

**Zeitschrift:** Bildungsforschung und Bildungspraxis : schweizerische Zeitschrift für Erziehungswissenschaft = Éducation et recherche : revue suisse des sciences de l'éducation = Educazione e ricerca : rivista svizzera di scienze dell'educazione

**Herausgeber:** Schweizerische Gesellschaft für Bildungsforschung

**Band:** 15 (1993)

**Heft:** 1

**Artikel:** Schülerselektion und Schulleistungsprognose auf dem Prüfstand

**Autor:** Perrig-Chiello, Pasqualina / Perrig, Walter J. / Gutzwiller, Peter

**DOI:** <https://doi.org/10.5169/seals-786329>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 17.04.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# Schülerselektion und Schulleistungsprognose auf dem Prüfstand

## Analyse der prognostischen Bedeutung des Basler Übertrittsverfahrens

*Pasqualina Perrig-Chiello, Walter J. Perrig und Peter Gutzwiller*

*Auch wenn sich Ausbildungssysteme verändern, bleibt die Selektionspraxis ein Diskussionsthema und ist für die Bildungsforschung ein ständiger Auftrag. Angesichts des bevorstehenden Schulsystemwechsels in Basel-Stadt will die vorliegende Arbeit das bestehende Übertrittsverfahren (Leistungstest und Schulnoten) von der Primarschule in die weiterführenden Schulen (Progymnasium bzw. Realschule) evaluieren und eine Bilanz ziehen. Die Ergebnisse:*

- Während die prognostische Validität der Schulnoten für die Leistungsvorhersage im Progymnasium hinreichend ist, versagt sie als Prädiktor für die Leistungen in der Realschule.*
- Die prognostische Validität des Übertrittstests liefert bei beiden Schultypen substantielle Erklärungsbeiträge. Seine Bedeutung ist aber für die Vorhersage der Leistungen in der Realschule grösser.*

*Zur Optimierung der bisherigen Selektionspraxis wird eine Kombination von fachspezifischer Notenberücksichtigung (summative Leistungsbeurteilung) und kriteriumsorientierter Leistungsmessung (im Sinne einer formativen Leistungsbeurteilung) vorgeschlagen. Die Notwendigkeit eines vernetzten Denkens in der Schülerleistungsbeurteilung wird unterstrichen.*

### 1. Einleitung

Notengebung, Schüler- und Leistungsbeurteilung sind Themen, die Emotionen wecken. Viele Leute sind gesättigt von Diskussionen über Sinn und Unsinn von Noten, Lehrerurteilen und Leistungsmessung. Andere meinen, dass das, was hinter Noten und Leistungsbeurteilungen steckt und die Folgen, die daraus

erwachsen, noch viel zu wenig reflektiert werde. In den Forschungsarbeiten der pädagogischen Psychologie finden sich im wesentlichen folgende zwei Perspektiven, aus denen Noten und allgemein Leistungsbeurteilung betrachtet werden können. Der eine Aspekt betont die *Determinanten der Schüler- und der Schülerleistungsbeurteilung*. Es geht dabei um die Frage, was Noten eigentlich ausdrücken und wie der psychologische Prozess der Leistungsbeurteilung, der dem Endprodukt Note vorausgeht, aussieht (Perrig-Chiello, 1981, 1983; Pierrehumbert, 1992).

Der zweite Aspekt widmet sich der Frage, welche *Funktionen die Schülernote* explizit zu erfüllen hat, oder welche impliziten Funktionen ihr dann zukommen. Dass sich Noten für eine Vielzahl von impliziten Funktionen geradezu anbieten (Berechtigungs-, Selektions-, Feedback-, Motivationsfunktion u.a.), mag einen Hauptgrund darin haben, dass der Gegenstand der Beurteilung nicht definiert, bzw. nicht definierbar ist (z. B. Ist der Gegenstand die effektive oder die potentielle Leistung?) (Flammer, Perrig und Rüegg, 1983). Aber auch bei den expliziten Funktionen scheiden sich die Geister. Exemplarisch sei hier auf die Auseinandersetzungen um die formative versus normative und zertifikative Schülerbeurteilung verwiesen (Perrenoud, 1991; SIPRI, 1983, 1985). Hat formative Schülerbeurteilung die Beobachtung, Förderung und Lernprozessessteuerung zum Ziel, so strebt die normative Beurteilung ein hierarchisches Klasselement der Lernenden an und verlangt, dass sich alle unter gleichen Bedingungen den gleichen Prüfungen unterziehen. Die zertifikative Beurteilung schlussendlich konzentriert sich am Ende eines Schuljahres oder eines Studienganges auf jene Kriterien, die mit den Anforderungen für ein Zertifikat (eine Berechtigung) in Beziehung stehen. Hinter diesen expliziten Funktionsansprüchen stehen (implizite) Weltanschauungen, die je nach Notenabnehmer (Kinder, Eltern) und Beurteilungsinstanz (Lehrpersonen, Schulbehörde) völlig im Widerspruch zueinander stehen können – was die ganze Beurteilungspraxis verkompliziert und immer wieder für bildungspolitischen Zündstoff sorgt. Auch wenn sich Ausbildungssysteme verändern, die Selektionspraxis wird ein Diskussionsthema bleiben. Selektionspraktiken sind für die jeweils Beteiligten (Kinder, Lehrpersonen, Eltern, Bildungspolitiker) ein aktuelles, heikles Thema, für die Bildungsforschung hingegen ein ständiger Auftrag, dieselben periodisch zu evaluieren und nötigenfalls zu adaptieren.

## 2. Selektionspraktiken und ihre wissenschaftliche Begründung

Die Funktion der schulischen Selektionspraktiken leitet sich aus den Bedürfnissen der Gesellschaft ab. Mit dem Zulassungsentscheid zu einer bestimmten höheren Schule wird gleichzeitig auch der Rahmen zukünftiger beruflicher Qualifikation und somit sozialer Position des Einzelnen abgesteckt. Angesichts dieser wichtigen, ja schicksalhaften Funktion von Selektionspraktiken erstaunt es nicht, dass hohe Anforderungen an sie gestellt werden. Wer mit Selektion zu tun hat, wünscht sich ein zuverlässiges, gerechtes, ökonomisches,

möglichst einfaches Instrument. In erster Linie haben die Beurteilten selbst grösstes Interesse an einer gerechten Beurteilung und natürlich auch Anrecht auf ein entsprechendes Beurteilungsinstrument. Die Bildungswissenschaft versucht, diesen Ansprüchen gerecht zu werden, indem geprüft wird, ob ein Selektionsinstrument die wissenschaftlichen Kriterien der Objektivität, der Zuverlässigkeit und der Gültigkeit erfüllt oder nicht.

In der pädagogisch-psychologischen Fachliteratur finden wir eine Flut von empirischen Studien, die der Frage nachgehen, inwiefern diesen wissenschaftlichen Gütekriterien in den einzelnen Leistungs- und Selektionsprozeduren auch wirklich Genüge getan wird. Durch die Darstellung der wichtigsten Beurteilungskriterien und eine Auswahl von Forschungsbefunden soll hier die Sensibilität im Hinblick auf die Problematik der verwendeten Selektionsinstrumente erhöht werden.

