüfmethodisch nutzbare Beschreibungs- und Verknüpfungsebene zwischen der Schulklassenebene (die allenfalls instabile Schätzungen ermöglicht) und der Schülerebene (die unterrichtsbedingte Gleichungsdifferenzen zwischen Klassen maskiert) zu finden. Dafür bot sich eine Typisierung der untersuchten 15 Schulklassen mit einigen wenigen Kategorien an. Die Klassifikation dieser Schulklassen nach dem Vorhersagegewicht der Testintelligenz für die Schülerleistung bei schweren Mathematikaufgaben verwies dabei auf eine Möglichkeit, eine geeignete Zwischenebene zu etablieren.

Der Zusammenhang von kognitiver Eingangsfähigkeit ('Testintelligenz') und kognitiven Schülerleistungen wird vielfach als Indikator herangezogen, um beschreiben zu können, ob und in welchem Ausmaß es in Schule und Unterricht gelingt, außerschulisch bedingte Schülerdifferenzen zu reduzieren und alternativ dazu das Vorhersagegewicht leichter veränderbarer Schülermerkmale (wie Vorkenntnisse oder Lernmotivation) zu erhöhen. In dieser Indikatorfunktion ist die Enge der Korrelation zwischen Schulleistungen und Testintelligenz einer der Aspekte der Effektivität und Qualität von Unterricht (vgl. Simons et al., 1975).

Danach wurde der Determinationskoeffizient ( $R^2$ ) in jeder Klasse zur Vorhersage der Schülerleistungen bei schweren Mathematikaufgaben durch die sechs Intelligenz-Subtests herangezogen, um jeweils drei Klassen mit besonders hohem und besonders niedrigem Testintelligenz-Vorhersagegewicht zu zwei Gruppen ( $n_1 = 76$ ;  $n_2 = 78$ ) zusammenzufassen. Diese Extremgruppen-Aufteilung ist in Tab. 3 dargestellt.

Zusätzliche Dateninspektionen sicherten zudem, daß die unterschiedlichen Vorhersagegewichte nicht auf Varianzheterogenität in Schulleistung und/oder Testintelligenz zwischen diesen Klassen zurückgeführt werden können. Es erschien deshalb zulässig, Klassen mit geringer

Tab. 3. Extremgruppenaufgliederung von sechs ausgewählten Klassen nach dem Determinationskoeffizienten ( $R^2$ ) bei Vorhersage der Schülertestleistung in schweren Mathematikaufgaben durch sechs Intelligenztests.

Schulklasse	gut $R^2$	geführte Klassen	schlecht $R^2$
1	.06		
2	.14		
3	.03		
4			.49
5			.35
6			.47
$\bar{x}$	.07		.43
N	76		78

Intelligenzdetermination ihrer Schülerleistungen im folgenden als ‚gut geführt‘, solche mit deutlich höherem Intelligenzgewicht hingegen als ‚schlecht geführt‘ zu bezeichnen.

Werden die Schülerleistungen in beiden Klassentypen nunmehr – wie schon vorher – durch die drei Prädiktorengruppen ‚Testintelligenz‘, ‚Vorkenntnisse‘ und ‚Schülermotivation‘ in jeweils getrennten Analysen vorherzusagen versucht, ergibt sich folgendes Bild (Tab. 4):

Vorkenntnisse tragen in ‚gut geführten‘ Klassen zur Vorhersage der Schülerleistungen bei; in ‚schlecht geführten‘ Klassen sind Schülermotivations-Variablen von größerer Bedeutung als in den Klassen mit geringer Determiniertheit der Schülerleistungen durch Intelligenzvariablen.

Tab. 4. Determinationskoeffizienten ( $R^2$ ) für die multiple Vorhersage der Schülerleistung bei schweren Mathematikaufgaben durch verschiedene Prädiktorensets.

Prädiktorenset	gut	geführte Klassen	schlecht
Testintelligenz	.016		.240
Vorkenntnisse	.040		.019
Schülermotivation	.165		.23
N =	76		78

Der Versuch, die beiden Klassentypen mit Hilfe von Clusteranalysen auf der Basis der von den Schülern perzipierten Motivierungs- und Instruktionsqualität des Unterrichts zu reproduzieren, erbrachte keine befriedigenden Ergebnisse. Von einer bei Verwendung verschiedener

Klassifikationskriterien konsistenten Gruppierung der Schulklassen kann demnach nicht ausgegangen werden. Unberücksichtigt davon bleibt allerdings die Frage nach der theoretischen Legitimation und dem praktischen Nutzen der von uns gewählten Klassifikationsmöglichkeit.

