



^b
**UNIVERSITÄT
BERN**

Virtueller Campus Erziehungswissenschaft
Abteilung Pädagogische Psychologie

Wirksamkeit von Schule und Unterricht

Leitung: Prof. Dr. Walter Herzog

Autorenteam: Prof. Dr. Walter Herzog, Prof. Dr. Armin Hollenstein, Simon Bärtschi, Annika Butters, Susanne Retsch, Hans Ryser

Unter Mitwirkung von: Irina Adamova, Barbara Andenmatten, Andreas Brunner, Sabine Campana, Tamara Carigiet, Daniela Blum, Franziska Gobat, Markus Heinzer, Robert Hilbe, Jolanda Huwiler, Daniel Ingrisani, Elena Makarova, Katrin Müller, Madeleine Pfäffli Schmid, Yvonne Seiler, Roman Suter, Peter Vetter

Universität Bern
Virtueller Campus Erziehungswissenschaft
Institut für Erziehungswissenschaft
Abteilung Pädagogische Psychologie
Muesmattstrasse 27
3012 Bern
www.edu.unibe.ch

6. Auflage 2010

Inhaltsverzeichnis

Einleitung: Wie wirksam sind Schule und Unterricht?	1
1 Bedingungen internationaler Schulleistungsvergleiche	3
1.1 Äussere Bedingungen internationaler Schulleistungsvergleiche	3
1.1.1 Internationalisierung der Politik	4
1.1.2 Empirische Sozialforschung	5
1.1.3 Demographischer Wandel.....	5
1.1.4 Wandel der politischen Steuerungskultur	6
1.1.5 Qualitätsansprüche	7
1.2 Methodische Ansprüche an internationale Schulleistungsvergleiche	9
1.2.1 Erzeugen Schulen überhaupt Wirkungen?.....	10
1.2.2 Welcher Art sind die Wirkungen, die Schulen erzeugen?.....	12
1.2.3 Wie lassen sich die Bereiche schulischer Wirksamkeit identifizieren?.....	13
1.2.4 Lassen sich die Wirkungen von Schule überhaupt messen?.....	15
1.2.5 Wonach bemisst sich der Erfolg einer Schule?	15
1.2.6 Was bedingt die Leistungen von Schülerinnen und Schülern?.....	18
1.2.7 Wenn Schulen wirksam sind, wie stellt man Wirksamkeit fest?.....	18
1.2.8 Wer soll untersucht werden?.....	22
1.2.9 Wie werden die Daten erhoben?	23
1.2.10 Wie werden die Daten ausgewertet?.....	24
1.2.11 Wie werden die Ergebnisse dargestellt?	24
1.3 Analysemodelle	25
1.3.1 Das Input-Output-Modell	25
1.3.2 Das Angebot-Nutzungs-Modell	26
2 Internationale Schulleistungsvergleiche am Beispiel der IEA Lesestudie (Study of Reading Literacy)	29
2.1 Generelles	29
2.1.1 Die Ziele international vergleichender Studien	29
2.1.2 Zur Durchführung international vergleichender Studien.....	31
2.1.3 Forschungslogik und ihre Probleme	33
2.1.3.1 Testkonstruktion: Vergleichbarkeit, Mehrdimensionalität und Vergleichsnorm.....	33
2.1.3.2 Stichprobenauswahl	38
2.1.4 Erkenntniswert	39
2.2 Ergebnisse	41

2.2.1	Die Lesekompetenz der Schweizer Schülerinnen und Schüler: Nationale Resultate in der dritten und der achten Klasse	41
2.2.1.1	Individuelle Merkmale	41
2.2.1.2	Familienmerkmale	41
2.2.1.3	Regionale Merkmale	44
2.2.1.4	Der Einfluss des Schultyps auf die Leseleistung	45
2.2.1.5	Zusammenfassung	45
2.3	Folgen und Konsequenzen	46
3	Internationale Schulleistungsvergleiche am Beispiel der TIMSS (Third International Mathematics and Science Study)	48
3.1	Generelles	48
3.1.1	Einleitung	48
3.1.2	Forschungslogik und ihre Probleme	48
3.1.2.1	Was wird verglichen?	49
3.1.2.2	Mehrdimensionalität des Untersuchungsgegenstandes	49
3.1.2.3	Wen soll man untersuchen?	50
3.1.3	Erkenntniswert	52
3.2	Ergebnisse	52
3.2.1	Zusammenfassung der wichtigsten Ergebnisse der Schweiz	52
3.2.1.1	Zentrale Untersuchungsergebnisse	53
3.2.1.2	Periphere Untersuchungsergebnisse	53
3.2.2	Zusammenfassung der wichtigsten Ergebnisse Deutschlands	54
3.2.3	Interpretation der Ergebnisse	55
3.2.3.1	Geschlechterunterschiede in der Mathematikleistung auf der Sekundarstufe I	55
3.2.3.2	Klassengrösse	55
3.2.3.3	Unterrichtszeit in der Mathematik und in den Naturwissenschaften auf der Sekundarstufe I	56
3.2.3.4	Öffentliche Bildungsausgaben	56
4	Internationale Schulleistungsvergleiche am Beispiel von PISA (Programme for International Student Assessment)	57
4.1	Generelles	57
4.2	Ergebnisse	61
4.2.1	Leistungskompetenzen	62
4.2.1.1	Lesekompetenz	62
4.2.1.2	Mathematik	66
4.2.1.3	Naturwissenschaften	70
4.2.2	Soziale Herkunft und Chancengleichheit	72
4.2.3	Kulturelle Vielfalt in der Schule: Herausforderung und Chance	74
4.2.4	Geschlechterunterschiede	76

5	Proximale Bedingungen schulischer Wirksamkeit	79
5.1	Einleitung	80
5.2	«Gute Schulen» (effective schools)	81
5.2.1	Eine Studie im Sekundarschulbereich	83
5.2.2	Eine Studie im Primarschulbereich.....	84
5.2.3	Eine zusammenfassende Darstellung der Forschungsliteratur	86
5.2.4	Ausblick	87
5.3	Prozess-Produkt-Forschung	88
5.3.1	Die Persönlichkeit als Wirkkraft	88
5.3.2	Das Prozess-Produkt-Paradigma	90
5.3.3	Ergebnisse der Prozess-Produkt-Forschung	91
5.3.3.1	Bedingungen auf Schülerseite: kognitive Bedingungen.....	92
5.3.3.2	Bedingungen auf Schülerseite: motivationale und emotionale Bedingungen	93
5.3.3.3	Bedingungen auf Schülerseite: relative Bedeutung der kognitiven und motivationalen Bedingungen	94
5.3.3.4	Zusammenfassende Beurteilung der Bedingungen auf Schülerseite.....	95
5.3.3.5	Bedingungen auf Seiten des Unterrichts: quantitative Bedingungen («time on task»).....	96
5.3.3.6	Bedingungen auf Seiten des Unterrichts: qualitative Bedingungen	96
5.3.3.7	Zusammenfassende Beurteilung der bisherigen Bedingungen.....	98
5.3.3.8	Bedingungen auf Seiten des Unterrichts: Kontextmerkmale und Lehrererwartungen.....	98
5.3.3.9	Bedingungen auf Seiten der Familie	98
5.3.3.10	Gesamtbeurteilung der genannten Bedingungen	99
5.4	Experimentelle Unterrichtsforschung	99
5.4.1	Der Begriff des Experiments	100
5.4.2	Grenzen der experimentellen Forschung in pädagogischen Situationen	100
5.4.3	Interventionsstudien	102
5.5	Metaanalysen	103
5.6	Modelle	107
5.6.1	Zwei einfache Modelle von Krapp (1976).....	108
5.6.2	Das Modell von Carroll (1973).....	110
5.6.3	Modelle im Zusammenhang mit Qualitätssystemen.....	112
5.6.4	Komplexe- und Mehrebenenmodelle.....	115
5.6.5	Einbezug von Kontextbedingungen	116
5.7	Rückblick und Ausblick.....	117
6	Literatur	123
6.1	Zitierte Internetseiten	129

Einleitung: Wie wirksam sind Schule und Unterricht?

Vielleicht wird die Geschichtsschreibung das Jahr 2002 als Jahr in Erinnerung behalten, in dem die Schweiz und das Nachbarland Deutschland einen Bildungsschock erlitten haben. Ausgelöst durch Ergebnisse einer international vergleichenden Studie, die den 15-jährigen Schweizer Schülerinnen und Schülern im Lesen und in den Naturwissenschaften lediglich mittelmässige Leistungen attestierten, haben die Medien ein düsteres Bild unseres Schulsystems gezeichnet. «Viele Schüler verstehen nur Bahnhof» titelte etwa provokativ der Zürcher «Tages Anzeiger» (5.12.2001, p. 8). Jeder fünfte Jugendliche sei kaum auf das Berufsleben vorbereitet. Die PISA-Studie, von der die Rede ist¹, deckte «Schwachstellen der Volksschule auf», hiess es in der «Sonntags Zeitung» (16.12.2001, p. 5). Drei solcher Schwachstellen wurden aufgeführt: (1) «Bei der Integration der ausländischen Kinder hat die Schule versagt.» (2) «Die Kinder werden in der Schweiz zu spät eingeschult.» (3) «Der Sprachunterricht erreicht sein Ziel nicht» (ebd.).

Im Lande Pestalozzis wurde beklagt, dass die Schule ihre Ziele nicht erreiche. Wieder sei ein Mythos geplatzt, schrieb der Berner «Bund» (10.1.2002, p. 13): «Die Schweizer Grundschulen [...] produzieren nur biederes Mittelmass.» Voller Besorgnis hiess es weiter: «Wenn der Bildungsstand der Schweizer nur für einen internationalen Mittelfeldplatz reicht, ist zu fragen, wie lange der Spitzenplatz im Pro-Kopfeinkommen zu halten sein wird» (ebd.). Relativierend war dann in einer Äusserung des Präsidenten der Schweizerischen Konferenz der Kantonalen Erziehungsdirektoren (EDK), dem St. Galler Regierungsrat Hans Ulrich Stöckling, zu entnehmen, die Ergebnisse der PISA-Studie würden «nicht allein auf Probleme in der Schule» hinweisen, «sondern auch auf gesamtgesellschaftliche» (Der Bund 8.3.2002, p. 13).

Demnach scheint die Schule nicht *allein* verantwortlich zu sein für den «Bildungsnotstand», von dem die «Sonntags Zeitung» bereits am 26. November 2000 berichtet hatte. Dieselbe «Sonntags Zeitung» warf den Erziehungsdirektoren später vor, das «wahre Ausmass der Schweizer Bildungsmisere» zu verkennen und das «PISA-Debakel» zerredend herunterzuspielen (Sonntags Zeitung 10.3.2002, p. 21).

Infolge der Resultate von PISA 2000, welche für die Schweiz eher unerwartet kamen, wurden die Ergebnisse der zweiten Serie von PISA (PISA 2003) sowohl von Bildungsverantwortlichen wie auch von den Medien ungeduldig erwartet. In dieser zweiten Serie schnitten die Schulabgängerinnen und Schulabgänger in der Mathematik und den Naturwissenschaften im internationalen Vergleich besser ab. In den Medien, welche das Thema wiederum eingehend behandelten, wurde dies einerseits als Zeichen dafür interpretiert, dass innert kurzer Zeit Fortschritte möglich sind (Der Bund 7.12.2004, p. 2). Andere wie etwa der Dachverband der Schweizer Lehrerinnen und Lehrer (LCH) zeigten sich zwar erfreut, aber dennoch zurückhaltend und warnten vor verfrühtem Enthusiasmus (Der Bund 8.12.2004, p. 8).