### *2.1. Die Objektivität der Schulnoten und der Leistungstests*

Dass Noten der Forderung nach Objektivität nur ungenügend nachkommen, mag nicht erstaunen. Die vielzitierte Studie von Ulshöfer (1948/49), die zeigt, dass ein und derselbe Abituraufsatz von verschiedenen Lehrpersonen mit allen Stufen der Notenskala (also von 1 bis 6) beurteilt wurde, war Ausgangspunkt vieler polemischer Kontroversen. Entgegen gängiger Erwartungen stellt sich aber das Problem der Objektivität nicht bloss innerhalb der sprachlich-historischen und musischen Fächer, sondern auch bei vermeintlich exakteren Fächern wie Mathematik (vgl. hierzu Ingenkamp, 1976). Dagegen ist die Durchführungs-, die Auswertungs- und die Interpretationsobjektivität der Leistungstests aufgrund der festgelegten Standardangaben bei den Aufgaben, bei der Auswertung und bei der Interpretation hoch.

### *2.2. Die Reliabilität der Noten und der Leistungstests*

Die Zuverlässigkeit ist bei den Noten in der Regel enttäuschend (vgl. Weiss, 1989). Da die Objektivität der Noten in der Regel auch schon gering ist, ist es nicht überraschend, wenn z.B. ein und dieselbe Lehrperson bei wiederholten Beurteilungen zu unterschiedlichen Ergebnissen kommt. Hier muss man sich natürlich fragen, ob eine wissenschaftliche Testsituation genau der Alltagssituation in der Schule entspricht, in der dieselbe Lehrperson mit derselben Klasse über Jahre zusammenarbeitet. Bei den Leistungstests ist die Zuverlässigkeit jedoch in der Regel hoch.

### *2.3. Die Validität der Noten und der Leistungstests*

Zentral ist für uns jedoch die Frage nach der Validität, genauer gesagt, der prognostischen Validität der Leistungsbeurteilung: Schulnoten sollen nicht nur aktuelle schulische Leistungen kennzeichnen, sondern auch Prognosen zukünftiger Leistungen.

tiger Leistungen ermöglichen. Danach geht man von der Annahme aus, dass durch eine gute Methode der Leistungsmessung und der Leistungsbeurteilung nicht nur gute Schüler und Schülerinnen von schwachen unterschieden werden können, sondern dass es auch gelingt, die für einen späteren Schul- oder Leistungserfolg verantwortlichen Grundvoraussetzungen zu messen. Von guten Schülern und Schülerinnen wird erwartet, dass sie auch in den weiterführenden Schulen gut sind, während von schwachen angenommen wird, dass diese wohl eher schwach bleiben. Untersuchungen, die dieser Frage nachgehen, gibt es in grosser Zahl. Die Ergebnisse haben insgesamt die hohen Erwartungen nicht erfüllt; die Vorhersage des Schulerfolgs ist meistens von einem Jahr zum folgenden relativ gut möglich, sie verschlechtert sich aber mit zunehmender Vorhersagedistanz massiv. Um diesen Sachverhalt zu illustrieren, sollen hier einige Resultate exemplarisch dargestellt werden.

### *2.3.1. Korrelationsanalysen zur Bestimmung der prognostischen Validität*

Deutsche Untersuchungen zeigten, dass die Korrelation zwischen den vor dem Übergang auf das Gymnasium erteilten Primarschulnoten und dem Gymnasialerfolg nach 5 Jahren in der Grössenordnung von  $r = 0.30$  liegen. Das gleiche gilt für die Korrelation zwischen Aufnahmeprüfung und Gymnasialerfolg (Schultze, 1964). Ähnliche Ergebnisse finden sich auch für die Schweiz bei Meili et al. (1977). Demgegenüber korrelieren Intelligenztestwerte mit dem späteren Schulerfolg auf dem Gymnasium zwischen  $r = 0.4$  und  $r = 0.8$ . Die durchwegs positiven Korrelationen zwischen Schulleistungs- und Intelligenzmassen belegen nach Ingenkamp, dass ein wesentlicher Faktor der Schulleistung die testmässig erfasste Intelligenz ist (Ingenkamp, 1985, 190), deren standardisierte Erhebung eine generell objektivere, zuverlässigere und gültigere Prognose liefert (Heller, 1984).

Wir wollen nun einige Erfahrungswerte von evaluierten Übertrittsverfahren anschauen:

- Das Solothurner Übertrittsverfahren wurde von Reiber (1973) evaluiert: Die Übertrittsprüfung korrelierte mit  $r = .28$  mit der Leistung im Gymnasium und das Lehrerurteil mit  $0.32$ .
- Eine Zusammenfassung verschiedener Schweizer Untersuchungen bezüglich Prädiktion durch Schulnoten und Aufnahmeprüfungen fasste Trier (1974) zusammen: Es fanden sich hier Korrelationen von  $r = 0.3$  bis  $r = 0.4$ , bzw.  $r = 0.2$  bis  $r = 0.3$ .
- Ingenkamp (1976) fasste ebenfalls verschiedene Untersuchungen in der BRD zusammen: Die Korrelationswerte für die Prädiktoren Deutsch und Rechnen schwanken zwischen  $r = 0.19$  und  $r = 0.54$ .
- Erstaunlich hoch fällt die prognostische Validität beim Zuger Übertrittsverfahren aus (Rohrer et al, 1981): Prädiktoren waren das Zeugnis, das Lehrerurteil, je zwei Leistungsprüfungen in Deutsch und Rechnen, zusätzlich noch ein Schuleignungstest. Das Kriterium war der Notendurchschnitt (Deutsch, Rechnen, Biologie und Algebra/Geometrie) in der zweiten Klasse der weiterführenden Schule. Bei den Schülern und Schülerinnen des Gymnasiums

und der Sekundarschule (die leistungsstarken Gruppen also) fiel die grösste Prädiktionskraft auf das Zeugnis ( $r = 0.47$ ) und auf die beiden Rechnungsprüfungen (beide bei  $r = 0.46$ ). Lehrerurteil, Deutschprüfungen und Schulleistungstest bewirkten nur eine geringfügige Verbesserung der Prädiktion. Bei den Realschülern und -schülerinnen (bei der leistungsschwachen Gruppe also) hingegen waren die Prädiktionswerte nicht mehr so eindeutig. Zeugnis und Prüfung zusammen sagen bei Sekundarschülern und -schülerinnen sowie bei Gymnasiasten und Gymnasiastinnen ca. 35% der Kriteriumsvarianz voraus, Lehrerurteil und Prüfung bei den Realschülern ca. 24% bzw. 28%. Somit bleiben 65% der Streuung des Schulerfolgs unerklärt und müssen auf andere Faktoren zurückgeführt werden.

- Zu vergleichbar hohen Validitätskoeffizienten kam man übrigens auch bei der Evaluation des Luzerner Übertrittsverfahrens (Casparis, 1984).

### 2.3.2. Kritische Bemerkungen zur Forderung hoher Reliabilitäts- und Validitätswerte

Nach der Analyse dieser Studien liegt die Schlussfolgerung nahe, dass im Grunde alle Prozeduren, rein statistisch gesehen, die späteren Erfolge nur in einem bescheidenen Masse vorhersagen können. Aber gerade an diesem Punkt möchten wir einhaken und auf ein grundlegendes Problem aufmerksam machen: Prinzipiell ist fraglich, ob eine bessere prognostische Validität eines Übertrittsverfahrens überhaupt möglich ist: Eine hohe prognostische Validität beinhaltet ja die Konstanz des Merkmals Schulleistung. Das würde z. B. bedeuten, dass eine Leistungsverbesserung (durch intensive Förderung zum Beispiel) eines schwachen Schülers zu einem Absinken der Vorhersagevalidität führt. Nun gibt es aber natürlich noch andere Möglichkeiten, die die Leistungskonstellationen zwischen Schüler und Schülerinnen verändern. Bei jüngeren Kindern könnten unterschiedliche Verläufe von Reifungsprozessen zu diesen Veränderungen führen, bei älteren hingegen können Kontexteinflüsse und eine spezifische Persönlichkeitsentwicklung die Leistungen verändern. So können die Lernfortschritte verschiedener Schüler und Schülerinnen unterschiedlich verlaufen.