Im nächsten Schritt sollte überprüft werden, ob die Items des Schülerfragebogens zur Instruktions- und Motivierungsqualität die beiden Klassentypen befriedigend diskriminieren.

Zu diesem Zweck schien das Verfahren der schrittweisen Diskriminanzanalyse deshalb adäquat, weil lediglich die trennschärfsten Items aus den beiden Variablenätzen in die geplante *underachiever*-Regressionsanalyse eingehen sollten (die Verwendung eines möglichst reduzierten Merkmalsatzes erwies sich im Hinblick auf eine tolerable Relation zwischen Stichprobengröße und Variablenanzahl als unumgänglich).

Von den Variablen der Motivierungsqualität erwiesen sich insgesamt drei Merkmale für die Diskriminanzfunktion als relevant, und zwar die Items: „Der Lehrer lobt uns, wenn wir gut mitarbeiten“. „Wenn in der Klasse Unruhe entsteht, dann bestraft der Lehrer die ganze Klasse. (Stimmt nicht)“. „Auch wenn wir mal frech gewesen sind, bleibt der Lehrer ruhig“. (In Klassen mit geringer Intelligenzdetermination der Schulleistungen wurden diese Items überwiegend zustimmend beantwortet).

Die Kennwerte der dabei ermittelten Diskriminanzfunktion sind niedrig (Wilk's  $\Lambda$  .94; Eigenwert .07), und die Kanonische Korrelation als Maß für den Zusammenhang zwischen der Diskriminanzfunktion und der Gruppierungsvariable liegt mit .25 ebenfalls nicht sehr hoch.

Die aufgrund der Diskriminanzfunktion vorgenommene Re-Klassifikation der Fälle entspricht gerade noch den Erwartungen: 61.4 % der Schüler werden der richtigen Gruppe zugeordnet – ein Ergebnis, das auch bei der aufgrund geringer Probandenzahlen notwendig gewordenen *Jackknife*-Kreuzvalidierung absolut stabil blieb.

Demgegenüber sind die Befunde für die Variablen der Instruktionsqualität weniger aufschlußreich. Aufgrund der schrittweisen Diskriminanzanalyse bilden wiederum drei Variablen die Grundlage für die Diskriminanzfunktion: „Wie schwer ist für Dich das Fach Mathematik im Vergleich zu Deinen anderen Unterrichtsfächern? (schwerer/leichter?)“; „Verstehst Du Deinen Mathematiklehrer, wenn er etwas Neues erklärt? (ja/nein)“; „Hilft Dir Dein Lehrer, wenn Du eine Mathematikaufgabe nicht verstanden hast? (ja/nein?)“.

Die Kennwerte der ermittelten Diskriminanzfunktion entsprechen denen der Motivierungsqualität (Wilk's  $\Lambda = .91$ ; Eigenwert =  $.09$ ; Kanonische Korrelation =  $.28$ ). Die Re-Klassifikation der Schüler fiel mit einer Trefferquote von  $58.6\%$  dagegen etwas schlechter aus und verringert sich bei der Jackknife-Prozedur auf  $54\%$ . Entscheidenden Anteil an der niedrigen Erfolgsquote hat die Fehlklassifikationsrate für die Schüler aus Klassen mit niedriger IQ-Determination der Schulleistungen: Aufgrund der Diskriminanzfunktion werden hier über  $70\%$  der Fälle inkorrekt gruppiert.

Insgesamt gesehen, erweisen sich demnach die ausgewählten Merkmale der Motivierungsqualität als ausreichend trennstabil, während über die Variablen der Instruktionsqualität nur Schüler aus Klassen mit hoher IQ-Determination der Schulleistungen gut re-klassifiziert werden konnten.