Bei aller Heterogenität zeigt die Verarbeitung der Ergebnisse der PISA-Studie in den Medien einen klaren Fokus: Es geht um die Frage der *Leistungsfähigkeit* und *Wirksamkeit* unseres Schulsystems. Wenn im Rahmen einer international vergleichenden Studie wie PISA die Leistungen von Schülerinnen und Schülern am Ende der obligatorischen Schulzeit erfasst werden, dann stellt sich zu Recht die Frage, welchen Anteil die *Schule* an den gewonnenen Resultaten hat. Dass die Schule nicht *allein* für die Kompetenzen 15-jähriger Jugendlicher verantwortlich ist, kommt in den zitierten Medienberichten zum Ausdruck. Dass sie aber in Bezug auf die Lesefertigkeiten sowie die mathematischen und naturwissenschaftlichen Kenntnisse der Schulabgängerinnen und Schulabgän-

¹ Konkret ist hier die Rede von PISA 2000, der ersten von drei PISA-Studien. Unterdessen sind auch die Resultate von PISA 2003 sowie PISA 2006 erschienen. PISA steht für «Programme for International Student Assessment». Im Kapitel 4 gehen wir ausführlich auf PISA ein.

ger eine durchaus *wesentliche* Ursache spielt, dürfte kaum zu bestreiten sein. Denn wo sonst sollten die Schülerinnen und Schüler die entsprechenden Kenntnisse und Kompetenzen her haben? Zu Recht nehmen wir an, dass Schule und Unterricht *Wirkungen* haben. Aber *welche* Wirkungen haben sie? Und wie können wir diese Wirkungen *feststellen*?

Diese Fragen liegen im Kern einer inzwischen ausgedehnten Forschung zur Wirksamkeit von Schule und Unterricht, mit der wir uns im Folgenden auseinandersetzen wollen. Dabei verfolgen wir das Ziel, Sie als Leserin und Leser dieses Grundlagentextes mit den wesentlichen Forschungsansätzen zur Wirksamkeit von Schule und Unterricht sowie deren Ergebnissen vertraut zu machen. Im Vordergrund stehen die internationalen Schulleistungsvergleiche. Diese Studien stehen – wie die eingangs zitierten Medienberichte zeigen – im Fokus des öffentlichen Interesses. Ihre Bedeutung wird in den nächsten Jahren zunehmen, wofür wir im 1. Kapitel einige Gründe darlegen werden (Abschnitt 1.1). Da diese Studien sehr oft zu kritischen Urteilen über unsere Schulen und die Lehrkräfte, die an ihnen unterrichten, führen, scheint es uns wichtig zu sein, angehenden Lehrerinnen und Lehrern zu zeigen, wie die Studien durchgeführt werden, was ihre Ziele und welches die verwendeten Methoden sind, wie man die Ergebnisse interpretieren muss und wo ihre Grenzen liegen (Abschnitt 1.2 und 1.3).

Die drei folgenden Kapitel stellen je eine der in den letzten Jahren durchgeführten internationalen Studien zur schulischen Wirksamkeit (Effektivität), an denen auch die Schweiz beteiligt war, vor. Es handelt sich um die «IEA Reading Literacy Study» (IEA Lesestudie; vgl. Kapitel 2), die «Third International Mathematics and Science Study» (TIMSS; vgl. Kapitel 3) und das «Programme for International Student Assessment» (PISA; vgl. Kapitel 4). Solche Studien sind sehr aufwändig; sie setzen eine entsprechende organisatorische und forschungsmethodische Infrastruktur voraus. Die Voraussetzungen der Studien werden im Kapitel 1 diskutiert. Bei allen Themen finden Sie Hinweise zu interessanten weiterführenden Texten.²

Bei allem Aufwand, der für solche internationalen Schulleistungsvergleiche getrieben wird, der Aussagewert der Studien ist begrenzt. Das zeigt sich vor allem, wenn man andere Forschungsansätze zum Vergleich bezieht. Dies ist das Thema des 5. Kapitels, in dem wir uns mit Forschungsstrategien auseinandersetzen, die auf einer weniger globalen Ebene ansetzen und sich vermehrt den *Prozessen* zuwenden, die für die Wirksamkeit von Schule und Unterricht verantwortlich sind. Wir nennen dies die proximalen Bedingungen schulischer Wirksamkeit – im Unterschied zu den distalen Bedingungen auf der Schulsystemebene.

Es könnte sein dass Sie sich fragen, welchen Nutzen das Thema «Wirksamkeit von Schule und Unterricht» für Ihre berufliche Praxis abwirft. In der Tat muten wir Ihnen zu, sich mit einer Materie auseinanderzusetzen, deren praktischer Nutzen nicht auf Anhieb ersichtlich ist. Wir glauben aber, dass die folgenden Ausführungen Sie auch dann ansprechen sollten, wenn sie wenig Interesse an internationalen Schulleistungsvergleichen haben. Denn auch als Lehrerin oder Lehrer sind Sie vor die Frage gestellt, welche Wirkungen von Ihrem Handeln ausgehen. Unausweichlich haben Sie ein bestimmtes Verständnis der Wirksamkeit dessen, was sie berufsmässig tun. Verstehen Sie sich als Input, der Kausalprozesse auslöst, die nach berechenbarer Wahrscheinlichkeit Lernprozesse erzeugen? Sehen Sie sich als jemanden, der Angebote macht, die von den Schülerinnen und Schülern genutzt werden können? Wie weit vermögen Sie Ihre Schülerinnen und Schüler zu beeinflussen? Wie gross ist Ihr Einfluss im Vergleich zu anderen pädagogischen Wirkfaktoren wie den Eltern, den Medien oder den Peers?

Was wir in diesem Grundlagentext mit einem wissenschaftlichen Anspruch zu behandeln versuchen, hat durchaus praktische Bedeutung, denn es tangiert unmittelbar die Vorstellungen und Erwartungen, die Sie als Lehrkraft bezüglich ihrer Wirkungen im Unterricht haben.

² Die Texte zur fakultativen Lektüre finden Sie online in der Lernumgebung des Moduls.

1 Bedingungen internationaler Schulleistungsvergleiche

Internationale Schulleistungsvergleiche wie TIMSS oder PISA sind nicht ohne Voraussetzungen. Es gibt solche politischer und es solche forschungsmethodischer Art. Mit beiden Voraussetzungen wollen wir uns im Folgenden vertraut machen. Wir diskutieren im Abschnitt 1.1 zunächst einige Bedingungen «äusserer» Art, womit wir im wesentlichen politische Rahmenbedingungen meinen wie die Internationalisierung der Politik bzw. die Globalisierung des gesellschaftlichen Zusammenlebens, den demographischen Wandel in Europa, die Verknappung der Mittel der öffentlichen Hand und den Wandel der politischen Steuerungskultur.

Im Abschnitt 1.2 werden wir uns mit Bedingungen «innerer» Art auseinandersetzen, womit die forschungsmethodischen Voraussetzungen international vergleichender Schulleistungsstudien gemeint sind. Es ist uns ein Anliegen, Sie als Leserin und Leser nicht einfach mit Ergebnissen von Studien zur schulischen Wirksamkeit einzudecken, sondern Sie auch in die Lage zu versetzen, solche Studien in ihrer Logik und Bedeutung zu beurteilen. Das setzt voraus, dass wir Sie wenigstens im Grundsatz mit den Forschungsmethoden vertraut machen, die in solchen Studien verwendet werden, denn nur so lässt sich erkennen, wo die Grenzen der Studien liegen. Bereits in diesem Kapitel werden wir uns schliesslich unter 1.3 auch schon mit einigen Fragen der Modellierung des Forschungsgegenstandes beschäftigen, d.h. mit der Frage, wie man sich denn die Wirksamkeit von Schule und Unterricht vorzustellen hat.

1.1 Äussere Bedingungen internationaler Schulleistungsvergleiche

International vergleichende Studien zu den Leistungen von Schülerinnen und Schülern werden erst seit knapp 40 Jahren durchgeführt. Das heisst nicht, dass es Erhebungen zur Qualität von Schule und Unterricht vorher nicht gegeben hat. So hat die Schweiz lange Zeit sogenannte Rekrutenprüfungen durchgeführt, deren Funktion (auch) in der Überprüfung des Bildungsstandes der Bevölkerung (allerdings beschränkt auf junge Männer) lag (vgl. Lustenberger 1996). Nicht immer konnten diese Rekrutenprüfungen für sich in Anspruch nehmen, wissenschaftlich abgestützt zu sein. Des Weiteren fehlte ihnen das entscheidende Kriterium der heutigen Schulleistungsvergleiche: die Internationalität.

Die erste internationale Studie dieser Art, die «First International Mathematics Study» (FIMS), stammt aus dem Jahre 1964. Einen Überblick über weitere internationale Schulleistungsstudien gibt Tabelle 1.1.

Tabelle 1.1: Überblick über einige internationale Schulleistungsvergleiche

FIMS	First International Mathematics Study	1964
	Six Subject Survey	1966-1973
FISS	First International Science Study (= Teil des Six Subject Survey)	1970/71
SIMS	Second International Mathematics Study	1980-1982
SISS	Second International Science Study	1983/84
IRLS	International Reading Literacy Study	1990/91
CES	Classroom Environment Study	1993/94

TIMSS	Third International Mathematics and Science Study	1994/95
PISA	Programme for International Student Assessment	2000/2003/2006

Weshalb werden internationale Vergleiche zu den Leistungen von Schülerinnen und Schülern erst seit relativ kurzer Zeit durchgeführt? Was sind die Bedingungen, die gegeben sein müssen, damit solche Studien realisiert werden können? Es gibt zwei triviale Gründe, die im Folgenden relativ kurz abgehandelt werden, und drei substantielle Gründe, die wir etwas ausführlicher diskutieren wollen. Im Einzelnen erörtern wir die folgenden fünf Bedingungen:

1. Bedingung: Internationale Organisationen, d.h. der Wille von Nationalstaaten, sich im Rahmen von supranationalen Vereinigungen (wie UNO, UNESCO, OECD oder IEA) zusammenzuschließen und zu kooperieren.
2. Bedingung: Entwicklung der Sozial- und Erziehungswissenschaften zu empirisch forschenden Disziplinen mit einem elaborierten Instrumentarium an Forschungs- bzw. Evaluationsmethoden.
3. Bedingung: Demographischer Wandel, der zu einer Verschiebung des Anteils alter Menschen zulasten von jungen führt und dadurch zu einer Umverteilung der öffentlichen Gelder nötig, die zudem knapper werden.
4. Bedingung: Wandel der politischen Steuerungskultur von der Input- zur Outputseite (New Public Management).
5. Bedingung: Eindringen des Qualitätsdiskurses ins Bildungswesen.

1.1.1 Internationalisierung der Politik

Ein erster, eher trivialer Grund für die kurze Tradition internationaler Schulleistungsvergleiche ist institutioneller Art. Internationale Vergleichsstudien setzen internationale Organisationen voraus, die solche Studien tragen, finanzieren und durchführen. Wenn wir in Rechnung stellen, dass die UNO erst 1945 gegründet wurde, dann ist offensichtlich, dass eine internationale politische Optik kaum älter als 50 Jahre ist. Zwar gibt es eine Vorläuferorganisation der UNO, den Völkerbund; wie jedoch die beiden Weltkriege zeigen, ist eine Mentalität der internationalen Kooperation nicht sehr alt.³ Die UNESCO, die für Bildung und Wissenschaft zuständige Unterorganisation der UNO, wurde ebenfalls 1945 gegründet. Die OECD, eine Organisation europäischer Staaten, die massgeblich hinter der PISA-Studie steht, wurde 1960 gegründet, wobei auch hier seit 1948 eine Vorgängerorganisation existierte.

Diese wenigen Daten genügen, um zu zeigen, dass ein Wille zu internationaler Zusammenarbeit noch nicht lange besteht, so dass auch im Bildungsbereich nicht erwartet werden kann, dass man sich schon lange dafür interessiert, wie das eigene Land im Vergleich zu anderen steht. Es war gleichsam immer schon klar, dass man *besser* ist als der Nachbar.⁴

³ Zur Zeit der Gründung der Vorläuferorganisation der UNO, dem Völkerbund, war die Strömung gegen eine internationale Zusammenarbeit im Bereich der Bildung noch so stark, dass der Völkerbund beschloss, pädagogische Fragen ganz aus seinem Tätigkeitsbereich auszuklammern.