Unter diesen Voraussetzungen stimmen Ergebnisse von Studien nachdenklich, wonach viele Lehrpersonen schon früh die Kinder «kategorisieren». In einer Studie von Steinkamp (1971) «wussten» 64 von 69 Klassenlehrern und Lehrerinnen schon im zweiten Schuljahr, welche ihrer SchülerInnen für das Gymnasium geeignet waren! Dass solche Lehrkräfte, die diese «klaren und sicheren» Prädiktionen machen, dann oft auch durch die späteren Ergebnisse bestätigt werden, ist aufgrund der Mechanismen der selbsterfüllenden Prophezeiungen (Pygmalioneffekte) nicht verwunderlich.

### 3. Analyse der Basler Übertrittsverfahren im alten Schulsystem

Anlass zur vorliegenden Studie gab dem Autorenteam die Annahme eines neuen Schulgesetzes durch den Basler Soverän 1988, das 1994 rechtskräftig sein wird. Der jetzige Zeitpunkt schien uns opportun, das seit 1972 praktizierte Übertrittsverfahren von der Primarschule in die weiterführenden Schulen (Gymnasium, Real- und Sekundarschule) zu evaluieren, eine Bilanz zu ziehen und aus den gemachten Erfahrungen mögliche Konsequenzen zu ziehen.

Die Primarschule dauert in Basel-Stadt vier Jahre. Erstmals kommen Basler Kinder (und ihre Eltern!) in der 3. Primarklasse mit Zeugnissen in Berührung. Das entscheidende Zeugnis für den Übertritt ist das erste der 4. Primarklasse, wobei der Durchschnitt der beiden erteilten Noten, Sprache und Rechnen, zählt. Die Notenlimiten für den Eintritt ins Progymnasium betragen zusammen 10 Punkte, für die Realschule 9 Punkte. Alle darunterliegenden Punktwerte erlangen die Qualifikation für die Sekundarstufe. Man beachte also, dass in Basel-Stadt die Realschule die zweithöchste Schulform darstellt, dies im Gegensatz zu vielen andern Kantonen. Gleichzeitig geben die Lehrer und Lehrerinnen der 4. Klasse eine Empfehlung für den Übertritt in eine bestimmte Abnehmerschule ab. Zusätzlich führt der Schulpsychologische Dienst BS mit sämtlichen Schülern und Schülerinnen der 4. Primarklasse einen einheitlichen Leistungstest durch. Somit stehen zwei Übertrittskriterien zur Verfügung:

1. Der Notendurchschnitt des Zeugnisses im 1. Semester der 4. Primarklasse (Sprache und Rechnen) und
2. das Ergebnis des Leistungstests.

Für diejenigen Schüler und Schülerinnen, deren Zeugnisnoten nicht den Minimalanforderungen der gewünschten Schule entsprechen, gilt der Test als Aufnahmeprüfung. Genügend gute Testleistungen gestatten also den Eintritt in Realschule bzw. Progymnasium auch beim Fehlen entsprechender Schulnoten. Es entspricht in Basel einem politischen Entscheid, dass 50% aller Kinder eines Jahrganges aufgrund ihres Testergebnisses für die Sekundarstufe selektioniert werden sollen, 30% für die Realstufe und 20% fürs Progymnasium.

Das Basler Übertrittsverfahren in der gegenwärtig noch geltenden «alten Schulordnung» ist ein Selektionsverfahren, das zu einem sehr frühen Zeitpunkt (die Kinder sind 10 Jahre alt) Schülerinnen und Schüler einem hierarchisch gegliederten Schulsystem zuweist, dessen Durchlässigkeit zwar theoretisch gegeben, praktisch aber wenig genutzt wird (z. B. aufgrund von Informationsmangel und/oder Unmotiviertheit der Eltern, mangelnder Laufbahnberatung der Kinder, usw). Aufgrund der Vorbemerkungen über mögliche Funktionen der Schulnoten und der Leistungsbeurteilungen erfüllt in diesem Übergangsverfahren die Note eine Berechtigungsfunktion (ein vorher festgelegtes Minimum von Punkten muss erreicht werden) und der Test erfüllt eine reine Selektionsfunktion (z. B. nur die besten 20% dürfen ins Progymnasium). Hier wäre ein erster Diskussionspunkt über die Rechtfertigung dieses Vorgehens angemessen. Denn gerade die Leistungstests, die zum Teil Annäherungen oder Teilkomponenten validierter psychologischer Testverfahren darstellen, erhe-

ben den Anspruch, kognitive Basalfunktionen, die für den Schulerfolg verantwortlich sind, objektiv und zuverlässig zu messen. Aus diesem Grund ist es besonders interessant zu prüfen, ob nun der Test oder die Noten der 4. Klasse den Schulerfolg in der Abnehmerklasse (in dieser ersten Untersuchung die ersten Klassen der Realschulen und Progymnasien) besser vorhersagen. Dabei gilt es zu bemerken, dass Leistungsprognose und Selektionsrechtfertigung eines Beurteilungsinstrumentes zwei unterschiedliche Sachverhalte darstellen. Wenn wir z.B. mit geringen Korrelationen zwischen den Übertrittsnoten und den Leistungen der Abnehmerklasse massive Leistungsverschiebungen innerhalb der untersuchten Gruppe finden, kann daraus noch nicht gefolgert werden, dass einige Kinder nicht in diese Schule gehören. Um solche Fragen zu beantworten, müssten die Kinder beider Schultypen mit den gleichen Aufgaben untersucht werden; eine Idee, deren Realisierung ernsthaft zu prüfen wäre, in dieser Studie aber nicht zu verwirklichen war. Andererseits kann man sagen, dass ein Selektionsverfahren mit geringer prognostischer Validität seinen Einsatz ohnehin nicht rechtfertigen kann.

### 3.1. Methode

#### *Die Stichprobe*

Bei unserer Stichprobe handelt es sich um den Übertrittsjahrgang 1985. Dieser Jahrgang wurde nach dem Zufallsprinzip gewählt. Insgesamt wurden die Noten und Testleistungen von 1315 Kindern in die Analyse einbezogen. In bezug auf die Prädiktionskraft der Schulnoten und des Leistungstests des 4. Schuljahrs auf den Lernerfolg im nächsten Schuljahr (1986) wurden die Leistungen der Kinder in der Realschule,  $n = 303$  (23%), und der Kinder im Progymnasium untersucht,  $n = 435$  (33%). Die Kinder der Sekundarschule, 577 (34%) wurden in diese Analyse nicht einbezogen, da sie eine sehr heterogene Gruppe bilden, die durch Negativ-Selektion in diesen Schultypus gelangt (alle, die die beiden oberen Schultypen nicht schaffen, haben keine andere Wahl als die Sekundarschule).

#### *Analysemethoden*

Zur Überprüfung des Selektionsverfahrens verwendeten wir Korrelations- und Regressionsanalysen.