#### *Schulklassenspezifische Determination von Schulleistungen bei leichten und schweren Mathematikaufgaben für leistungsstarke und -schwache Schüler*

Dieser Analyseschritt sah vor, die bisher bereits berichteten Unterschiede im Vorhersagemuster der Schülerleistungen näher aufzuschlüsseln. Dafür wurden Kommunalitätenanalysen (vgl. Cooley & Lohnes, 1976; Kerlinger & Pedhazur, 1973) jeweils getrennt bei leichten und schweren Mathematikaufgaben für leistungsschwache vs. -starke Schüler gerechnet. Die Gruppierung der Schüler nach der Höhe ihrer Mathematikleistung erfolgte über eine Median-Aufspaltung an der Rohwertverteilung im Gesamtmathematiktest. Die resultierende Vorhersagestruktur gibt Tab. 5 wieder. Danach variiert die durch die berücksichtigten Variablen insgesamt aufklärbare Kriteriumsvarianz der Schülerleistungen bei schweren und leichten Aufgaben erheblich; sie ist bei schlechten Schülern in ‚schlecht geführten‘ Klassen für schwere Testaufgaben am niedrigsten ( $.41$ ), am höchsten hingegen bei schlechten Schülern und leichten Testaufgaben in Klassen mit niedrigem Vorhersagegewicht der Intelligenzleistungen. Im einzelnen (1) nimmt das Erklärungsgewicht der Testintelligenz bei guten Schülern im Vergleich von leichten zu schweren Aufgaben jeweils zu, bei schwächeren Schülern (bedingt durch ihre geringere Leistungsstreuung bei schweren Testaufgaben) hingegen eher ab; (2) nimmt die Bedeutung von Vorkenntnissen bei leichten gegenüber schweren Aufgaben nur bei schwächeren Schülern in ‚gut geführten‘ Klassen zu; (3) tragen Schülermotivations-Variablen zur Vorhersagbarkeit der Mathe-

Tab. 5. Ergebnisse der Kommunalitätenanalysen bei leichten (l) und schweren (s) Mathematikaufgaben, aufgeschlüsselt für gute und schlechte Schüler aus ‚gut‘ bzw. ‚schlecht‘ geführten Klassen.

	gut geführte Klassen				schlecht geführte Klassen			
	gute Schüler		schlechte Schüler		gute Schüler		schlechte Schüler	
	l	s	l	s	l	s	l	s
<i>Spezifische Varianz</i>								
Testintelligenz	.18	.32	.20	.16	.12	.15	.14	.12
Vorkenntnisse	.08	.07	.05	.17	.09	.11	.12	.04
Schülermotivation	.02	.07	.07	.13	.02	.08	.22	.11
Instruktionsqualität	.02	.06	.01	.02	.04	.04	.02	.01
Motivierungsqualität	.04	.17	.06	.16	.13	.11	.08	.03
Fähigkeit x Anstrengung	.01	.08	.01	.00	.14	.02	.12	.05
<i>Konfundierte Varianz</i>								
Insgesamt erklärte Kriteriumsvarianz	.24	.18	.33	.02	-.01	.10	-.04	.05
Nicht erklärte Kriteriumsvarianz	.59	.59	.73	.66	.54	.61	.70	.41
Nicht erklärte Kriteriumsvarianz	.41	.41	.27	.34	.46	.39	.30	.59

matikleistung bei schweren Testaufgaben – z.T. deutlich – mehr bei als bei leichten Aufgaben; nur bei schlechten Schülern in Klassen mit hoher Intelligenzdetermination der Schulleistung ergibt sich ein differentes Bild (der Determinationskoeffizienten schrumpft von den leichten zu den schweren Aufgaben von .22 auf .11).

Die in Tab. 5 berichteten Ergebnisse der Kommunalitätenanalyse beruhen auf einer nur linear-additiven Verknüpfung der berücksichtigten Erklärungsvariablen. Nun ist aber gerade für schwere Mathematikaufgaben plausiblerweise anzunehmen, daß ihre Lösungswahrscheinlichkeit vom Zusammenwirken der Anstrengungs- und Fähigkeitsvariablen abhängt (Koppelungsmodell) und nicht nur wahlweise durch eine dieser Variablen zu erklären ist (Kompensationsmodell); vgl. Krug & Rheinberg (in diesem Band).