⁴ Dieses Urteil mag etwas übertrieben klingen. Immerhin gibt es in Genf seit 1925 ein «Bureau International d'Éducation» (BIE), das sich nicht zuletzt unter Jean Piaget, der von 1929 bis 1967 dessen Direktor war, der vergleichenden Analyse von Bildungssystemen zugewandt hat. Trotzdem kann man nicht sagen, dass das BIE viel zur Förderung einer eigenständigen international vergleichenden Forschung beigetragen hat.

1.1.2 Empirische Sozialforschung

Ein zweiter, wiederum eher trivialer Grund für die relative Neuheit internationaler Schulleistungsvergleiche ist methodischer Natur. Vergleiche von Schulsystemen und deren Wirkungen sind nicht leicht durchzuführen. Gefordert ist nicht nur eine *organisatorische* Infrastruktur, sondern auch eine adäquate *Forschungsmethodologie*. Die Methodologie für solche Studien gibt es aber wiederum erst seit relativ kurzer Zeit. Die Sozialwissenschaften (inklusive die Erziehungswissenschaften) sind vergleichsweise jungen Datums. Ihre Geschichte ist kaum älter als 100 Jahre. Wobei die ersten Jahrzehnte eher an theoretischen Fragen orientiert waren – so waren die soziologischen Klassiker im wesentlichen Theoretiker der Disziplin, wie Emile Durkheim in Frankreich, Max Weber oder Georg Simmel in Deutschland und Talcott Parsons in den USA. Eine *empirische* Disziplin ist die Soziologie – und vergleichbares gilt noch mehr für die Erziehungswissenschaft – erst in den fünfziger und sechziger Jahren des 20. Jahrhunderts geworden. Hier stehen Figuren wie Paul Lazarsfeld und René König im Vordergrund, letzterer vor allem als Herausgeber des ersten deutschsprachigen «Handbuchs der empirischen Sozialforschung» (1. Auflage 1967) und als Mitbegründer des «Forschungsinstituts für Soziologie der Universität zu Köln».⁵

Erst seit den 1950er und 1960er Jahren stehen Forschungsmethoden zur Verfügung, die der anspruchsvollen Aufgabe internationaler Schulleistungsvergleiche genügen. Wobei gleich anzumerken ist, dass gerade in methodischer Hinsicht auch heute noch viele Fragen offen stehen und die Studien in ihrem Aussagewert oft deutliche Grenzen aufweisen.⁶

Neben der Internationalisierung der Politik und der Ausrichtung der Sozialwissenschaften als empirische Disziplin müssen weitere Bedingungen gegeben sein, damit internationale Schulleistungsvergleiche durchgeführt werden können. Es sollen im Folgenden drei weitere Bedingungen diskutiert werden, die sich allerdings nicht immer trennscharf voneinander abgrenzen lassen.

1.1.3 Demographischer Wandel

Das Interesse an internationalen Schulleistungsvergleichen ist auf dem Hintergrund demographischer Probleme, mit denen sich moderne Gesellschaften konfrontiert sehen, gewachsen. Gemeint ist die sich abzeichnende Verschiebung der Anteile junger und alter Menschen an der Gesamtbevölkerung: Der Anteil jüngerer Leute nimmt kontinuierlich ab (hervorgerufen durch geringere Geburtenraten), der Anteil älterer Leute hingegen nimmt zu (hervorgerufen durch eine höhere Lebenserwartung).

Es trifft zwar nicht zu, dass internationale Schulleistungsvergleiche eine direkte Antwort auf die demographischen Probleme moderner, westlicher Gesellschaften sind, trotzdem werden solche Studien nicht einfach zum Selbstzweck durchgeführt. Sie stehen in Relation zu bestimmten Themen der politischen Agenda. Dazu gehören die Finanzierungsprobleme des Staates, die diesen dazu zwingen, seine Ausgaben vermehrt zu legitimieren. Bezüglich der Bildung heisst dies: Überprüfung der *Effektivität* (Wirksamkeit) der öffentlichen (staatlich finanzierten) Schulen, Nachweis der *Qualität* der Bildungseinrichtungen und Steigerung der *Effizienz* der staatlichen Mittel. Zusammennehmen kann man diese Punkte unter dem Prinzip der *Rechenschaftslegung*: Die öffentliche Hand verlangt vom Bildungssystem in zunehmendem Masse, dass es über seine Leistungen Rechenschaft ablegt. Dies deshalb, weil der Staat die ihm verfügbaren und knapper werdenden Mittel neu verteilen muss bzw. die Investitionen in die Bildung angesichts des Bedarfs an öffentlichen Geldern für die Betagten der Rechtfertigung bedürfen.

⁵ Auch hier ist eine Vorgängerinstitution zu nennen, nämlich das «Forschungsinstitut für Sozialwissenschaften», das 1919 gegründet und von der Stadt Köln getragen wurde. Das Institut wurde im Jahr 1934 aufgelöst.

⁶ Darauf gehen wir im Abschnitt 1.2 ansatzweise und im Kapitel 5 ausführlich ein.

1.1.4 Wandel der politischen Steuerungskultur

In enger Beziehung zu den politischen Konsequenzen des demographischen Wandels steht die zurzeit beobachtbare Akzentverschiebung der politischen Steuerungskultur. Tatsächlich zeichnet sich seit einiger Zeit ein politischer Kulturwandel ab. Darauf verweisen Stichworte wie «New Public Management» (NPM), «Leistungsaufträge» und «Evaluation».

Im Bildungswesen ist man bisher davon ausgegangen, dass aus einer Investition linear ein entsprechender Ertrag abgeleitet werden kann. Die Steuerung des Bildungssystems erfolgte über die *Inputseite*: Zuweisung von Finanzen, Revision der Lehrpläne, Reform der Schulstrukturen, Erhöhung bzw. Kürzung der Lektionenzahl oder der Lehrerlöhne, Erneuerung der Schulhäuser, Verbesserung der Lehrmittel sowie generell die Erhöhung bzw. Kürzung der Mittel der öffentlichen Hand, die ins Bildungswesen (und nicht anderswohin) fließen. Dabei wurde unterstellt, dass «mehr desselben» (mehr Geld, mehr Lohn, mehr Unterrichtszeit etc.) *automatisch* eine Steigerung der Effekte von Bildung (des Outputs) zur Folge hat. Diese Input-Steuerung wird vor unseren Augen abgelöst durch eine gegenteilige Optik: durch den Blick auf den Output und den Versuch, das Bildungssystem über die Kontrolle seiner «Produktivität» zu steuern. Man will nur dann mehr ins System investieren, wenn die Investitionen durch den Output gerechtfertigt werden. Man macht m.a.W. die Zuweisung der Mittel von den Erträgen abhängig, die sie abwerfen.

Die Umlenkung der politischen Aufmerksamkeit vom Systeminput zum Systemoutput setzt voraus, dass der Output *überprüft* werden kann. Allgemeiner gesagt, ein System, das *Wirkungen* erzeugt, dessen Wirkungen aber nicht *messbar* sind, lässt sich nicht über Output-Steuerung kontrollieren. Oder nochmals anders formuliert: Output-Steuerung impliziert *Evaluation* und *Qualitätssicherung*.

Evaluation steht zweifellos nicht nur in Beziehung zum Wechsel von der Input- zur Output-Steuerung des Bildungssystems. Evaluation ist auch eine Sache der Unterrichtsebene – also nicht nur der Makroebene von Schule, sondern auch der Mikroebene der pädagogischen Prozesse. Man unterscheidet daher zwischen interner und externer Evaluation. Interne Evaluation ist das, was eine Lehrkraft oder auch eine Schule tut, um die Qualität ihres Unterrichts zu überprüfen. Externe Evaluation ist mehr oder weniger dasselbe, wenn es von einer unabhängigen Instanz von aussen gemacht wird. Allerdings ermöglicht die externe Evaluation auch einen Blick über die Unterrichtsebene hinaus. Sie erlaubt die *vergleichende* Beurteilung einer Schule oder ganzer Schulsysteme. Insofern stellen internationale Schulleistungsvergleiche Instrumente der *externen Evaluation von Bildungssystemen* dar, die durchaus auf die Ebene von Einzelschulen konkretisiert werden können. Dementsprechend heisst auch die IEA, d.h. jene Institution, die bereits eine Reihe von solchen internationalen Leistungsvergleichen durchgeführt hat, «International Association for the *Evaluation of Educational Achievement*».

Allerdings stehen zum Teil auch aus methodischen Gründen bei internationalen Evaluationen nicht so sehr die Einzelschulen im Vordergrund, sondern eher der *Systemvergleich*, wobei – wie die PISA-Studie zeigt – die Systemebene verschieden gewählt werden kann. Dies können Länder (Nationalstaaten) sein, es können aber auch kleinere Einheiten sein, wie in der Schweiz die Kantone oder in Deutschland die Bundesländer. Dass auch diese Vergleiche medienwirksam sind, zeigen die Ergebnisse von PISA 2000 aus den Kantonen Bern, Zürich und St. Gallen. Zwar sind die Unterschiede objektiv gesehen kaum von Belang, im «Bund» war trotzdem zu lesen: «St. Galler liegen vor den Bernern, die Berner schlagen die Zürcher» (2.10.2002, p. 17). Auch der «Tages Anzeiger» äusserte sich diesbezüglich wenig zurückhaltend: «Zürcher Schüler am schwächsten» (2.10.2002, p. 21). Weit nüchterner urteilte die «Neue Zürcher Zeitung»: «Kaum Leistungsunterschiede trotz Föderalismus» (2.10.2002, p. 47), während die «NZZ am Sonntag» wiederum eine

deutlichere Sprache fand und – gemünzt auf den damaligen Zürcher Bildungsdirektor – titelte: «Schlechte Noten für den Oberlehrer» (6.10.2002, p. 19).⁷

Zusammengenommen geht es bei dieser vierten Bedingung für internationale Schulleistungsvergleiche um einen politischen Kulturwandel. Die Gelder der öffentlichen Hand, die aufgrund der genannten demographischen Veränderungen *knapper* werden, sollen dem Bildungssystem nicht länger einfach zugeteilt werden, sondern orientiert an dem, was das System effektiv leistet, d.h. orientiert am *Output*, der folglich *evaluiert* werden muss. Internationale Schulleistungsvergleiche sind in diesem Sinne Instrumente zur Evaluation des Outputs von Bildungssystemen auf einer relativ globalen Ebene.

Es ist davon auszugehen, dass der Wandel der politischen Steuerungskultur über die Institutionalisierung von internationalen Leistungsvergleichen indirekt verstärkt wird. So ist Weinert (2001) der Ansicht, dass sich mit dem zunehmenden Gewicht, das international vergleichende Studien bekommen, die Argumentationslage der Bildungspolitik ändert. Weinert schreibt, man dürfe

»[...] erwarten, dass sich die Begründungen bildungspolitischer Entscheidungen gegenüber der Öffentlichkeit künftig stark verändern werden. Populäre Ideen, erhoffte Bildungswirkungen und suggestive Behauptungen werden in künftigen Diskussionen über Veränderungen von Schulstrukturen nicht mehr reichen. Man wird zumindest eine Expertise über den empirischen Forschungsstand und den internationalen Erfahrungsbestand sowie Aussagen über die geplante Begleitforschung verlangen.» (Weinert 2001a, p. 29).

Dieser Einschätzung Weinerts ist ebenso zuzustimmen wie seinen folgenden Bemerkungen:

«Leistungsmessungen in Schulen werden künftig eine grössere Rolle als bisher spielen. Ihren Ergebnissen kommt für die Fundierung und Überprüfung bildungspolitischer Entscheidungen wie für die Qualitätsentwicklung der einzelnen Schulen und des Unterrichts eine wachsende Bedeutung zu. Das gilt allerdings nur unter der Voraussetzung, dass die Studien wissenschaftlich und praktisch halten, was sie versprechen. Methodisch unzuverlässige, gegen variable Kontextbedingungen unsensible und die eigentlichen pädagogischen Fragestellungen verfehlende Untersuchungen können für die Entwicklung des Schulwesens eher schädlich sein und bei den interessierten Politikern und betroffenen Lehrern grosse Glaubwürdigkeitseinbussen in die Wissenschaft als einer neutralen und nützlichen Instanz verursachen.» (ebd., p. 30).