### 3.2. Ergebnisse

#### *Ergebnisse der Korrelationsanalyse*

In diesem ersten Schritt wurden folgende Variablen korreliert:

- Noten der 4. Primarklasse in Rechnen, Deutsch und die Notensumme und
- Übertrittstest (Verbal-, Rechnenteil und total) mit der



- Notensumme des 1. bzw. 2. Semesters der Abnehmerklassen in der Realschule und dem Progymnasium.

Die Ergebnisse sind aus Tabelle 1 (Realschule) und Tabelle 2 (Progymnasium) ersichtlich. Aus dem Vergleich der beiden Korrelationstabellen können wir folgende zusammenfassende Ergebnisdarstellung machen:

*Tabelle 1:* Korrelationen der Schulnoten und des Leistungstests der 4. Primarklasse mit den Schulnoten in der 1. Realklasse (n = 303)<sup>1</sup>.

	Notensumme 1. Sem.	Notensumme 2. Sem.	Ü-Test Verb.	Ü-Test Rechn.	Ü-Test Tot.	Notensumme 4. Prim.
Note Deutsch 4. Prim.	<u>.287</u>	<u>.259</u>	<u>.194</u>	-.179	.018	.6215
Note Rechn. 4. Prim.	.091	.148	-.029	<u>.227</u>	.136	.704
Ü-Test Verb.	<u>.372</u>	<u>.279</u>	1.0	.316	.788	.116
Ü-Test Rechn.	.105	.072	.316	1.0	.792	.051
Ü-Test Total	<u>.328</u>	<u>.242</u>	.788	.792	1.0	.120
Notensumme 4. Prim.	<u>.277</u>	<u>.302</u>	.116	.052	<u>.120</u>	1.0

<sup>1</sup> Die unterstrichenen Korrelationen haben für die Interpretation eine besondere Bedeutung.

*Tabelle 2:* Korrelationen der Schulnoten und des Leistungstests der 4. Primarklasse mit den Schulnoten in der 1. Progymnasialklasse (n = 435)<sup>1</sup>.

	Notensumme 1. Sem. PG	Notensumme 2. Sem. PG	Ü-Test Verb.	Ü-Test Rechn.	Ü-Test Total	Notensumme 4. Prim.
Noten Deutsch 4. Prim.	<u>.494</u>	<u>.35</u>	<u>.307</u>	.039	.244	.808
Noten Rechn. 4. Prim.	<u>.318</u>	.217	.150	<u>.244</u>	.237	.771
Ü-Test Verb.	<u>.376</u>	.220	1.0	.514	.885	.293
Ü-Test Rechn.	.253	.139	.514	1.0	.815	.174
Ü-Test Total	<u>.409</u>	<u>.247</u>	.885	.815	1.0	.305
Notensumme 4. Prim.	<u>.518</u>	<u>.362</u>	.293	.174	<u>.305</u>	1.0

<sup>1</sup> Die unterstrichenen Korrelationen haben für die Interpretation eine besondere Bedeutung.

1. Der Zusammenhang zwischen den Noten bzw. dem Übertrittstest der 4. Primarklasse mit den Leistungen in der weiterführenden Schule variiert je nach gewähltem Schultypus: Für das Progymnasium (1. Sem.) fallen die Korrelationen ausgeprägter aus als für die Realschule. So können wir fürs Progymnasium akzeptabel hohe positive Korrelationen zwischen den Variablen: Noten Deutsch 4. Kl., Noten Rechnen 4. Kl., Notensumme 4. Kl., Übertrittstest Verbalteil, Übertrittstest Total und der Notensumme in der 1. Gymnasialklasse feststellen; für die Realschule jedoch nur beim Übertrittstest (Verbal und Total).
2. Für das 2. Semester der weiterführenden Schule wird die Prädiktionskraft der Selektionsinstrumente abgeschwächt.

Die bisherigen Analysen bescheinigen dem Basler Selektionsverfahren und seinen einzelnen Selektionskriterien eine prognostische Validität, die man auch bei der Evaluation anderer Verfahren im In- und Ausland findet. Auf den ersten Blick ist das Verfahren in bezug auf die Erfolgsvorhersagen also nicht besser, aber auch nicht schlechter als die Selektionsprozeduren der eingangs zitierten Schulsysteme. Dies betrifft vor allem die Vorhersagekraft der Schulnoten. Der Prädiktionwert des Tests ist eher gering, wenn wir vom Ziel ausgehen, mit einem psychologischen Leistungstest inhaltsübergreifende kognitive Leistungsfunktionen zu messen, die dem Schulerfolg zugrunde liegen. Wenn dieser Test so etwas wie Intelligenzfunktionen messen sollte, würden wir erfahrungsgemäss Korrelationen in der Grössenordnung von .50 und mehr erwarten. Der Befund, wonach die Leistungen der Schüler und Schülerinnen in der leistungsschwächeren Realschule weniger gut vorausgesagt werden können als in den leistungsstärkeren Schulen, bestätigt sich auch in unserer Analyse.

Beachtlich ist der starke Verlust an Prädiktionskraft vom ersten zum zweiten Semester. Wenn man annimmt, dass sich diese Tendenz bei späteren Messungen fortsetzt, wäre die prognostische Validität der Selektionsverfahren bald auf null gesetzt. Erfahrungsgemäss trifft dies nicht zu. Also müssen wir annehmen, dass beim Schulbeginn in der weiterführenden Schule etwas Besonderes passiert. Wir vermuten, dass die erforderliche Neuorientierung von den Kindern unterschiedlich bewältigt wird, was die in der Normalschule relativ stabilen Leistungskonstellationen der Schüler und Schülerinnen relativ rasch durcheinander bringt. Man könnte dies so interpretieren, dass die neue Ausbildungskonstellation eine neue Chancenverteilung impliziert, die von den Kindern auch tatsächlich unterschiedlich genutzt wird. Um diese Hypothese der besonderen Bedeutung des «neuen» Schulanfangs weiter zu verfolgen, müssten wir unsere Analyse auf spätere Leistungen ausdehnen.

Erklärungsbedürftig ist auch die geringe prognostische Validität der Selektionskriterien bei den schwächeren Schülern und Schülerinnen der Realschule. Dies kann auch wiederum verschiedene Gründe haben. Auch hier kann natürlich die Neuorientierung zu neuen Leistungsrangordnungen führen (z. B. aufgrund anderer Skalenhandhabung). Auffallend ist, dass die Leistungsergebnisse der Rechenkomponenten ihre Prädiktionskraft vollständig verlieren. Dies könnte bedeuten, dass die spezifischen Leistungsanforderungen bei schwächeren und besseren Schülern und Schülerinnen in der Realschule sich nicht in

differentiellen Ergebnissen niederschlagen. Bei der Beurteilung der berechneten Korrelationen gilt es auch noch zu bemerken, dass durch die Zuweisung in die beiden Schultypen (Gymnasium und Realschule) Homogenisierungen vorgenommen werden, die zu einer Reduktion des gemessenen Zusammenhangs führen können. Dies gilt aber natürlich für alle Studien, die die prognostische Validität der Benotung untersuchen.

### *Ergebnisse der Regressionsanalyse*

Unser Hauptanliegen ist hier die tiefere Klärung der Frage, wie gross der Vorhersagewert der im Basler Übertrittsverfahren verwendeten *Prädiktoren* (Noten der 4. Klasse, Übertrittstest) auf das *Kriterium Schulerfolg* in der weiterführenden Schule (hier operationalisiert als Noten im 1. bzw. 2. Semester des Progymnasiums bzw. der Realschule) ist. Hierfür verwendeten wir die Methode der multiplen schrittweisen Regression.