Die Überprüfung dieser Annahme war im Rahmen der vorliegenden Untersuchung insofern möglich, als auch ein Interaktionswert (Intelligenz X Schülermotivation) berechnet wurde und in die Kommunalitätenanalyse einging. Dieser ließ Aussagen über die Adäquanz des Koppelungsmodells, nicht aber über die des Kompensationsmodells zu, das an dieser Stelle ausgespart werden muß.

Tab. 6. Determinationskoeffizienten ( $R^2$ ) für Fähigkeit x Anstrengung – (Koppelungsmodell).

	gut geführte Klassen		schlecht geführte Klassen		$\bar{x}$
	gute Schüler	schlechte Schüler	gute Schüler	schlechte Schüler	
leichte Aufgaben	.02	.05	.04	.03	.035
schwere Aufgaben	.13	.06	.10	.12	.125
$\bar{x}$	.075	.055	.07	.075	

Tab. 6 gibt für gute und schlechte Schüler aus Klassen mit niedriger und hoher Intelligenzdetermination die Varianzaufklärungsquote dieser Interaktion getrennt für leichte und schwere Aufgaben wieder. Die Gegenüberstellung der Determinationskoeffizienten für beiden Aufgabentypen macht zunächst deutlich, daß das Koppelungsmodell für schwere Aufgaben im allgemeinen angemessener ist. Aus den Befunden läßt sich aber auch erkennen, daß dies bei schlechten Schülern dann nicht zutrifft, wenn ein didaktisch guter Unterricht (= geringe Intelligenzdetermination der Schulleistung) erteilt wird; denn in diesen

Klassen ist eine Differenz zwischen leichten und schweren Aufgaben nicht erkennbar.

### *Vergleich der Klassifikationen ‚erwartungswidriger Schulleistungen‘ auf der Basis verschiedener Prädiktorensätze*

Mit Hilfe dieses Auswertungsschrittes kann geprüft werden, wie sich die üblicherweise vorgenommene *over-* und *underachievement*-Klassifikation verändert, wenn auf der Prädiktorseite zusätzlich zu den Intelligenztestvariablen Indikatoren des Unterrichts aufgenommen werden. Es stellt sich hier insbesondere die Frage, ob sich durch eine solche Maßnahme die Klassifikationsstabilität generell und systematisch in dem Sinne ändert, daß mehr Schülerleistungen als erwartungsgemäß klassifiziert werden.

*Konventionelle Feststellung der erwartungswidrigen Schulleistung.* Zunächst wurde die zur Ermittlung von *over-* und *underachiever* übliche Prozedur benutzt, indem eine Regressionsanalyse zur Vorhersage der Schüler-Mathematikleistung (leichte und schwere Aufgaben zusammengenommen) durch die sechs Intelligenz-Subtests berechnet und für jeden Schüler der Residualwert (beobachtete minus vorhergesagte Testleistung) bestimmt wurde. Damit war gleichzeitig die Identifikation von Schülern mit erwartungswidrig hohen positiven und negativen Residualwerten möglich. Nachdem für die Auswahl der *over-/underachiever* ein statistisches Signifikanzkriterium ( $\alpha = .10$ ) festgelegt worden war, ließen sich von insgesamt 118 Schülern 26 mit erwartungswidrigen Schulleistungen ermitteln, wobei 12 als *under-* und 14 als *overachiever* zu klassifizieren waren. Der Determinationskoeffizient ( $R^2$ ) für die spezifizierete Regressionsgleichung betrug .29.

*Hinzunahme von Prädiktoren der Instruktions- und Motivierungsqualität.* Im nächsten Schritt wurde die Testintelligenz-Regressionsgleichung dann um jene (jeweils drei) Variablen der Instruktions- und Motivierungsqualität ergänzt, die in den Diskriminanzanalysen zur Separierung der beiden Extremgruppen als besonders bedeutsam gesichert worden waren. Dies ergibt folgendes Bild:

Die Vorhersagegenauigkeit ( $R^2$ ) der Schüler-Mathematikleistung verbesserte sich erwartungsgemäß durch Hinzunahme dieser Unterrichtsmerkmale von .29 über .31 auf .34.