1.1.5 Qualitätsansprüche

Nach einer Zeit der Expansion des schweizerischen Bildungswesens in den 1970er und 1980er Jahren stossen Bund und Kantone bei der Finanzierung von Schulen und Hochschulen in jüngster Zeit an die Grenzen ihrer Möglichkeiten. Wachstum ist nur mehr möglich durch mehr Effizienz bei der Nutzung der vorhandenen Mittel. Die Perspektive der Quantität weicht derjenigen der *Qualität*. Seit einiger Zeit ist die Rede von Schulqualität, von Kriterien «guter» Schulen, von Qualitätsentwicklung, Qualitätssystemen und Qualitätsmanagement – gerade auch in Bezug auf Bildungsinstitutionen. Obwohl nicht immer klar ist, was unter Qualität genau zu verstehen ist, geht es bei Qualitätsurteilen letztlich um *Bewertungen*: Die Qualität von Schule und Unterricht ist gut oder schlecht – oder sie liegt irgendwo dazwischen.

⁷ Kaum beachtet wurde in den deutschsprachigen Schweizer Medien das vergleichsweise gute Abschneiden der welschen Schülerinnen und Schüler in der Erhebung PISA 2000. Die Freiburger Schülerinnen und Schüler schnitten sowohl im Lesen als auch in Naturwissenschaften und Mathematik am besten ab. Dicht dahinter folgte der Kanton Wallis. Vergleichsweise schlecht waren dagegen die Leistungen in den Kantonen Genf und Jura.

Die Unklarheiten über die Kriterien, die der Qualitätsbegriff zu erfüllen hat, ergeben sich aus seiner Herkunft. Qualität ist nicht eigentlich ein Begriff der deutschen Sprache - jedenfalls nicht seiner aktuellen Herkunft nach. Qualität ist die Eindeutschung des angelsächsischen Wortes «quality». Anders als im Deutschen, wo wir Qualität mit einer besonderen Eigenschaft assoziieren, wobei wir an positive Merkmale (an Kriterien der Güte) denken, meint «quality» im Englischen ganz einfach Standard bzw. das Verfahren zur Festsetzung eines Standards (vgl. Winkler 2000, p. 148). «Total Quality Management» meint in diesem Sinne nicht spezifische Merkmale von Leistungen oder Produkten, sondern die bloße Forderung, dass der Vorgang der Leistungserbringung einer gründlichen Prüfung hinsichtlich seiner Optimierungsmöglichkeiten unterzogen wird. Dementsprechend formal fallen die Ansprüche aus, die an solche Prüfverfahren gestellt werden. Viele Qualitätssysteme geben nur eine Orientierung hinsichtlich der Verfahren, die eingesetzt werden sollen, um ein Produkt zu prüfen. Eindrücklich in dieser Hinsicht sind die ISO-Normen, die im wesentlichen Betriebsabläufe erfassen, egal um welchen Betrieb oder um welches Produkt es sich handelt.

Der rein formale Charakter des Qualitätsbegriffs kommt in der Definition zum Ausdruck, welche die ISO («International Organization for Standardization») vorschlägt: «Qualität ist die Gesamtheit von Eigenschaften und Merkmalen eines Produkts oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Bedürfnisse beziehen» (ISO 1988). In dieser Definition wird kein einziger Inhalt angesprochen. Die Sprache, die verwendet wird, ist an Unternehmen ausgerichtet, die für Märkte produzieren. Die Rede ist denn auch von einem Produkt bzw. einer Dienstleistung, die Bedürfnissen genügen sollen.

Die Analyse dessen, was man leistet bzw. zu leisten beansprucht, in Form von Produkten und operationalisierten Leistungskriterien, ist im Falle von Bildungsinstitutionen zumindest ungewohnt, wenn nicht befremdlich. Bildung im ursprünglichen Sinn des Wortes verweist auf Mündigkeit, kritische Vernunft, Moralität und Verantwortung. Lassen sich diese Konzepte operationalisieren? Lassen sie sich wie die Produkte eines Unternehmens zählen und messen? Gibt es Standards für Bildungsinstitutionen, die sich operational definieren lassen?⁸

Man sieht, worauf die Fragen hinaus wollen: Internationale Schulleistungsvergleiche folgen derselben Logik wie der Qualitätsdiskurs. Sie setzen voraus, dass man den Output von Bildungssystemen und Bildungsprozessen *messen* kann. So kann verständlich werden, wenn gelegentlich gesagt wird, die Funktion internationaler Schulleistungsvergleiche liege im Benchmarking, d.h. in Vergleichen, die eine *Rangierung* von Schulsystemen erlauben.⁹

Damit führt der Qualitätsdiskurs zurück zum bereits diskutierten Wandel der politischen Steuerungskultur. Das Thema Qualität steht im Zusammenhang mit Diskussionen um die Kostendämpfung im Bildungswesen, um Verwaltungsreformen (NPM), um die Einführung von Marktelementen im öffentlichen Bildungswesen («teilautonome» Schulen, Trennung von strategischer und operativer Führung, Globalbudgets, Leistungsaufträge, Schulleitbilder etc.) – alles Diskussionen, die aus der Forderung nach Reduktion der Staatsquote hervorgehen. Die Output-Steuerung des Bildungswesens ist nichts anderes als eine Marktsteuerung: Belohnt wird, was sich als *gut* herausstellt, wobei das Kriterium der Güte nur allzu leicht in dem gesehen wird, was sich gut verkaufen lässt.

Ein interessanter Effekt des Wandels der politischen Steuerungskultur liegt darin, dass die Streitfrage, ob Schulen öffentlich oder privat organisiert werden sollen, an Bedeutung verliert. Indem der Staat seine Mittel nicht mehr von oben nach unten sickern lässt – in der Hoffnung, was

⁸ In jüngster Zeit wird in der Schweiz von der Lehrerinnen- und Lehrerbildung gefordert, dass sie bestimmten Standards zu genügen habe. Diese Standards werden in operationalisierter Form zu Wertmassstäben erhoben.

⁹ «Benchmarking» lässt sich schlecht ins Deutsche übersetzen. Ein Benchmark ist ein Vergleichsstandard zur Bestimmung der Leistungsfähigkeit eines Unternehmens oder eines Produktionsablaufs.

oben ins System eingegeben wird, werde unten fruchtbar sein –, sondern die operative Umsetzung seiner Bildungspolitik den einzelnen Schulen überlässt, spielt es letztlich keine Rolle, wer Träger dieser Schulen ist. Wenn die Einzelschule die operative Verantwortung für ihr Handeln übernehmen muss, dann kann sie sowohl öffentlich als auch privat sein. Der Staat kann die Zügel lockern und seine Mittel breiter verteilen. Er kann sich vom ideologischen Streit um die private versus öffentliche Bildung verabschieden. Wenn es um die Effektivität und die Effizienz des Mitteleinsatzes geht, dann ist er frei, auch Schulen in «freier Trägerschaft» in die Konkurrenz um seine knapper werdenden Mittel einzubeziehen.

Damit ist natürlich die grundsätzliche Frage nicht auch schon entschieden, ob nämlich eine marktförmige Organisation des Bildungswesens nicht auch Nachteile haben kann, wie z.B. die Auflösung der klassischen Idee der Allgemeinbildung. Bildung und Markt sind keineswegs Konzepte, die ohne Weiteres deckungsgleich sind.

Qualität umfasst daher nicht nur einen Gütemassstab, sondern auch Kontrolle. Kontrolle nicht im pädagogischen, sondern im politischen Sinn. Es geht um Rechenschaftslegung, die von einer Institution verlangt wird, angesichts einer sich verschlankenden Verwaltung und einem Staat, der die direkte Kontrolle der Subventionsnehmer zurücknimmt bzw. mit Globalbudgets und Leistungsvereinbarungen arbeitet, die den Leistungsbeauftragten mehr Eigeninitiative und Eigenverantwortung zugestehen.

Es ist nicht falsch, wenn der Begriff der Qualität – wie dies gelegentlich geschieht – mit dem Begriff der *Effektivität* gleichgesetzt wird (vgl. Ditton 2000, p. 74). Daneben geht es aber auch um *Effizienz*. Effektivität bezieht sich auf die Wirksamkeit einer Institution, Effizienz auf den optimalen Einsatz der Mittel zur Erreichung der Ziele. Damit ist der Begriff der Qualität auch mit demjenigen der *Evaluation* verbunden. Denn die Indizien, die Qualitätsaussagen ermöglichen, basieren auf Datenerhebung, Dokumentation und Messung, d.h. auf Evaluation (vgl. Ditton 2002; Gonon, Hügli, Landwehr, Ricka & Steiner 1999, p. 13).

Nun soll auch bei dieser Bedingung nicht der Eindruck erweckt werden, die internationalen Leistungsvergleiche seien eine *direkte* Konsequenz der Einführung von NPM. So ist es zweifellos nicht. Aber sie stehen in einem Klima, das durch NPM gefördert wird, wenn nicht sogar ausgelöst worden ist. Qualität im Sinne von Qualitätssicherung, Qualitätskontrolle und Qualitätsentwicklung ist ein Output-Kriterium, das Evaluation fordert und Messbarkeit impliziert, d.h. *Operationalisierbarkeit* von Bildung und Vergleichbarkeit von Bildungsinstitutionen bzw. deren Leistung, womit ein Ranking (wie es bei Banken üblich ist) und ein Benchmarking ermöglicht wird.

Damit sind fünf Bedingungen genannt, die gegeben sein müssen, damit internationale Schulleistungsstudien in nennenswertem Ausmass durchgeführt werden. Zweifellos sind nicht alle fünf Bedingungen von gleichem Gewicht. Trotzdem ist anzunehmen, dass die internationalen Schulleistungsvergleiche ohne diese Bedingungen nicht den Stellenwert hätten, der ihnen heute zukommt und der ihnen in naher Zukunft aufgrund der genannten politischen Konstellationen vermehrt zukommen wird.

1.2 Methodische Ansprüche an internationale Schulleistungsvergleiche

Wie schon angedeutet stellen internationale Schulleistungsvergleiche hohe methodische Ansprüche (vgl. Abschnitt 1.1.2). Im Folgenden geht es darum, einige Probleme methodischer Art, die sich bei solchen Studien stellen, etwas ausführlicher zu erörtern. Da der Erkenntniswert internationaler Schulleistungsvergleiche von deren wissenschaftlicher Qualität abhängig ist, liegt in der Art und Weise, wie die Studien durchgeführt werden, d.h. in den verwendeten *Methoden*, der entscheidende Ansatzpunkt zur Beurteilung dieser Studien. Deshalb setzen wir uns bereits in diesem Abschnitt, d.h. noch bevor wir uns den Studien in den folgenden Kapiteln *en détail* zuwenden, mit einigen Grundfragen der Durchführung internationaler Schulleistungsvergleiche auseinander.

Vielleicht ist Ihnen als Leserin oder Leser aufgefallen, dass wir in diesem Kapitel bis zu dieser Stelle vor allem von «Schulleistungsvergleichen» gesprochen und den Begriff «schulische Wirkungen» eher vermieden haben. Wir haben zwar in der Einleitung zu diesem Skript davon gesprochen, dass es thematisch um die Wirksamkeit von Schule und Unterricht gehen soll und um Studien, die beanspruchen, Aussagen über die Wirksamkeit von Schule und Unterricht zu machen. Es ist jedoch keineswegs einfach, Wirkungen von Schule und Unterricht festzustellen. Es ist gerade eine Frage der methodischen Qualität einer Studie, ob wir ihre Ergebnisse als Aussagen über schulische Wirksamkeit verstehen dürfen. Wir wollen daher die folgende Diskussion der forschungsmethodischen Probleme von Schulleistungsvergleichen unter die generelle Frage stellen, wie eine Studie angelegt werden muss, damit ihr verlässliche Aussagen über die Wirkungen von Schule und Unterricht entnommen werden können. Es könnte ja sein, dass wir mit solchen Studien irgendwelche Erkenntnisse gewinnen, nur nicht solche über die Wirkungen von Schule und Unterricht.