### *Die Ergebnisse in der Realschule*

In Tabelle 3 sind die Ergebnisse der Regressionsanalyse der Prädiktoren Rechnen, Deutsch, Test-Verbalteil und Test-Rechenteil und der vier Kriterien, Deutsch, Rechnen, Fremdsprache und Notensumme in der Realschule dargestellt.

*Tabelle 3:* Ergebnisse der Regressionsanalysen in der Realschule. Erstes Sem. (2. Sem. in Klammer), Prädiktoren sind die Einzelleistungen in der 4. Klasse<sup>1</sup>.

	<i>Multiples R</i>	<i>R<sup>2</sup></i>
1) Kriterium: Deutsch		
Prädiktoren: 1. Ü-Test Verbalteil	.41 (.412)	.18 (.163)
2. Note Deutsch	.49 (.479)	.24 (.229)
2) Kriterium: Fremdsprache		
Prädiktoren: 1. Note Deutsch	.196 (.19)	.038 (.036)
2. Note Rechnen	.23 (.252)	.053 (.063)
3) Kriterium: Rechnen		
Prädiktoren: 1. Ü-Test Rechenteil	.261	.068
2. Note Rechnen	.333 (.252)	.111 (.063)
3. Ü-Test Verbalteil	.389 (.299)	.151 (.089)
4) Kriterium: Notensumme		
Prädiktoren: 1. Ü-Test Verbalteil	.386 (.279)	.149 (.078)
2. Note Deutsch	.456 (.348)	.208 (.121)
3. Note Rechnen	.473 (.392)	.224 (.154)

<sup>1</sup> Die vollständige Prädiktorenliste war: 1) Note Rechnen, 2) Note Deutsch, 3) Ü-Test Rechnen, 4) Ü-Test Verbal. In dieser Tabelle werden nur die signifikanten Prädiktoren aufgelistet. Ein fehlender Wert in der Tabelle bedeutet, dass hier keine statistisch signifikante Vorhersage möglich war.

Auffallend ist, dass die Selektionskriterien der vierten Klasse mit insgesamt 5% erklärter Varianz praktisch überhaupt keinen prognostischen Wert für die Vorhersage der Leistungen im Fremdsprachenunterricht haben. Zudem fällt auf, dass die anderen Noten am besten durch die Testleistungen prognostiziert werden. Bei Deutsch ist es der Test-Verbalteil, der 18% der insgesamt 24% an erklärter Varianz aufdeckt. Bei der Notensumme ist wiederum der Test-Verbalteil mit 14% der beste Prädiktor und beim Rechnen ist es der Test-Rechenteil, aber mit nur 7% Varianzaufklärung. Die geringe Kraft der Notenprädiktion könnte bedeuten, dass das Lehrpersonal in der 4. Primarklasse bei seiner Leistungsbeurteilung grosses Gewicht auf diskriminative Einstufungen der guten Schüler und Schülerinnen legt, um eine gerechte Auslese für das Progymnasium zu treffen, und dass die Noten bei den schwächeren oder schwächsten Kindern weniger gut diskriminieren. Um diese Interpretation zu überprüfen, würden weitere Analysen nötig sein.

Die nächste Analyse sollte Aufschluss darüber geben, ob die Prädiktionskraft der Noten und der Tests steigt, wenn man die Summenwerte (Notensumme und Testsumme) als Prädiktoren verwendet. Die Ergebnisse dieser Analyse sind in Tabelle 4 dargestellt. Sie zeigen, dass am Gesamtmuster nichts verändert wird. Allerdings sinkt die erklärte Varianz bei der Vorhersage der Deutsch- und der Gesamtnote.

Tabelle 4: Ergebnisse der Regressionsanalysen in der Realschule. Erstes Sem. (2. Sem. in Klammer). Prädiktoren: Gesamtnote und Gesamttest<sup>2</sup>.

	<i>Multiples R</i>	<i>R<sup>2</sup></i>
1) Kriterium: Deutsch		
Prädiktoren: 1. Ü-Test Total	.295 (.308)	.087 (.094)
2. Gesamtnote 4. Kl.	.332 (.351)	.110 (.123)
2) Kriterium: Fremdsprache		
Prädiktoren: 1. Gesamtnote 4. Kl.	.217 (.248)	.047 (.061)
2. Ü-Test Total	.248	.061
3) Kriterium: Rechnen		
Prädiktoren: 1. Ü-Test Total	.352	.124
2. Gesamtnote 4. Kl.	.414	.171
(1. Gesamtnote 4. Kl.)	(.266)	(.070)
(2. Ü-Test Total)	(.317)	(.100)
4) Kriterium: Notensumme		
Prädiktoren: 1. Ü-Test Total	.334	.111
2. Gesamtnote 4. Kl.	.419	.175
(1. Gesamtnote 4. Kl.)	(.301)	(.091)
(2. Ü-Test Total)	(.365)	(.133)

<sup>2</sup> In der Tabelle werden nur die signifikanten Prädiktoren aufgelistet. Ein fehlender Wert in der Tabelle bedeutet, dass hier keine statistisch signifikante Vorhersage möglich war.

## Die Ergebnisse im Progymnasium

In Tabelle 5 und 6 sind die Ergebnisse dargestellt. In Tabelle 5 fällt auf, dass die Fremdsprachenleistung wiederum durch kein Selektionskriterium der 4. Klasse befriedigend vorhergesagt wird. Im Vergleich zur Realschule zeigt sich nun überraschend, dass nicht mehr die Tests, sondern die Noten der jeweiligen Fächer die besten Vorhersagen für die Leistungen im Progymnasium machen. So erklärt die Deutschnote in der 4. Primarklasse 30% der Variation in der Deutschnote (2. Sem.) des Progymnasiums und 24% der Variation in der Notensumme, was aufgrund von Erfahrungswerten recht hohe Erklärungsanteile darstellt. Bei der Note im Rechnen ist die Rechnungsnote der beste Prädiktor und klärt 15% Varianz auf. Dieser Befund bestärkt die schon früher geäußerte Vermutung, dass sich das Lehrpersonal für die Beurteilung der stärkeren Schüler und Schülerinnen grosse Mühe gibt, was dann die Noten zu guten Prädiktoren für den späteren Schulerfolg im Progymnasium macht.

Tabelle 5: Ergebnisse der Regressionsanalysen im Progymnasium. Erstes Sem. (2. Sem. in Klammer). Prädiktoren: Einzelnoten und Subtests<sup>3</sup>.

	<i>Multiples R</i>	<i>R<sup>2</sup></i>
1) Kriterium: Deutsch		
Prädiktoren: 1. Note Deutsch	.556 (.433)	.309 (.187)
2. Ü-Test Verbalteil	.610 (.486)	.373 (.236)
2) Kriterium: Fremdsprache		
Prädiktoren: 1. Note Deutsch	.309 (.308)	.095 (.095)
2. Note Rechnen	.339 (.327)	.115 (.107)
3) Kriterium: Rechnen		
Prädiktoren: 1. Note Rechnen	.384 (.342)	.147 (.117)
2. Ü-Test Rechenteil	.468 (.396)	.219 (.156)
3. Note Deutsch	.487 (.428)	.237 (.183)
4) Kriterium: Notensumme		
Prädiktoren: 1. Noten Deutsch	.493 (.35)	.243 (.122)
2. Ü-Test Verbalteil	.546	.298
3. Note Rechnen	.576	.332
4. Ü-Test Rechenteil	.583	.340
(2. Note Rechnen)	(.374)	(.14)
(3. Ü-Test Verbalteil)	(.39)	(.152)

<sup>3</sup> Die vollständige Prädiktorenliste war: 1) Note Rechnen, 2) Note Deutsch, 3) Ü-Test Rechnen, 4) Ü-Test Verbal. In dieser Tabelle werden nur die signifikanten Prädiktoren aufgelistet. Ein fehlender Wert in der Tabelle bedeutet, dass hier keine statistisch signifikante Vorhersage möglich war.