Die Klassifikation von *over-* und *underachiever* veränderte sich durch die Hinzunahme dieser Variablen erheblich; nur 13 von insgesamt 26

Schülern blieben als *over-/underachiever* stabil, während die übrigen 13 Schüler ihre Position veränderten.

Die insgesamt verbesserte Vorhersagepräzision führte aber zu keiner durchgängigen Reduktion der individuellen Abstands- bzw. Residualwerte: Für 9 der 26 Schüler ließ sich die Mathematikleistung bei Berücksichtigung der Unterrichtsqualität sogar deutlich schlechter vorhersagen als zuvor.

*Zusätzliche Berücksichtigung (des Typs) der Schulklasse.* Dieser überraschende Befund kann auf Klassenebene insofern weiter differenziert werden, als hier schulklassenspezifische Varianten ins Auge fallen. Während in zwei der fünf untersuchten Klassen durch die Zunahme von Unterrichtsmerkmalen überwiegend Residuumsanstiege zu verzeichnen sind, bietet sich in zwei weiteren Klassen das entgegengesetzte (und eigentlich annahmekonforme) Bild: Die Residuenabnahme zeigt sich in Klassen, die durch eine hohe IQ-Determination der Schulleistung charakterisiert sind, während die stabilen *under-* und *overachiever* in den Klassen mit niedriger Intelligenzdetermination zu finden sind.

### Diskussion

Bei der theoretischen Interpretation der vorliegenden Befunde muß zunächst berücksichtigt werden, daß die analysierten empirischen Daten nicht zum Zwecke einer schrittweisen Bedingungsanalyse erwartungswidriger Schulleistungen erhoben wurden, sondern für die Bearbeitung einer anderen Fragestellung. Die nachträgliche Verwendung (eines kleinen Teiles) des Datensatzes für die vorliegende Studie führt dazu, daß nicht alle theoretischen Konstrukte und Annahmen empirisch angemessen repräsentiert und realisiert werden konnten. Da die Mängel der Untersuchung aber eher zulasten der überprüften Hypothesen gehen, ist es zulässig, die ermittelten Resultate mit gebotener Zurückhaltung und Vorläufigkeit in verallgemeinernder Weise zu interpretieren.

Erstes Ergebnis der Studie ist das extrem geringe Vorhersagegewicht, das auf die Variablengruppen Intelligenz, Vorkenntnisse und Lernmotivation für die Prädiktion von Mathematikleistungen in der untersuchten Stichprobe entfällt, wenn die Schülerdaten ohne Berücksichtigung der individuellen Schulklassenzugehörigkeit verknüpft werden. Zweifellos sind die dabei geschätzten Korrelationswerte deutlich niedriger als die in der Literatur üblicherweise berichteten

(vgl. Bloom, 1976); dennoch liegen sie innerhalb des Bereichs, wie er in früheren empirischen Studien (vgl. Löschenkohl, 1973) abgesteckt wurde. Dabei muß offen bleiben, ob die geringe Determination der Schulleistungsunterschiede durch die berücksichtigten Variablen auf die Tatsache zurückzuführen ist, daß es sich bei der untersuchten Population von Hauptschülern um eine stark ausgelesene und deshalb mehr oder minder merkmalthomogenisierte Personpopulation handelt (vgl. Tent, 1969; Gaedike, 1975); daß die Kriteriumsvariable mit einem inhaltsspezifischen und instruktionssensitiven Mathematiktest erfaßt wurde, dessen Ergebnisse aber bei schulklassenunspezifischer Auswertung ihre Spezifität (und damit Validität) verlieren; daß die Merkmalsvariation zwischen den Schulklassen zufällig besonders groß ist; und/oder daß die Zuverlässigkeit der verwendeten Erhebungsinstrumente in Zweifel gezogen werden muß.

Die in unserem Falle sehr geringe (und in anderen Untersuchungen erkennbar große Variation) klassenunspezifische Determination von Schülerdifferenzen in ihrer Mathematikleistung spricht deshalb zunächst ebenso für die Verwendung der Analyseeinheit Schulklasse wie für vielfach berichtete Differenzen schulklassenspezifischer Leistungsdetermination (Simons, 1969) und die theoretisch wie empirisch ausreichend belegte Variationsbreite von Unterrichtseffekten zwischen Schulklassen (Treiber, 1980).