Die übergeordnete Frage nach den Kriterien, denen eine Untersuchung genügen muss, damit sie Aussagen über schulische Wirksamkeit machen kann, wollen wir in 11 Teilfragen ausdifferenzieren, die uns auf methodische Probleme solcher Studien hinweisen sollen. Es geht um folgende Fragen:

1. Erzeugen Schulen überhaupt Wirkungen?
2. Welcher Art sind die Wirkungen, die Schulen erzeugen?
3. Wie lassen sich die Bereiche schulischer Wirksamkeit identifizieren?
4. Lassen sich die Wirkungen von Schulen überhaupt messen?
5. Wonach bemisst sich der Erfolg einer Schule?
6. Was bedingt die Leistungen von Schülerinnen und Schülern?
7. Wenn Schulen wirksam sind, wie stellt man Wirksamkeit fest?
8. Wer soll untersucht werden?
9. Wie werden die Daten erhoben?
10. Wie werden die Daten ausgewertet?
11. Wie werden die Ergebnisse dargestellt?

Wir beginnen mit eher grundsätzlichen Fragen, um dann allmählich auf konkretere Fragen der Durchführung und Auswertung solcher Studien einzugehen.

1.2.1 Erzeugen Schulen überhaupt Wirkungen?

Wenn es um die «*Evaluation of Educational Achievement*» geht¹⁰, dann wird unterstellt, dass die Erziehung («education») tatsächlich Leistungen («achievement») erzeugt, d.h. Schule und Unterricht *Wirkungen* haben, die sich messen (evaluieren) lassen. Internationale Schulleistungsvergleiche implizieren, dass Schule und Unterricht wirksam (effektiv) sind, d.h. Effekte (welcher Art auch immer) erzeugen. Ist diese Annahme berechtigt? Die Frage hat grundsätzlichen Charakter. Sie wirft gewissermaßen die Sinnfrage solcher Studien auf: Schulen lassen sich sinnvollerweise nur dann in Bezug auf ihren Output miteinander vergleichen, wenn sie für ihren Output (kausal) verantwortlich sind. Es ist jedoch weder a priori klar, dass Erziehung und Unterricht Wirkungen

¹⁰ IEA = International Association for the Evaluation of Educational Achievement

haben, noch ist es leicht nachzuweisen, dass Schule und Unterricht den Output tatsächlich erzeugen, der ihnen zugeschrieben wird.

Die Frage nach der Wirksamkeit pädagogischer Massnahmen ist innerhalb der Pädagogik allerdings kaum kontrovers. Schon rein begrifflich impliziert Erziehung Wirksamkeit. Erziehung setzt ihrer Etymologie nach einen Zustand voraus, aus dem jemand herausgezogen, und ein Ziel, zu welchem er hinaufgezogen wird. In einer vergleichenden Untersuchung verschiedener Sprachen stellt Josef Dolch fest, die Übereinstimmung im pädagogischen Wortgebrauch liege darin beschlossen, «[...] auf etwas unter Anwendung von Kraft derart ein[zu]wirken, dass es von einer Lage oder Befindlichkeit in eine dem Einwirkenden irgendwie näher seiende andere, bessere oder richtiger, erwünschter erscheinende gebracht wird» (Dolch 1961, p. 167). Weiter meint Dolch, das einfache Ziehen habe mit dem Erziehen die Bewegung als wesentliches gemeinsam, darüber hinaus fehle aber im ersten Ausdruck («ziehen») das wichtige Moment des dauernden Erfolgs, des Verbleibs am neuen Ort oder im neuen Zustand, das für den zweiten («erziehen») zentral sei. Dementsprechend heisst es bei einem Pädagogen des 19. Jahrhunderts, beim Herbartianer Tuiskon Ziller, die Erziehung sei «[...] eine absichtliche, planmässige Einwirkung auf einen Menschen, und zwar auf den einzelnen Menschen als solchen in seiner frühesten Jugend, eine Einwirkung zu dem Zweck, dass eine bestimmte, aber zugleich bleibende geistige Gestalt dem Plane gemäss bei ihm ausgebildet wird» (Ziller 1884, p. 8 - im Original hervorgehoben).

Die Erziehung beinhaltet also nach ihrer Grundbedeutung die Idee der nachhaltigen Einwirkung auf ein Individuum, damit etwas Wertvolles aus ihm werde. Der Berner Seminarlehrer und spätere Pädagogik-Professor Hans Rudolf Rüegg schreibt in einem Text aus dem Jahre 1866, wer über die Erziehung «wissenschaftlichen Aufschluss» geben wolle, der müsse wissen, was die Erziehung aus dem Zögling machen soll. «Das Kunstwerk, welches sich unter dem erzieherischen Einfluss gestalten soll, muss in völliger Klarheit vor dem Auge unsers Geistes stehen» (Rüegg 1866, p. 14). Dementsprechend gilt es gemäss Rüegg, die Erziehung als eine «ununterbrochene Kette von Einwirkungen» zu begreifen (ebd., p. 20). Wie auf einer Reise würden wir bei der Erziehung von einer Hauptstation zur anderen gelangen, «[...] indem wir auf der sicher vorgezeichneten Richtung eine Zwischenstation nach der andern zurücklegen. Das Reiseziel ist der allgemeine Erziehungszweck, die Hauptstationen sind die besondern Zwecke, die Zwischenstationen aber bezeichnen jene einzelnen Zwecke, nach deren Verwirklichung das Werk der Erziehung in täglicher Anstrengung ringt» (ebd., p. 21).

Nicht nur die Erziehung, auch der Unterricht wird in der pädagogischen Tradition als Einwirkung verstanden. So kann man bei Johann Friedrich Herbart, einer Gründungsfigur der modernen Pädagogik, lesen, der Unterricht habe «[...] seine pädagogischen Wirkungen nicht etwa nur gelegentlich bey diesem und jenem Schüler, sondern bey jedem, so weit es dessen Eigenthümlichkeit zulässt, mit möglichster Gewissheit und Vollständigkeit zu erreichen [...]» (Herbart 1964, p. 525). In einer Rezension von Kants posthum erschienenen Vorlesungen über Pädagogik heisst es weiter, die transzendente Freiheit könne man dem Zögling nicht einflössen.¹¹ Vielmehr habe die Pädagogik zu zeigen, wie sich erzieherische Wirkungen ereignen «[...] und mit einer Gewissheit sich ereignen müssen, auf die man rechnen könne» (Herbart 1991, p. 308).

Noch vor Herbart hatte Johann Amos Comenius darüber spekuliert, wie sich die Schule zu einer mechanischen Maschine gestalten liesse, die ihre Erfolge mit berechenbarer Sicherheit erreichen würde. In seiner «Pampaedia» schrieb er: «Das ganze Schulwesen soll wie eine Kette sein, wo ein Glied das andere hält und so ein zusammenhängendes Ganzes entsteht - wie eine Uhr, in der ein Rädchen so in das andere greift, dass durch eine Bewegung das ganze Uhrwerk leicht und

¹¹«Transzendente Freiheit» meint bei Kant eine empirisch nicht nachweisbare Bedingung, unter der Menschen stehen. Da sie empirisch nicht nachweisbar ist, handelt es sich um eine rein gedankliche («transzendente») Konstruktion. Herbart schien dies eine pädagogisch unmögliche Position zu sein, da man in der Erziehung auf empirisch fassbare Gegebenheiten angewiesen sei.

harmonisch bewegt wird» (Comenius 1965, p. 135). Wenn der Unterricht früh genug einsetzt, «[...] so ist es nicht schwerer, eine beliebig grosse Schülerzahl alles zu lehren, als [...] mit Hilfe der Archimedischen Maschine Häuser, Türme und andere Lasten fortzurücken [...]» (Comenius 1993, p. 75). Indem Comenius auch den Menschen mit einem mächtigen Uhrwerk verglich, konnte er behaupten,

«[...] dass die Methode der menschlichen Bildung mechanisch sei, das heisst, dass sie alles so bestimmt vorschreibe, dass das, was so gelehrt, gelernt, getrieben wird, unmöglich ohne Ergebnis bleiben könne, ähnlich wie bei einer gut konstruierten Uhr, einem Wagen, einer Mühle oder sonst einer für die Bewegung kunstvoll angefertigten Maschine» (Comenius 1959, p. 111).

Folglich entwarf Comenius «[...] eine nach mechanischen Gesetzen konstruierte didaktische Maschine» (ebd., p. 105), die der Lehrkraft versprach, ihren Unterricht so arrangieren zu können, dass sie ihre Ziele unausweichlich erreicht.

Man mag sich wundern ob solcher Wirkungserwartungen. Die Theoretiker der Pädagogik scheinen sich die Erziehung und den Unterricht als Instrumente zu denken, die es dem Praktiker erlauben, schier uneingeschränkt wirksam zu sein. Auch ein Erziehungswissenschaftler des 20. Jahrhunderts, Helmut Fend, schreibt, mit «Unterricht und Erziehung» seien «[...] immer der Sachverhalt und der Anspruch verbunden, Heranwachsende aktiv aus einem unbefriedigenden bestehenden Zustand herauszuführen - auch und gerade wenn der unbefriedigende Zustand infolge mangelnder Lerngelegenheiten unverschuldet war - und einem normativ gesetzten Zielzustand zuzuführen» (Fend 1981, p. 377). Noch deutlicher ist Wolfgang Brezinka, der verlangt, es sei Aufgabe der wissenschaftlichen Pädagogik, Theorien zu entwickeln, «[...] die Voraussagen von folgender Art ermöglichen: wenn bestimmten Educanden gegenüber unter bestimmten Umständen in bestimmter Weise gehandelt wird, dann werden bestimmte Veränderungen in ihrer Persönlichkeit erfolgen» (Brezinka 1981, p. 13). Der Wert der Erziehung sei dahingehend zu bestimmen, «[...] dass sie tatsächlich die Ursache der gewollten Wirkung ist» (ebd., p. 211).

Wir wollen nicht fragen, inwiefern solche Erwartungen an die Mechanisierbarkeit der Erziehung zu Recht bestehen. Wir wollen auch nicht fragen, wie es die Erziehungswissenschaft anstellt, zu beweisen oder nur schon plausibel zu machen, dass ihre Wirkungserwartungen eine begründete Basis haben. Wir werden uns im 5. Kapitel etwas eingehender mit der Frage beschäftigen, wie weit Schule und Unterricht auf die Entwicklung und das Lernen von Kindern und Jugendlichen Einfluss nehmen können. Im Moment wollen wir lediglich festhalten, dass es offenbar zum Kernbestand pädagogischer Überzeugungen gehört, dass Erziehung und Unterricht von grosser Wirksamkeit sind.

1.2.2 Welcher Art sind die Wirkungen, die Schulen erzeugen?

Schulleistungsvergleiche unterstellen, dass der Output von Schule in Form von «Leistung» («achievement») erfasst werden kann. Je nachdem wie man den Begriff Leistung versteht, werden damit ausschliesslich *kognitive* Wirkungen von Schule bezeichnet. Leistung ist dann unter Umständen begrenzt auf den Bereich des (schulischen) *Wissens*. Was durch die Maschen der Leistungsvergleiche zu fallen scheint, sind Wirkungen im Bereich von *Bildung* im umfassenden Sinn von Persönlichkeit, Moral, sozialer Kompetenz etc. Gehören solche Effekte nicht zu den Leistungen von Schule? Sind sie nicht messbar? Fallen sie aus dem Sichtfeld internationaler Schulleistungsvergleiche heraus?

Die Frage muss nicht mit «nein» beantwortet werden, denn neuere Studien versuchen, die Wirkungen von Schule auf der Ebene von «Kompetenzen» zu erfassen. So ist in der PISA-Studie von *Lesekompetenz* die Rede, ebenso von Kompetenzen in Mathematik und Naturwissenschaften.