Wie bei der Analyse in der Realschule haben wir sodann geprüft, ob die Vorhersagekraft der Selektionsinstrumente steigt, wenn wir die Notensumme und die Testsumme als Prädiktoren verwenden. Diese Ergebnisse sind in Tabelle 6 dargestellt. Auch hier verändert sich das Gesamtmuster nicht. Die Gesamtnote bleibt der bedeutend bessere Prädiktor als der Gesamttestwert für den Schulerfolg im Progymnasium. Die Vorhersage der Deutschleistung wird wesentlich schlechter. Bei den anderen Fächern bleibt die Situation etwa gleich.

*Tabelle 6:* Ergebnisse der Regressionsanalysen im Progymnasium. Erstes Sem. (2. Sem. in Klammer). Prädiktoren: Noten- und Test-Totel <sup>4</sup>.

	<i>Multiples R</i>	<i>R<sup>2</sup></i>
1) Kriterium: Deutsch		
Prädiktoren: 1. Gesamtnote 4. Kl.	.489 (.393)	.239 (.155)
2. Ü-Test Total	.547 (.447)	.299 (.199)
2) Kriterium: Fremdsprache		
Prädiktoren: 1. Gesamtnote 4. Kl.	.332 (.313)	.110 (.098)
3) Kriterium: Rechnen		
Prädiktoren: 1. Gesamtnote 4. Kl.	.337 (.364)	.142 (.133)
2. Ü-Test Total	.467 (.415)	.213 (.172)
4) Kriterium: Notensumme		
Prädiktoren: 1. Gesamtnote 4. Kl.	.518 (.361)	.268 (.13)
2. Ü-Test Total	.581 (.389)	.338 (.151)

<sup>4</sup> In der Tabelle werden nur die signifikanten Prädiktoren aufgelistet. Ein fehlender Wert in der Tabelle bedeutet, dass hier keine statistisch signifikante Vorhersage möglich war.

Dieses Ergebnismuster legt die Vermutung nahe, dass die fachspezifische Prädiktion einer Leistung besser ist als die Vorhersage über summierte Teilergebnisse. Dies ist besonders bei der Vorhersage der Deutschnote der Fall.

### *Die fächerspezifische Vorhersage der Gesamtleistung in Realschule und Progymnasium*

In der letzten durchgeführten Regressionsanalyse haben wir die Frage weiter untersucht, ob eine fächerspezifische Vorhersage besser ist als eine Vorhersage mittels einer Leistungssumme. Dabei haben wir als Kriterium die Notensumme in der Realschule und im Progymnasium gewählt. Als Prädiktoren haben wir die Einzelnoten und die Subtest-Resultate zusammen mit den Notensummen und den Testsummen verwendet. Die Ergebnisse dieser Analyse sind in Tabelle 7 dargestellt.

*Tabelle 7: Ergebnisse der Regressionsanalysen im Progymnasium und in der Realschule. Zweites Semester. Prädiktoren: Einzelnoten, Subtests und Summenwerte. Kriterium: Notensumme <sup>5</sup>.*

*a) Realschule*

Prädiktoren	<i>Multiples R</i>	R <sup>2</sup>
1. Übertrittstest/Verbalteil	.28	.08
2. Note 4. Klasse Deutsch	.35	.12
3. Note 4. Klasse Rechnen	.39	.15

*b) Progymnasium*

Prädiktoren	<i>Multiples R</i>	R <sup>2</sup>
1. Note 4. Klasse Deutsch	.35	.12
2. Übertrittstest Total	.38	.15
3. Note 4. Klasse Rechnen	.40	.16

<sup>5</sup> In der Tabelle werden nur die signifikanten Prädiktoren aufgelistet.

Hier zeigt sich, dass nicht nur die Leistungen in den Einzelfächern, sondern die Gesamtleistung in den weiterführenden Schulen durch die Einzelnoten in den Fächern der Klasse vor dem Übertritt besser vorhergesagt werden als durch die Gesamtleistung. Bei den Realschülerinnen und -schülern ist der beste Prädiktor die Leistung im Verbalteil des Übertrittstests, im Progymnasium ist es die Deutschnote in der 4. Klasse.

## 5. Schlussfolgerungen

Wir wollen die Hauptergebnisse, mögliche Schlussfolgerungen und Konsequenzen, die sich aufgrund der durchgeführten Analysen ergeben, zusammenfassend festhalten:

1. Während die *prognostische Validität der Schulnoten* für die Erfolgsvorhersage im Progymnasium hinreichend ist, versagt die Schulnote als Leistungsprädiktor in der Realschule.

Dieses Ergebnis lässt verschiedene Interpretationen zu. Ein möglicher Grund für diesen Befund könnte darin bestehen, dass das Lehrpersonal in der Primarschule vor allem bei den besseren Schülerinnen und Schülern auf eine differenzierte Leistungsbeurteilung achtet. Dabei könnte die Berechtigungsfunktion der Note (wer 10 Punkte bekommt, kann ins Progymnasium) als wichtiges Entscheidungskriterium gelten. Die Notwendigkeit, schwächere

Schüler und Schülerinnen, die dieses Kriterium nicht erreichen, aufgrund ihrer Leistungsunterschiede ebenso feinstrukturiert zu beurteilen, entfällt. Ein anderer Grund für dieses Ergebnis könnte auch in der weiterführenden Schule selbst liegen. Danach könnten die im Progymnasium geforderten Leistungen eher dem Leistungsvermögen entsprechen, das in der Primarschule in der Note bewertet wurde, als die in der Realschule geforderten Leistungen.

2. Die *prognostische Validität des Übertrittstests* liefert bei beiden Schultypen substantielle Erklärungsbeiträge. Seine Bedeutung ist aber in der Realschule sehr viel grösser.

Dieses Ergebnis zeigt, dass der Übertrittstest in der Erfolgsvorhersage bezüglich der Abnehmerschule eine wichtige korrektive Funktion hat. Dies bedeutet auch, dass die Schulnoten und die Tests nicht dieselben Grundfähigkeiten messen, die für den späteren Schulerfolg entscheidend sind. Wenn wir die Prädiktionskraft des Tests im direkten Vergleich mit der Schulnote betrachten, bekommt man im Progymnasium und in der Realschule unterschiedliche Ergebnisse. Im Progymnasium machen die Noten bessere Vorhersagen, in der Realschule die Testleistungen. Wie wir in der Einleitung festgestellt haben, hat im aktuellen Selektionsprozess die Note eine Berechtigungsfunktion und der Test eine reine Selektionsfunktion. Aufgrund der Ergebnisse im Progymnasium könnte man diese Festlegung noch rechtfertigen. Die Ergebnisse in der Realschule würden eher dafür plädieren, dass dem Test Berechtigungsfunktion zukommt. Dies würde zur Annahme führen, dass bei Erreichen eines bestimmten Testergebnisses die leistungsmässigen Voraussetzungen vorhanden sind, um in der Schule erfolgreich zu sein. Allgemein stellt man fest, dass die beiden Testteile (Rechen- und Verbalteil) Schulleistungen schlechter vorhersagen als dies vollständige Intelligenzmessungen erfahrungsgemäss tun. Dies ist auch nicht weiter erstaunlich, da ein Intelligenztest noch weitere Dimensionen misst, als dies die verwendeten Leistungstests tun.