Während sich Schulklassen als Analyseeinheiten demnach zumindest für die *Beschreibung* von Schulleistungen und Schulleistungsverläufen anbieten, ergibt sich beim Versuch der multivariaten *Erklärung* solcher Schulklassenunterschiede aber als Problem, daß Gleichungskoeffizienten zur Überprüfung komplexer theoretischer Erklärungen wegen der geringen Stichprobengröße von Schulklassen nur unzuverlässig geschätzt werden können (vgl. Cronbach, 1976). Dies belegt auch die Inspektion von Tab. 2, in der die Abweichungen der adjustierten Determinationskoeffizienten von unkorrigierten Koeffizienten dargestellt sind. Eine Überprüfung theoretischer Annahmen durch separate Schulklassenanalysen erscheint damit ebenso fragwürdig wie die direkte Realisierung des Vorschlages von Simons & Möbus (1976), „das Phänomen der erwartungswidrigen Schulleistung . . . auf schulklassenspezifische Bedingungen (zu) relativieren“ (S. 11).

In der vorliegenden Studie versuchten wir, das Dilemma zwischen dem notwendigen Spezifikationsgrad von Schulleistungserklärungen und der durch die Stichprobengröße ihrer Analyseeinheiten (z.B. der Schulklasse) gesetzten Prüfbarkeitsgrenzen dadurch zu lösen, daß wir Typen von Schulklassen als neue Analyseeinheit einführten. Eine solche Strategie hat in der Pädagogischen Psychologie bei der Klassifikation

von Lehrermerkmalen, Unterrichtsmethoden und Klassenklima eine lange, neue Versuche nicht unbedingt ermutigende Tradition. Unsere Aufgabe war es deshalb, ein Klassifikationsmerkmal zu bestimmen, das sich theoretisch plausibel begründen läßt und sich in früheren Untersuchungen als hinreichend reliabel und valide erwiesen hat. Wir entschieden uns, die Schulklasse nach dem Betrag ihrer multiplen Determination ( $R^2$ ) der Schulleistungen bei schweren Mathematikaufgaben durch sechs Intelligenz-Subtests zu typisieren. Begründet wird diese Wahl durch Befunde, nach welchen die Enge des Zusammenhangs zwischen Intelligenz und Schulleistung als ein schulklassentypisches Merkmal für die didaktische Qualität des Unterrichts gelten kann (Simons et al., 1975; Treiber, 1980), dem es in variablem Ausmaß gelingt, Schulleistungen von kognitiven Schülereingangsfähigkeiten zu entkoppeln (vgl. Bloom, 1976). Die erzielten Ergebnisse rechtfertigen nach unserer Auffassung die vorgenommene Klassifikation der Schulklassen in solche mit hoher und niedriger Determiniertheit der Mathematikleistungen durch Intelligenztestwerte. Zwar ließen sich die beiden Klassentypen durch davon unabhängig durchgeführte Clusteranalysen der von den Schülern wahrgenommenen Instruktions- und Motivierungsqualität ihres Unterrichts nicht als homogene Cluster reproduzieren; doch konnten die beiden Typen zumindest durch schülerperzipierte Unterrichtsmerkmale diskriminiert werden. Es ist deshalb zu vermuten, daß in so typisierten Klassen auch differente Lehr-Lern-Bedingungen bestehen, die in einer variablen Vorhersagbarkeit von Schulleistungen durch kognitive Schülerfähigkeiten resultieren.

Der nächste Analyseschritt bestand deshalb konsequent darin, die Suche nach spezifischen Vorhersagemustern für Mathematikleistungen auszuweiten, und zwar für leichte vs. schwere Aufgaben sowie im Vergleich leistungsstärkerer vs. -schwächerer Schüler in didaktisch gut vs. schlecht geführten Schulklassen. In Übereinstimmung mit Befunden von Snow (1976) ließen die jeweils getrennt gerechneten Kommunalitätenanalysen dabei eine spezifische Determiniertheit von Schulleistungen erkennen. So nimmt z.B. das Gewicht der Vorkenntnisse bei schlechten Schülern in ‚gut geführten‘ Klassen bei schweren gegenüber leichten Aufgaben besonders stark zu. Eine ähnliche Relation ergibt sich für die Motivierungsqualität des Unterrichts, deren Bedeutung bei schweren Mathematikaufgaben sowohl für gute wie für schlechte Schüler deutlich steigt. Trotz der Plausibilität solcher und anderer Detailresultate sollte eine theoretische Integration der Untersuchungsbefunde jedoch erst nach ihrer Replikation durch besser kontrollierte Studien erfolgen. Dies gilt insbesondere für die