Im Titel des schweizerischen Schlussberichts von PISA 2000 ist von «Grundkompetenzen» die Rede, wobei der Begriff der Kompetenz durchaus klärungsbedürftig ist. Weinert (2001) schlägt vor, bei der Erfassung schulischer Wirksamkeit folgende Kompetenzen zu unterscheiden: (a) fachliche Kompetenzen (z.B. physikalischer, fremdsprachlicher, musikalischer Art), (b) fachübergreifende Kompetenzen (z.B. Problemlösen, Teamfähigkeit) und (c) Handlungskompetenzen, die neben kognitiven auch soziale, motivationale und moralische Aspekte umfassen (vgl. Weinert 2001a, p. 28). Statt von «fachübergreifenden Kompetenzen» ist oft auch von «Schlüsselqualifikationen», «soft skills» und «überfachlichen Kompetenzen» die Rede.

Bei diesen fachübergreifenden Kompetenzen stellt sich allerdings die Frage, inwiefern man feststellen kann, dass sie tatsächlich ein Effekt von Schule sind (vgl. auch Punkt 3). Es ist schon schwierig, im Falle von fachlichen Kompetenzen nachzuweisen, dass sie (allein) auf schulische Wirksamkeit zurückgehen, noch schwieriger ist es, diesen Nachweis im Falle von überfachlichen Kompetenzen zu erbringen. Es dürfte naheliegend sein, dass bei überfachlichen Kompetenzen auch ausserschulische Einflüsse von Bedeutung sind (z.B. der Erziehungsstil der Eltern oder der Kommunikationsstil in der Familie).

Darüber hinaus stellen sich forschungsmethodische Fragen: Wie misst man schulische Leistungen bzw. fachspezifische oder fächerübergreifende Kompetenzen? Wie misst man Handlungskompetenzen? Lassen sich in allen Fächern Kompetenzen definieren? Sind alle Wirkungen von Schule operationalisierbar bzw. messbar? Diesen Fragen gehen wir im Folgenden nach.

1.2.3 Wie lassen sich die Bereiche schulischer Wirksamkeit identifizieren?

Man kann auch anders fragen: In welchen Bereichen ist überhaupt zu erwarten, dass Schule und Unterricht wirksam sind? Oder: Wieviel Wirksamkeit ist Schulen und Unterricht zuzugestehen?

Man kann sich auf den Standpunkt stellen, überprüfen lasse sich nur, was explizit als Ziel angestrebt werde. Um es extrem zu formulieren: Man kann die Chinesischkenntnisse am Ende der obligatorischen Schulzeit nicht überprüfen, wenn in einem Schulsystem wie dem schweizerischen Chinesisch gar nicht unterrichtet wird. Schliesslich geht es um *Schulleistungen*: Evaluieren sollen die Wirkungen von *Schule* und nicht von ausserschulischen Faktoren wie zum Beispiel des Anregungsgehalts oder der Konversationskultur der Familie. Um beim Beispiel zu bleiben: Es ist durchaus möglich, dass Kinder am Ende der obligatorischen Schulzeit Chinesisch können, obwohl Chinesisch nicht unterrichtet worden ist, dann nämlich, wenn sie die Sprache von ihren Eltern gelernt haben.

Das ist in doppelter Hinsicht ein wichtiger Punkt.

Erstens muss man nachweisen können, dass die gemessenen Leistungen tatsächlich dem Einfluss der Schule zuzuschreiben sind und eben nicht der Familie oder den Medien oder den Peers. Bei den überfachlichen Kompetenzen und bei den Handlungskompetenzen ist dieser Nachweis unter Umständen ausserordentlich schwer zu erbringen. Aber selbst beim *fachlichen* Wissen ist der Einfluss der Eltern nicht auszuschliessen. Es gibt genügend Belege dafür, dass der Schulerfolg auch vom sozialen bzw. Bildungsstatus der Eltern abhängig ist.¹² Wo Schulen miteinander verglichen werden und die Schulen für die erhobenen Differenzen im Leistungsausgang verantwortlich gemacht werden, da ist darauf zu achten, dass nicht ausserschulische Faktoren für die Leistungsunterschiede verantwortlich sind. Zu diesen ausserschulischen Faktoren können auch Merkmale («Eingangsvoraussetzungen») der Schülerschaft gehören. Wenn im Einzugsbereich einer Schule Kinder aus privilegierten Familien oder besseren soziokulturellen Verhältnissen dominieren bzw. – bei tiefem Leistungsniveau – der Anteil an fremdsprachigen Kindern oder Kindern mit Verhaltensproblemen

¹² Darauf gehen wir im 5. Kapitel ausführlicher ein.

besonders gross ist, dann ist dies in Rechnung zu stellen, wenn nach den Ursachen von Leistungsunterschieden zwischen Schulen gefragt wird.

Dass es sich dabei um einen Punkt von grosser praktischer Bedeutung handelt, leuchtet unmittelbar ein. In internationalen Schulleistungsvergleichen schneiden asiatische Länder wie Japan, Korea oder Taiwan in Fächern wie Mathematik und Naturwissenschaften regelmässig gut bis sehr gut ab. Dahinter verbergen sich allerdings eher Effekte der Familie als der Schule. Denn wie andere Studien zeigen, sind die Erwartungen asiatischer Eltern gegenüber den Schulleistungen ihrer Kinder vergleichsweise hoch. Auch neigen sie dazu, dem Faktor Anstrengung (eine motivationale und beeinflussbare Bedingung) bei der Erklärung von Schulleistungen mehr Gewicht einzuräumen als dem Faktor Begabung (eine kognitive und eher unbeeinflussbare Bedingung). Folglich wäre es nicht nur problematisch, sondern geradezu fahrlässig, wenn Bildungspolitiker in westlichen Ländern die Schulleistungen dadurch verbessern wollten, indem sie Unterrichtsmethoden, wie sie an japanischen, koreanischen oder chinesischen Schulen vorherrschen, übernehmen wollten (vgl. Chen & Stevenson 1995; Stevenson & Stigler 1994).

Zweitens: Wenn international, aber – denken wir an die Schweiz – auch wenn national (interkantonal) verglichen wird, müssen die Curricula bzw. Lehrpläne der Schulsysteme der Länder bzw. Kantone, die miteinander verglichen werden, einigermaßen *homogen* bzw. *übereinstimmend* sein. Das ist im Allgemeinen bei Fächern wie Mathematik und Naturwissenschaften eher der Fall als bei Geschichte oder Literatur. Es dürfte daher kein Zufall sein, wenn in den bisherigen internationalen Vergleichsstudien genau diese Fächer sowie die Lesekompetenz im Vordergrund gestanden haben. Die «curriculare Validität» ist bei Fächern mit einem hoch standardisierten Stoffwissen eher gegeben als in den sogenannten weichen Fächern.

Um die Curricula genauer erfassen zu können, wird in der IEA-Studie nach intendiertem (angestrebtem), implementiertem (umgesetztem) und erreichtem Curriculum unterschieden. Das intendierte Curriculum ist das im Lehrplan vorgeschriebene Curriculum. Das implementierte Curriculum ist das, was von den Lehrkräften im Unterricht umgesetzt wird. Schliesslich umschreibt das erreichte Curriculum das, was die Schülerinnen und Schüler tatsächlich gelernt haben (vgl. Bos & Postlethwaite 2002, p. 247; Moser & Notter 2000, p. 53f.). Wenn man nur erfasst, was die Schülerinnen und Schüler erreicht haben, dann dürfte es schwierig sein, die erhobenen Leistungsunterschiede eindeutig der Schule zuzuschreiben, denn unter Umständen stehen im Lehrplan andere Dinge als diejenigen, die man gemessen hat. Ein bedeutsames Kriterium der methodischen Qualität internationaler und nationaler Schulleistungsvergleiche liegt demnach im Einbezug der curricularen Aspekte von Schule und Unterricht.

Wenn man wie angedeutet vorgeht, d.h. von den explizit formulierten («intendierten») Lehrzielen in den Curricula ausgeht und deren Umsetzung überprüft, dann löst man die Frage, welche Schuleffekte erwartet werden können, pragmatisch, aber nicht befriedigend. Denn es gibt auch *nichtintendierte* Wirkungen von Schule – wie jede Handlung Nebeneffekte erzeugt, die nicht beabsichtigt sind und sich schwer kontrollieren lassen. Ein Beispiel für eine solche nichtintendierte Wirkung von Schule ist die Steigerung des IQ (Intelligenzquotienten).¹³ Es gibt gute Gründe für die Annahme, dass der Schulbesuch zu einer merklichen Anhebung des Intelligenzniveaus führt (vgl. Ceci 1991). Die Intelligenz ist demnach nicht nur eine *Voraussetzung* für einen erfolgreichen Schulbesuch, sondern sie wird umgekehrt durch den Schulbesuch auch positiv beeinflusst. Dergleichen steht aber in *keinem* Curriculum, ist also keine intendierte Wirkung von Schule; und trotzdem scheint die Schule entsprechend wirksam zu sein. Analoge Beispiele lassen sich in Bezug auf die Persönlichkeit der Schülerinnen und Schüler, ihre soziale Kompetenz oder ihre Lernmotivation anführen, auch wenn man diesbezüglich schon eher Hinweise in den Lehrplänen finden kann.

¹³ Das übliche Verfahren zur Messung der Intelligenz sind Intelligenztests, die auf standardisierten bzw. genormten Häufigkeitsverteilungen beruhen. Der Intelligenzquotient wird dann in Bezug auf den Mittelwert einer solchen Häufigkeitsverteilung berechnet, wobei sich eingebürgert hat, dass der Mittelwert bei 100 liegt.

In der Bildungsforschung werden solche nicht-intendierten Wirkungen von Schule und Unterricht unter dem Stichwort «heimlicher Lehrplan» («hidden curriculum») abgehandelt. Damit bringt man zum Ausdruck, dass nicht nur der offizielle, sondern auch der inoffizielle (eben: heimliche) Lehrplan Wirkungen erzeugen kann, obwohl diese nicht intendiert sind und demnach in den Curricula auch nicht auffindbar sind. Soll man sie deshalb aus den schulischen Leistungsstudien weglassen?

Auch die Sozialisationsforschung geht oft von einer ähnlichen Problemstellung aus: Sie interessiert sich weniger für die Lehrpläne und deren Absichten als für das pädagogisch unartikulierte Geschehen in Schule und Unterricht. Was sind zum Beispiel die Auswirkungen der Beurteilung und Benotung der schulischen Leistungen? Welche Konsequenzen ergeben sich für das Selbstbild und den Selbstwert von Schülerinnen und Schülern, die gehäuft schlechte Noten erhalten? Was bedeutet es, wenn man in der Schule in erster Linie nach Leistungskriterien beurteilt wird? Solche und weitere Fragen nehmen Bezug auf Wirkungen von Schule und Unterricht, ohne dass den Wirkungen ein intendiertes Ziel zugrunde liegen würde. Die Rede ist von *Sozialisationseffekten*. Es sind gleichsam Nebenwirkungen, die von der spezifischen Kultur und vom organisatorischen Arrangement einer Schule ausgehen. Wie aber *erfasst* man solche nicht-artikulierten Wirkungen von Schule? Und wie kann man sie eindeutig als Wirkungen von *Schule* identifizieren?

1.2.4 Lassen sich die Wirkungen von Schule überhaupt messen?

Etwas moderater kann man auch fragen: *Welche* Wirkungen von Schule lassen sich messen? Es geht hier um die grundsätzliche Frage, ob sich Phänomene wie Bildung, Wissen oder Kompetenz in einem vergleichbaren Sinn messen lassen, wie dies für physikalische Gegenstände der Fall ist. Für physikalische Objekte stehen uns standardisierte Messinstrumente zur Verfügung wie das Metermass, die Uhrzeit oder ein Thermometer. Im Falle von psychischen Phänomenen wie schulischen Leistungen oder Merkmalen der Persönlichkeit stellt sich zumindest die Frage, ob wir über vergleichbare Messinstrumente verfügen. Die Frage verweist auf die sozialwissenschaftliche Messtheorie. Wie bereits erwähnt, sind in internationalen Vergleichsstudien bisher vor allem Leistungen in Mathematik, Naturwissenschaften und Lesen untersucht worden. Ist das Zufall? Lassen sich andere schulische Wirkungen, obwohl es sie gibt, vielleicht gar nicht messen?