3. Die fachspezifische Prognose von Leistungen ist besser als die Prognose durch die Gesamtleistung.

Die Befundlage weist darauf hin, dass Leistungsvorhersagen am besten sind, wenn man fachspezifisch oder tätigkeitsspezifisch misst. Auch unsere Studie kann zwar zeigen, dass Notensummen aufgrund von Notensummen vorhergesagt werden können. Unsere Daten enthalten aber eine Reihe von Hinweisen, dass die fachspezifische Prognose die besten Ergebnisse bringt. Dies deutet darauf hin, dass es eher schwierig ist, durch Summation der Einzelleistungen auf eine zugrundeliegende «Generalfähigkeit» zu kommen, die für späteren, generellen Schulerfolg verantwortlich ist. Konkret müsste man sich nach diesem Befund überlegen, ob man das jetzige Summations-Selektionsmodell (10 Punkte aus Rechnen und Deutsch) nicht durch ein Einzelkomponenten-Selektionsmodell ersetzt, in dem die Kinder bei beiden Fächern mindestens



eine 5 erreichen müssen. Dasselbe würde für die Testergebnisse sprechen, falls der Test auch Berechtigungsfunktion bekommen sollte, was aufgrund der Prädiktionskraft durchaus diskutiert werden könnte. Wichtig ist es, hier festzuhalten, dass unsere Analyse die besten Werte bei der fachspezifischen Prognostik (z. B. von Deutsch auf Deutsch) erbracht hat. Dieses Postulat wird auch von Cardinet (1986) unterstrichen, wonach es unerlässlich ist, zwischen der Funktion von Beurteilungen und der Abstimmung der Instrumente und Verfahren auf die angestrebten Ziele der Beurteilungen zu unterscheiden. Die Notwendigkeit, diesem Umstand Rechnung zu tragen, wird auch durch unseren Befund betont, wonach die Vorhersage der Leistungen im Fremdsprachenunterricht durch die verwendeten Selektionskriterien gänzlich ungenügend ist.

4. Die Selektionspraktiken sollen periodisch evaluiert, reflektiert und nötigenfalls verändert werden.

Die bisherigen Analysen führen uns zu einigen Überlegungen darüber, wie eine optimierte Selektionspraxis aussehen könnte.

#### *Die Bedeutung der fachspezifischen Leistungsprognose*

Wenn sich Lehrer und Lehrerinnen wirklich bemühen, den ganzen Klassenverband in einem Fach auf einer Notenskala leistungsgerecht abzubilden, wird dies ein akzeptabel valides Urteil auf den späteren Facherfolg erlauben. Aufgrund der Ergebnisse unserer Analyse müsste aber verhindert werden, dass irgendwelche künstliche Grenzen bei verschiedenen Schülergruppen (Kinder über 10 Punkte, Kinder unter 10 Punkte) zu einer unterschiedlichen Notenskalen-Handhabung führen. Auch unsere Studie hat die prognostische Validität der Schulnoten bestätigt und gezeigt, dass in das Lehrerurteil offensichtlich Komponenten einfließen, welche durch irgendwelche Testverfahren nicht gemessen werden können.

#### *Die Bedeutung der kriteriumsorientierten Leistungsmessung*

Auch die Testleistungen haben in unserer Analyse ihren prognostischen Wert bewiesen. Auch wenn ihre Korrelationen mit den Schulleistungen nicht ganz so hoch sind wie diejenigen von Intelligenztests, waren sie doch imstande, beträchtliche Varianzanteile aufzuklären. Eine umfassende Intelligenzabklärung jedes Kindes wird für die Übertrittsselektion aus Praktikabilitätsgründen nicht ins Auge gefasst werden können. Dies wird eine Massnahme für die Einzeldiagnose und -beratung bleiben müssen. Aus unserem Befund aber, wonach Verbal- und Rechenteil des Übertrittstests eigene Vorhersagekraft auf den Facherfolg aufweisen, ist es naheliegend, den Lehrernoten kriteriumsorientierte Leistungsmasse auf Testbasis zur Seite zu stellen. Ziel müsste es sein, das Leistungsniveau jedes einzelnen Kindes in einem spezifischen Fach feststellen zu können. Ausgangspunkte sind immer spezifische Lernziele. Dabei nutzt man die den Tests zukommenden Vorteile einer der Note gegenüber grösseren Objektivität und Zuverlässigkeit. Ein weiterer Vorteil der kriteriumsbezogenen

Tests würde darin liegen, dass diese Leistungsmasse gleichzeitig die Funktionen einer Berechtigungsfunktion sowie einer Selektionsfunktion erfüllen würden, da die vorausgesetzten Standardisierungs- und Normierungsmöglichkeiten gegeben sind.

### *Summative und formative Beurteilung: Gegensätze oder Ergänzungen?*

Mit der vorgeschlagenen Kombination von fachspezifischer Notenberücksichtigung und kriteriumsorientierter Leistungsmessung hätte man die Voraussetzungen erfüllt, um eine verantwortungsvolle Selektionspraxis zu betreiben. Für ein Schulsystem, das heterogene Klassen mit differenzierenden Niveauebenen vorsieht (wie im geplanten, neuen Basler Schulsystem), stünden damit leistungsgerechte Selektionskriterien zur Verfügung, welche die Zuweisung für die Schüler und Schülerinnen jeder Leistungsstärke zu den optimalen Förderbedingungen erlauben. Ein solches Modell orientiert sich an der fachspezifischen Leistungsförderung, was aufgrund der Ergebnisse dieser Studie – zumindest aus der Perspektive der Leistungsprognose – ein gutes Modell darzustellen scheint.

Eine solche formative Beurteilung schliesst nun aber nicht aus, dass auch bei jedem Kind eine Prognose in bezug auf sein Gesamtleistungsvermögen und damit eine Selektion für gesamtleistungsdifferenzierende, weiterführende Schulen vorgenommen wird. Aufgrund unserer Ergebnisse schlagen wir vor, die Bewertungsfunktion der Leistungswerte nicht an Notensummen oder an Testleistungssummen zu binden, sondern den Leistungen in den Einzelfächern diese Bewertungsfunktion zukommen zu lassen. So müssten eben Schüler und Schülerinnen als Eintrittsberechtigung ins Gymnasium nicht einen bestimmten Summenwert aus Rechnen, Deutsch und Fremdsprache erreichen, sondern sie müssten in jedem der Fächer einen bestimmten Leistungsstand (Kriterium) erreichen. Unser Vorschlag steht in klarem Gegensatz zu den Forderungen nach generellen Prädiktoren (Schlüsselqualifikationen für die Schulbildung also). Hierzu gibt es Gründe theoretischer wie praktischer Art: In der psychologischen und pädagogischen Diagnostik kommt man immer mehr dazu, mittels Anforderungs- und Förderprofilen sowie Neigungsdifferenzierungen die Ressourcen eines Lernenden besser zu erfassen und zu aktivieren (Perrig-Chiello, 1992; Klingen, 1984). Zudem treffen wir in der Praxis zunehmend Mittelstufenmodelle, wonach heterogene Klassen mit fachspezifischen Differenzierungen (Niveauebenen) konzipiert sind (wie dies beispielsweise im neuen Basler Schulsystem der Fall sein wird).