Überprüfung des Koppelungsmodells für variable Ausprägungsgrade von Fähigkeiten und Motivstärken, das in der vorliegenden Studie eine allenfalls schwache Überprüfung und Bestätigung erfuhr.

Vorläufige Konsequenzen solcher Befunde für die Analyse erwartungswidriger Schulleistungen sind allerdings schon jetzt erkennbar. So erscheint vor allem ihre klassenunspezifische Identifikation äußerst problematisch. Insofern wird die Kritik am traditionellen Konzept und seiner Operationalisierung auch durch die vorliegende Arbeit bestätigt. Andererseits erscheint es aussichtslos, das Auftreten erwartungswidriger Schulleistungen innerhalb der Analyseeinheit Schulklasse aufklären zu wollen. Vielmehr ergibt sich als Folgerung, den variablen Zusammenhang zwischen Schülerdispositionsmerkmalen und Schulleistung in Kenntnis der Schul- und Unterrichtsqualität zumindest für Schulklassentypen zu spezifizieren. Dabei bietet sich dann ein zweistufiges Verfahren an: Zuerst müßte die Identifikation von Schulklassen erfolgen, in denen viele Schüler schlechtere Leistungen erzielen als Lernende mit vergleichbarer Intelligenz in anderen Klassen. Als zweiter Schritt hätte die klassentypspezifische Analyse erwartungswidriger Schulleistungen auf der Schülerebene zu erfolgen. Dabei muß nach unseren Befunden mit dem scheinbar paradoxen Ergebnis gerechnet werden, daß sich Lernende, die stabil als ‚underachievers‘ klassifiziert werden, vor allem in didaktisch gut geführten Schulklassen finden. Diese Schüler leisten offenbar weniger als gleichintelligente Mitschüler, obwohl in ihrer Klasse ein qualifizierter Unterricht stattfindet. Unter diesen Umständen werden sich die Bedingungsanalysen und Interventionsstrategien dann stärker auf die individuellen Schülermerkmale richten können, da Merkmale schulischer Lehr-Lern-Bedingungen an explanativem wie technologischem Wert verlieren. In schlecht geführten Klassen ist hingegen eine stabile Klassifikation erwartungswidriger Leistungen ersichtlich erschwert, weil (defizitäre) Einflußfaktoren des Unterrichts und der Schüler konfundiert sind. Deshalb wird sich der Erklärungs- und Interventionsfokus hier zunächst stärker auf Indikatoren der Schul- und Unterrichtsqualität konzentrieren.

Die Erwartungswidrigkeit individueller Schulleistungen ergibt sich bei diesem Vorgehen demnach erst in Kenntnis relevanter Schülerdispositionsvariablen und Merkmale schulischer Lernökologie. Ob die Ergebnisse der vorliegenden Arbeit dabei eine verbesserte Möglichkeit eröffnen, auf der Basis expliziter Verknüpfungsmodelle interagierender Schulleistungsbedingungen sowie mit multiplen Prädiktoren für spezifizierte Kriterien erwartungswidrige Schulklassen- und Schülerleistungen theoretisch präziser zu analysieren als bisher, muß die

künftige Forschung erweisen. Immerhin spricht einiges dafür, daß es zur propagierten Eliminierung des gesamten Forschungsansatzes (Wahl, 1975; Schlee, 1976) wissenschaftliche Alternativen gibt, die weitere pädagogisch-psychologische Arbeit mit dem Konzept erwartungswidriger Schulleistungen zulässig und aussichtsreich erscheinen lassen. In der vorliegenden Arbeit wurde versucht, eine dieser Alternativen vorzustellen.