1.2.5 Wonach bemisst sich der Erfolg einer Schule?

Eine weitere Frage betrifft die *Art* der erwarteten Effekte. Es mag trivial scheinen, wenn das *Niveau* (Leistungs- oder Kompetenzniveau) als wichtiges Kriterium für Schuleffekte erachtet wird: Je höher das Leistungsniveau, das Schülerinnen und Schüler erreichen, desto wirksamer die Schule. Tatsächlich ist dies die dominierende Sichtweise, in der die internationalen Schulleistungsvergleiche durchgeführt werden. Das zeigen auch die Ranglisten, die von den beteiligten Ländern gemacht werden (vgl. Kapitel 2 bis 4).

Das *Leistungsniveau* ist jedoch nicht das einzige Erfolgskriterium. Man kann sich auf den Standpunkt stellen, dass ein Schulsystem dann von besonders guter Qualität ist, wenn es einen möglichst *homogenen* Output erzeugt, wenn m.a.W. möglichst alle Schülerinnen und Schüler *dasselbe* Bildungsniveau erreichen. Kriterium ist in diesem Fall nicht das Niveau, sondern die *Varianz* der Leistungen: Nicht das (durchschnittliche) *Niveau* der Leistungen zählt, sondern deren *geringe Streuung*. Bei den internationalen Schulleistungsvergleichen ist dieser Ansatz allerdings nicht anzutreffen.

Da es sich bei der Unterscheidung von Niveau (Mittelwert) und Streuung (Varianz) um ein wichtiges Kriterium der Beurteilung von Schulleistungsstudien handelt, wollen wir es etwas

vertiefen. Es ist sehr wohl möglich, dass Schule und Unterricht das Lernen bzw. den Erwerb von Wissen und Können merklich beeinflussen, diesbezüglich also hohe Wirksamkeit haben, die individuellen Unterschiede im Leistungsvermögen aber nicht bedeutsam verändern. Es wäre wie in einem Fahrstuhl: Nach dem Besuch des Kindergartens treten die Schülerinnen und Schüler in den Aufzug Schule ein, fahren kollektiv die Etagen der Jahrgangsstufen hoch und steigen – so verschieden wie sie eingestiegen sind – in der neunten Klasse wieder aus. Sie sind als Kollektiv vorangekommen – der Mittelwert ihres Leistungsniveaus hat sich angehoben –, gemessen an den individuellen Differenzen ihres Leistungsvermögens sind sie jedoch stehen geblieben. Das Gleichnis zeigt, dass wir die Güte einer wirksamen Schule zumindest an zwei Massstäben messen können: am Leistungsniveau und an den Leistungsdifferenzen. Die beiden Kriterien sind unabhängig voneinander und können deshalb auch kumuliert werden: Eine besonders gute (wirksame) Schule wäre demnach eine, die nicht nur das Leistungsniveau der Schülerinnen und Schüler anhebt, sondern auch die Leistungsdifferenzen zwischen den Schülerinnen und Schülern abbaut. Allerdings zeigt die Forschung, dass es nicht leicht ist, beiden Kriterien gleichzeitig und gleichermaßen gerecht zu werden.

Die Frage nach den Kriterien, die bei der Beurteilung der Wirksamkeit von Schule beizuziehen sind, tangiert *ethische* Probleme. Was verstehen wir unter einer gerechten Schule? Wann ist der Anspruch auf Gleichheit der Bildungschancen erfüllt? Ist es Aufgabe der Schule, die intellektuellen und motivationalen Unterschiede, mit denen Kinder in die Schule eintreten, auszugleichen? Ist eine Schule dann gut, wenn sie ihren «Output» im Sinne von Bloom gleich macht?

Um auf die letzte Frage kurz einzugehen, so ist in Rechnung zu stellen, dass genetische, konstitutionelle, familiäre und andere ausserschulische Einflüsse eine Heterogenität der Schülerschaft erzeugen, die durch noch so effiziente schulische Massnahmen kaum zu reduzieren ist. Anders formuliert: Unabhängig von jedem schulischen Einfluss sind Kinder *Individuen* und unterscheiden sich in ihrer Individualität in einem so hohen Ausmass, dass nicht erwartet werden kann, dass Schule und Unterricht diese Unterschiede wesentlich zu schmälern oder gar auszumerzen vermögen. Eher ist vom Gegenteil auszugehen, dass nämlich eine gute Schule und ein effizienter Unterricht dazu beitragen, die Heterogenität bzw. Individualität der Schülerinnen und Schüler zu verstärken. Ist es nicht ein Hauptziel pädagogischen Handelns, jede und jeden nach ihren bzw. seinen Möglichkeiten (Begabungen) zu fördern und zu sich selbst zu verhelfen? Gut wären demnach Schule und Unterricht dann, wenn ihr «Output» eine maximale Streuung aufweist!¹⁴

Nun ist der Anspruch auf Gleichheit damit nicht erledigt, denn ethische Fragen werden in Bezug auf die Schule nicht zuletzt auf das Kriterium der *sozialen* Gleichheit bezogen. Wenn sich schulische Leistungen einerseits im Niveau, das (durchschnittlich) erreicht wird (erstes Kriterium), und andererseits im Ausmass der Streuung (Varianz) der erreichten Werte (zweites Kriterium) niederschlagen, dann liesse sich eine Art Kompromiss zwischen den beiden Kriterien dahingehend formulieren, dass ein Schulsystem dann effektiv bzw. gerecht ist, wenn die erzielten Leistungen zwar streuen, vielleicht sogar maximal streuen, die Unterschiede zwischen den Schülerinnen und Schülern aber nicht durch *Kriterien der sozialen Herkunft* (wie soziale Schicht, Bildungsniveau der Eltern, Nationalität, Rasse etc.) bedingt sind. Massstab der Wirksamkeit des Schulsystems wäre dann der *Abbau von ungleichen Bildungschancen*. In der Tat ist dies ein Kriterium, das oft beigezogen wird. So ist in den Studien, die von James Coleman und Christopher Jencks in den 1960er und 1970er Jahren in den USA durchgeführt wurden, nicht einfach nach der Leistungsfähigkeit amerikanischer Schulen gefragt worden, sondern nach deren Vermögen, Ungleichheiten der

¹⁴ Empirische Studien zeigen regelmässig, dass die Varianz der Schulleistungen, wie sie zwischen Schulen besteht, kleiner ist als die Varianz innerhalb einer Schule. Im ersten Fall liegt sie bei rd. 30%, im zweiten bei rd. 70% (vgl. Fend 1981, p. 265). Weiterführendes zu den Begriffen Varianz und Varianzaufklärung findet sich unter Punkt 1.2.7 in diesem Abschnitt.

Bildungschancen von Kindern aus niederen und höheren sozialen Schichten bzw. weisser und schwarzer Hautfarbe auszugleichen.

Es gibt weitere Kriterien, die zur Beurteilung der Qualität des schulischen Outputs beigezogen werden können. Ein Schulsystem kann zum Beispiel danach beurteilt werden, wie gross der Prozentsatz einer Alterskohorte ist, die noch zur Schule geht, eine weiterführende Schule besucht oder einen Maturitätsabschluss macht. Was das letztere Kriterium anbelangt, so ist in der Schweiz die Chance, ein Maturitätszeugnis zu erwerben, in den verschiedenen Kantonen sehr unterschiedlich. Während im Jahr 2008 im Kanton St. Gallen knapp 14% der 19-Jährigen einen Maturitätsabschluss machten, waren es im Kanton Genf fast 30% (vgl. Schweizerische Eidgenossenschaft 2010). Was sagen diese Zahlen über die *Effektivität* schweizerischer Gymnasien? Was sagen sie über die *Gerechtigkeit* des schweizerischen Schulsystems?

Ein anderes Kriterium bildet der Zuwachs in der Zeit. Wir sind bisher nicht darauf eingegangen, dass Erziehung und Unterricht *zeitliche* Phänomene sind. Wenn wir jedoch in Rechnung stellen, dass es Zeit braucht, um eine Schule (erfolgreich) zu besuchen, dann lässt sich als Kriterium der Wirksamkeit von Schule auch die *Dauer* einsetzen, die eine Schule braucht, um ein bestimmtes Ziel zu erreichen. Eine Schule, die in kürzerer Zeit bzw. mit weniger Lektionen dasselbe Ziel erreicht, für das eine andere Schule mehr Zeit benötigt, ist dann eine wirksamere Schule. Statt der Dauer liesse sich auch direkt der *Lernzuwachs* als Kriterium verwenden. In der Tat werden Schulleistungsvergleiche gelegentlich dahingehend kritisiert, die ausschlaggebende Differenz zwischen effizienten und ineffizienten Schulen liege nicht im Leistungsniveau als solchem, sondern in der *Niveausteigerung* über eine festgelegte Zeit.

Eine wichtige Unterscheidung in zeitlicher Hinsicht stellt die Frage dar, ob Leistungseffekte innerhalb der Schulzeit oder nach deren Abschluss erhoben werden. Im zweiten Fall wird die Schule im Hinblick auf ihre Instrumentalität für ausserschulische Kriterien überprüft: den Besuch einer postobligatorischen Bildungsinstitution (Gymnasium oder Berufslehre), den Berufserfolg, den sozioökonomischen Status, die Lebensfreude etc. Gelegentlich werden diese mittelbaren Konsequenzen von Schule als Outcome bezeichnet, im Unterschied zum Output als dem unmittelbaren Schuleffekt (vgl. Abschnitt 5.6).

Weitere mögliche Kriterien zur Bestimmung der (unmittelbaren bzw. mittelbaren) schulischen Wirksamkeit sind: die Repetentenquote in einem Schulsystem, die Ausbildung von überdauernden Lernhaltungen bei Schülerinnen und Schülern, die Integration von leistungsschwachen, geistig oder körperlich behinderten, fremdsprachigen sowie hochbegabten Schülerinnen und Schülern in Regelklassen, die Weiterbildungsbereitschaft der Bevölkerung, die Reduktion der Analphabetenquote in einem Land, das Ausmass an Illetrismus etc.

Zur Bestimmung der Kriterien schulischer Wirksamkeit kann man auch nach dem Auftrag von Schule fragen. Hat die Schule lediglich einen Unterrichtsauftrag? Oder auch einen Erziehungsauftrag? Soll sie nur Wissen vermitteln? Oder auch Werte, Überzeugungen und Einstellungen? Soll sie die Persönlichkeit bilden? Hat sie nur kognitive Ziele? Oder auch affektive, motivationale und motorische? Geht es ihr nur um Sachkompetenzen? Oder auch um Sozial- und Selbstkompetenz? Soll die Schule die Stoffhuberei oder das kritische Denken fördern? Ist sie um ihrer selbst willen da oder um «aufs Leben vorzubereiten»? etc.

Die Frage der Wirksamkeit von Schule ist zurückgebunden an die Frage der Ziele, die in einem Bildungssystem erreicht werden sollen. Dabei geht es auch um politische Haltungen. Grundsätzlich ist davon auszugehen, dass an die Schule heterogene Ansprüche gestellt werden, so dass es verfehlt wäre, die schulische Wirksamkeit lediglich an *einem* Kriterium zu messen. Es ist auch unwahrscheinlich, dass die Bedingungen einer wirksamen Schule für jede Art von Zielkriterium gleich sind. Vielmehr müssen wir davon ausgehen, dass es «die» wirksame Schule und «den» effektiven

Unterricht allein schon deshalb nicht gibt, weil verschiedene Zielkriterien auf unterschiedliche Weise erreicht werden.¹⁵

1.2.6 Was bedingt die Leistungen von Schülerinnen und Schülern?

Diese Frage bildet gleichsam die notwendige Ergänzung zur Frage nach den Wirkungen: Wenn es Effekte gibt, was sind dann deren Ursachen? Welches sind die Ursachen der Wirksamkeit von Schule und Unterricht? Sind es die Lehrerinnen und Lehrer? Sind es die Schülerinnen und Schüler? Ist es die Qualität des Unterrichts? Ist es die Organisation der Schule?