### 5. Notwendigkeit eines vernetzten Denkens auch in der Schülerleistungsbeurteilung: Zukunftsperspektiven

Zum Schluss noch ein letztes Anliegen, das über unser Thema hinaus geht und doch stark mit ihm verhängt ist. Schülerbeurteilung «passiert» nicht im luftleeren Raum wie die vielen messtechnischen Analysen suggerieren mögen. Schülerleistungsbeurteilung steht vielmehr für Interaktion zwischen zwei Systemen, die sich a priori nicht freiwillig als Partner «gewählt» haben, nämlich

das System «Schule» und das System «Familie». Veränderungen in einem System haben automatisch Veränderungen im anderen System zur Folge. Die Schülerleistungsbeurteilung – mag sie verfahrenstechnisch noch so optimal sein – trägt diesem, scheinbar so trivialen Faktum kaum Rechnung und konzentriert sich auf die Dyade Lehrer – Schüler. Das jeweils andere System wird meist als Konkurrent angesehen, den man aufmerksam, ja argwöhnisch kontrolliert, anstatt durch transparenten, institutionalisierten Austausch eine entspannte Kooperation zu schaffen. Unser Postulat geht dahin, mindestens für die formative Beurteilung das System Familie einzubeziehen. Dem Argument, dies sei unökonomisch und mühsam, kann entgegengehalten werden, dass somit mancher Gang zum Schulpsychologen (wegen Leistungsversagen, Überforderung, Unterforderung, Schulangst, Schwänzen, Schlafstörungen, etc.) eingespart werden könnte.

### Literatur

- Cardinet, J. (1986) *Evaluation scolaire et pratique*. Bruxelles: De Boeck.
- Casparis, C. (1984) Eine Evaluation des Luzerner Übertrittsverfahrens. *Bildungsforschung und Bildungspraxis*, 6, 1, 17–29.
- Fingerhut, W. und Langfeldt, H. P. (1975). Leistungsbeurteilung durch Notengebung. In: Heller, K. (Hrsg.). *Leistungsbeurteilung in der Schule*. Heidelberg: Quelle und Meyer.
- Flammer, A., Perrig-Chiello, P. und Rüegg, Th. (1983). *Zeugnisnoten vor dem Übertritt in die Sekundarstufe*. Freiburg: Universitätsverlag.
- Heller, K. A. (Hrsg.) (1984/4) *Leistungsdiagnostik in der Schule*. Bern: Huber.
- Ingenkamp, K. (Hrsg.), (1976). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- Klingen, L. H. Leistungsbewertung im Kurswahlssystem der Sekundarstufe. In Heller, K. A. (Hrsg.) (1984/4) *Leistungsdiagnostik in der Schule*. Bern: Huber.
- Meili, R., Aebi, H. J., Heizmann, M.-L., Schoefer, E. (1976). *Intelligenz und Schulleistung in höheren Mittelschulen der Schweiz*. Forschungsabteilung des Psychologischen Institutes Bern.
- Perrenoud, Ph. (1991) Formative Schülerbeurteilung: Welcher Platz in der Didaktik? *Beiträge zur Lehrerbildung*, 9, 3, 309–330.
- Perrig-Chiello, P. (1981). Kausalattribution und Schülerleistungsbeurteilung durch den Lehrer. Dissertation, Universität Fribourg, 1980.
- Perrig-Chiello, P. (1983) Kausalattribution und Schülerleistungsbeurteilung durch den Lehrer. *Bildungsforschung und Bildungspraxis*, 5, 1.
- Perrig-Chiello, P. (1992) Die Diagnostik ist tot – es lebe die Diagnostik. *Psychoscope*, 5, 13, 27–28.
- Pierrehumbert, B. (Ed.) (1992) *L'échec à l'école: échec de l'école?* Lausanne: Delachaux et Niestlé.
- Reiber, G. (1973). Lehrerurteil, Aufnahmeprüfung und Test gegenüber Schulerfolg. In: *Schweizerische Lehrerzeitung*, 11, 443–445.
- Rohrer, R., Keller, B. und Flammer, A. (1981). Evaluation des Zuger Übertrittsverfahrens. Forschungsbericht Nr. 24, Psychologisches Institut der Universität Fribourg, Schweiz.
- SIPRI (1983) Werkstattbericht 2. Funktionen und Formen der Schülerbeurteilung. Genf: EDK.
- SIPRI (1985) Von der Primarschule in die Sekundarstufe I in der Schweiz. 3 Hefte. Bern: EDK.
- Steinkamp, G. (1972). Lehrer voller Vorurteile? In: Hielscher, H. (Hrsg.) *Die Schule als Ort sozialer Selektion*, Heidelberg: Quelle und Meyer.

- Schultze, W. (1964). Über den Voraussagewert der Auslesekriterien für den Schulerfolg am Gymnasium. *Forschungsbericht 1 der Max-Träger-Stiftung*, Frankfurt.
- Trier, U. P. (1974). Die Problematik von Entscheidungsgrundlagen beim Übertritt von der Primarschule in die weiterführende Schule. *Schweizer Schule*, 61, 999–1004.
- Ulshöfer, R. (1949). Zur Beurteilung von Reifeprüfungsaufsätzen. In: *Der Deutschunterricht*, 1, 8, 84–102.
- Weiss, R. (1989) *Leistungsbeurteilung in den Schulen – Notwendigkeit oder Übel?* Wien: Jugend und Volk.

## Sélection des élèves et pronostic des performances scolaires au banc d'essai. Analyse de la valeur pronostique du dispositif d'admission bâlois.

### *Résumé*

Les systèmes scolaires changent – les problèmes d'évaluation scolaire restent. En face d'un nouveau système scolaire Bâlois, les auteurs ont évalué le système de sélection actuel pour le passage de l'école primaire aux écoles secondaires (Progymnasium et Realschule) qui repose sur un test de performance et des notes. Les résultats:

- La validité pronostique des notes n'est suffisante que pour le Progymnasium, mais pas pour la Realschule.
- Pour le test de performance on peut dire que la validité pronostique est attestée tant pour le Gymnasium que pour la Realschule (tout en étant plus élevée pour la Realschule).

Des suggestions sont faites pour améliorer la pratique de sélection: on propose un modèle combinatoire d'évaluation sommative et formative, tout en soulignant la nécessité de la pensée systémique dans le processus de l'évaluation scolaire.

## Selection of pupils and prediction of scholastic achievement on the testing bench. Analysis of the predictive value of the admission procedure in Basel.

### *Summary*

Schoolsystems may change – but the problem of evaluation and selection of students is always there. This study evaluates the selection system (performance test and marks) used in Basel for the transition from grade school to secondary school (Gymnasium and Realschule). The results are:

- Even though the prognostic validity of marks is sufficient for predicting performance in Gymnasium, it is not for Realschule (where weaker students are taught).
- The predictive validity of the performance test on the other hand exists (particularly for the weaker students in the Realschule).

To optimize this selection-procedure a model is proposed which combines summative and formative evaluation. At the same time the importance of systemic thinking in students evaluation and selection is stressed.