## Literatur

- Atkinson, J.W., Motivational determinants of intellectual performance and cumulative achievement. In J.W. Atkinson & J. Raynor (Hg.), *Achievement and performance*. Washington, D.C.: Winston, 1974. S. 389–410.
- Bloom, B.S., *Human characteristics and school learning*. New York: McGraw Hill, 1976.
- Cooley, W.W. & Lohnes, P.R., *Evaluation research in education*. New York: Wiley, 1976.
- Cronbach, L.J., (with assistance of J.E. Deken & N. Webb), *Research on classrooms and schools: Formulation of questions, design and analysis*. Occasional Paper. Stanford Evaluation Consortium. Stanford: School of Education, University of Stanford, 1976.
- Cronbach, L.J. & Snow, R.E., *Aptitudes and instructional methods. A handbook for research on interactions*. New York: Irvington, 1977.
- Gaedike, A.K., Untersuchungen zur Validität des kognitiven Fähigkeits-Tests für 4. bis 13. Klassen (KFT 4–13). Unveröffentlichte Dissertation, Bonn Pädagogische Hochschule Rheinland, 1975.
- Ginsburg, H.P., *Children's arithmetic: The learning process*. New York: Van Nostrand, 1977.
- Heller, K., Gaedike, A.K. & Weigl, H., *Kognitiver Fähigkeitstest (KFT 4–13)*. Weinheim: Beltz, 1976.
- Horn, W., *Prüfsystem für Schul- und Bildungsberatung*. Göttingen: Hogrefe, 1969.
- Hylla, E. & Kraak, B., *Aufgaben zum Nachdenken (AzN 4+)*. Weinheim: Beltz, 1970.
- Keeves, J.P., *Educational environment and student achievement*. Stockholm: Almquist & Wiksell, 1972.
- Kerlinger, F.N. & Pedhazur, E.J., *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston, 1973.
- Klauer, K.J., Methodenstudie zur Überprüfung von Lehrstoffhierarchien. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 1976, 8, 252–262.
- Lave, J., Tailor-made experiments and evaluating the intellectual consequences of apprenticeship training. *Quarterly Newsletter of the Institute for Comparative Human Development*, 1977, 1, 1–3.
- Löschekohl, E., Gibt es einen allgemein faßbaren Zusammenhang zwischen Schulleistung und Intelligenz? *Psychologie in Erziehung und Unterricht*, 1973, 20, 145–155.
- Neisser, U., General, academic, and artificial intelligence. In L.B. Resnick (Hg.), *The nature of intelligence*. Hillsdale, N.J.: Erlbaum, 1976, S. 135–147.
- Schlee, J., *Legasthenieforschung am Ende?* München: Urban & Schwarzenberg, 1976.

- Simons, H., Untersuchungen zur Generalität und Spezifität erwartungswidriger Schulleistungen. In M. Irle (Hg.), *Bericht über den 26. Kongreß der Deutschen Gesellschaft für Psychologie*, Tübingen 1968. Göttingen: Hogrefe, 1969. S. 446–454.
- Simons, H. & Möbus, C., Zur Fairness von Intelligenztests. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 1976, 8, 1–12.
- Simons, H., Weinert, F.E. & Ahrens, H.J., Untersuchungen zur differentialpsychologischen Analyse von Rechenleistungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 1975, 7, 153–169.
- Snow, R.E., Learning and individual differences. *Review of Research in Education*, 1976, 4, 50–105.
- Tent, L., *Auslese für weiterführende Schulen. Möglichkeiten und Grenzen*. Göttingen: Hogrefe, 1969.
- Thorndike, R., *The concepts of over- and underachievement*. New York: Columbia University, Teachers College Press, 1963.
- Treiber, B., *Qualifizierung und Chancenausgleich in Schulklassen*. Bd. 1 u. 2. Frankfurt: Lang, 1980.
- Treiber, B., Bedingungen individuellen Unterrichtserfolges. Projekt-Abschlussbericht an die Deutsche Forschungsgesellschaft. Heidelberg: Psychologisches Institut (In Vorb.).
- Wahl, D., *Erwartungswidrige Schulleistungen*. Weinheim: Beltz, 1975.
- Yule, W., Rutter, M., Berger, M. & Thompson, J., Over- and underachievement in reading: Distribution in the general population. *British Journal of Educational Psychology*, 1974, 44, 1–12.