Doch wie wir bereits unter Punkt 1.2.3 gesehen haben, stellt sich eine weitere Frage: Verdanken sich die Leistungen der Schülerinnen und Schüler überhaupt der Schule und dem Unterricht? Oder sind es außerschulische Faktoren (wie die Eltern, die Geschwister, die Peers, die Medien), die dafür verantwortlich sind?

Es wäre naiv anzunehmen, Leistungen von Schülerinnen und Schülern würden allein durch Schule und Unterricht bedingt. Ebenso abwegig wäre die Annahme, Schule und Unterricht seien ohne jede Bedeutung für schulische Leistungen, alles komme auf die außerschulischen Bedingungen an. Die Wahrheit dürfte irgendwo dazwischen liegen. Aber wo genau «dazwischen»? Die Frage zeigt, wie schwierig die Analyse der Effektivität von Schulen ist. Denn wie lassen sich die relativen Anteile schulischer und außerschulischer Bedingungen für die Leistungen von Kindern und Jugendlichen erfassen? Wie muss man vorgehen, um aus dem Gesamtspektrum der *möglichen* Einflussfaktoren jene herauszuschälen, die *wirklich* von Bedeutung sind?

Nicht nur ist es keineswegs trivial, intendierte und nicht-intendierte Effekte von Schule zu unterscheiden (vgl. Punkt 3), es ist auch alles andere als einfach, die (intendierten oder nicht-intendierten) *schulischen* Effekte von den *außerschulischen* zu differenzieren und in ihrem relativen Anteil zu bestimmen. Und es ist nochmals alles andere als leicht, bei den schulischen Effekten den Anteil der Lehrkräfte vom Anteil der Schülerinnen und Schüler zu unterscheiden oder die Einflüsse, die auf der Unterrichtsebene wirken, von denjenigen abzugrenzen, die von der Organisationsebene der Schule ausgehen.

1.2.7 Wenn Schulen wirksam sind, wie stellt man Wirksamkeit fest?

Bei dieser Frage geht es um die Untersuchungsanlage bzw. das *Design* von schulischen Leistungsvergleichen. Wie müssen die potentiellen Bedingungen der Leistungen von Schülerinnen und Schülern untersucht werden, damit wir verlässliche Aussagen über den Einfluss der einzelnen Faktoren gewinnen? Was sind die methodischen Voraussetzungen, um Wirksamkeit aufzudecken?

Wenn wir Wirkungen auf ihre Ursachen zurückführen wollten, müssten wir unser Forschungsobjekt über eine gewisse Zeit hinweg beobachten, denn Kausalität erkennen wir anhand der *Folgen*, die sich einstellen, wenn wir eine Bedingung so und nicht anders setzen. Wirkungen stellen sich *im Verlaufe der Zeit* ein.

Damit stellt sich eine grundsätzliche Problematik internationaler Vergleichsstudien. Denn dabei haben wir es fast ausnahmslos mit Survey-Forschung («Umfrageforschung») zu tun, d.h. mit sogenannten *Querschnittstudien*. Es sind keine Längsschnittstudien, die allererst Vergleiche über die Zeit hinweg erlauben würden. Und es sind keine experimentellen Studien, die die aktive Manipulation der Bedingungen erlauben würden, aus denen die vermuteten Wirkungen hervorgehen. Wenn daher in internationalen Leistungsvergleichen von «Wirkungen» oder «Kausalität» oder

¹⁵ Auf diesen zuletzt genannten Punkt werden wir im 5. Kapitel etwas näher eingehen.

auch nur von «Bedingungen» die Rede ist, dann handelt es sich immer um *relative* Aussagen. Eigentlich können wir nur sagen, dieses Schulsystem, diese Schulorganisation, diese Unterrichtsmethode oder dieses Lehrerverhalten ist wirksamer als eines bzw. eine der anderen, die wir untersucht haben. Nie können wir sagen, dieser Faktor ist *per se* kausal verantwortlich für diesen Effekt. Streng genommen können wir nicht einmal diese Art von Aussage machen, denn über die *Richtung*, in der die Bedingungen wirken, kann aufgrund des Querschnittscharakters dieser Studien kein Urteil gefällt werden.

Das hat auch mit dem statistischen Charakter der Daten zu tun, die in diesen Untersuchungen erhoben werden. Wir wollen diesen Punkt etwas vertiefen.

Die statistischen Verfahren, die in internationalen Vergleichsstudien zum Zuge kommen, lassen sich unter dem Begriff der «Varianzaufklärung» zusammenfassen. «Varianz» bedeutet, dass ein Merkmal – im Falle von Schulleistungen: das Merkmal schulische Leistung – in verschiedener Ausprägung vorkommt. Beim Beispiel Lesekompetenz oder Mathematikleistung gibt es eine Streuung der Werte von tiefen Leistungen bis zu hohen Leistungen. Auch die in der Schule übliche Notenskala ist ein solches Mass für die Ausprägung von Leistungen. Die Note 1 bedeutet den tiefsten, die Note 6 den höchsten Wert, den die Leistungen in einem Fach erreichen können. Wenn die Notenwerte in einer Prüfung zwischen 1 und 6 streuen, dann ist ihre Varianz entsprechend gross. Wenn sie lediglich zwischen 4 und 6 streuen, ist sie vergleichsweise gering. Wobei das Notenbeispiel zeigt, dass Varianz bzw. Streuung immer *mehrere* Messungen voraussetzt: entweder mehrere Messungen *bei ein und demselben Schüler* oder mehrere Messungen *bei verschiedenen Schülerinnen und Schülern*.

Die Varianz ist also ein Kennwert, um das Ausmass der Streuung eines (metrischen) Kriteriums relativ zu einer Gruppe zu kennzeichnen.¹⁶ Je grösser die Gruppe, deren Leistungen erfasst werden, desto eher ist mit einer grösseren Streubreite (Varianz) der Werte zu rechnen.

»Varianzaufklärung« heisst nun, dass man versucht, bedingende Faktoren ausfindig zu machen, die für das Ausmass der Streuung verantwortlich sind. Solche Faktoren können im genannten Beispiel sein: Intelligenz, Interesse am Fach, Lernfreude, Anzahl Unterrichtsstunden, Schultyp (zum Beispiel Real- versus Sekundarschule), Unterstützung durch die Eltern etc. So könnte sich herausstellen, dass intelligentere Schülerinnen und Schüler *bessere* Noten in Mathematik haben, weniger intelligente *schlechtere*. Dann würde man sagen, der Faktor Intelligenz «erklärt» die Unterschiede in den Mathematikleistungen. Man sagt auch, die Intelligenz sei ein «Prädiktor» für das Leistungsvermögen in Mathematik. In einem allgemeineren Sinn nennt man die Intelligenz auch eine «unabhängige» Variable und die Leistung die «abhängige» Variable. Allerdings wird es selten gelingen, schulische Leistungen auf einen *einzigsten* Prädiktor zurückzuführen. Im allgemeinen spielen mehrere Einflüsse eine Rolle, und es ist Aufgabe der Forschung, Art und Anzahl der determinierenden Faktoren ausfindig zu machen bzw. deren Zusammenwirken aufzudecken.

Streng genommen ist die Rede von der Varianzaufklärung durch Prädiktoren nicht präzise genug. Denn Prädiktoren können nur dann genutzt werden, um die Varianz von schulischer Leistung oder einem anderen Zielkriterium aufzuklären, *wenn sie ihrerseits variieren*. So ist es unmöglich, den Faktor Intelligenz zur Vorhersage von Mathematikleistungen zu gebrauchen, wenn alle Schülerinnen und Schüler einer Klasse oder Untersuchungsgruppe *dieselbe* Intelligenz aufweisen. Nur wenn sich die Schülerinnen und Schüler im Ausmass ihrer Intelligenz unterscheiden, und dies wiederum in einem nennenswerten Ausmass¹⁷, kann die Intelligenz – aufgrund der Logik der dabei

¹⁶ «Metrisch» meint, dass das Merkmal abgestuft vorliegt und zwischen den Abstufungen gleiche Intervalle bestehen. Diesen beiden Bedingungen genügen Intervallskalen und sog. Rationalskalen wie das Metermass oder die Kelvin-Temperaturskala. Wie das Beispiel der Notenskala zeigt, kann es umstritten sein, ob bei sozialwissenschaftlichen Messungen die Bedingung der Gleichheit der Intervalle tatsächlich erfüllt ist. Ist der Abstand zwischen den Noten 3 und 4 derselbe wie zwischen den Noten 5 und 6?

¹⁷ Aus einer Prädiktorvariable mit nur wenig Streuung kann kein grosser Effekt (z.B. Korrelationswert) hervorgehen.

verwendeten statistischen Verfahren – für die Aufklärung bzw. Prädiktion der Mathematikleistungen beigezogen werden. In sprachlicher Hinsicht wäre deshalb die präzise Ausdrucksweise die, dass die *Unterschiede* in einer Variable A (der Prädiktorvariable; im Beispiel: die Unterschiede in der Intelligenz) die *Unterschiede* in einer anderen Variable B (der Kriteriumsvariable; im Beispiel: die Unterschiede in den Mathematikleistungen) erklären. Da es nicht gerade eingängig ist, so zu reden, wird oft die saloppe Redeweise bevorzugt, wonach Variable A Variable B erkläre.

Das Beispiel zeigt, dass in Fällen, in denen es uns misslingt, die Varianz einer abhängigen Variable durch einen bestimmten Faktor aufzuklären, nicht gefolgert werden darf, dass der Prädiktor *ohne* kausale Bedeutung für die betreffende Variable ist. Um beim Beispiel zu bleiben: Auch in einer Gruppe von Schülerinnen und Schülern, die alle einen Intelligenzquotienten von 120 (ein relativ hoher Wert) haben, dürften sich die Mathematikleistungen unterscheiden. Wir haben m.a.W. eine Varianz, die wir aufklären können. Dabei greifen wir vielleicht auf die Faktoren Interesse am Fach, Lernfreude und Leistungserwartungen der Eltern zurück. Wenn wir mit diesen drei Faktoren die Varianz der Mathematikleistungen in dieser Gruppe *gleich intelligenter* Schülerinnen und Schüler zu erklären vermögen, dann heisst dies nicht, dass die Intelligenz kausal ohne Bedeutung ist. Es dürfte intuitiv plausibel sein, dass schulische Leistungen (gerade in einem Fach wie Mathematik) vom Intelligenzniveau beeinflusst werden. Wo wir es mit gleich intelligenten Schülerinnen und Schülern zu tun haben, wird dieser Einfluss aber nicht sichtbar, weil die fehlende Streuung der Intelligenzwerte dazu führt, dass die Intelligenz vom statistischen Analyseverfahren gleichsam übersehen wird.

Daraus ergibt sich eine wichtige Folgerung in Bezug auf die Aufdeckung von Kausalprozessen im Rahmen von internationalen Schulvergleichen. Damit der Anteil von Schule und Unterricht im Kausalgefüge schulischer Leistungen überhaupt «sichtbar» werden kann, bedürfen die Faktoren Schule bzw. Unterricht einer gewissen Varianz. Nur wenn wir – analog zur Intelligenz im obigen Beispiel – unterschiedliche Schulen bzw. unterschiedliche Unterrichtsformen haben, dürfen wir damit rechnen, dass Schule bzw. Unterricht beim Versuch der Erklärung der Leistungsvarianz überhaupt ins Gewicht fallen. Ein völlig homogenes Schulsystem liesse den Anteil von Schule und Unterricht am Schulerfolg der Schülerinnen und Schüler gar nicht erkennen – jedenfalls solange nicht, wie die Daten mittels statistischer Verfahren der «Varianzaufklärung» ausgewertet werden.

Nun gibt es gute Gründe anzunehmen, dass in westlichen Gesellschaften die Unterschiede zwischen den Schulen weit geringer sind als in Ländern der Dritten Welt. Dies aus dem einfachen Grund, weil wir es mit nationalen Schulsystemen zu tun haben, die weitgehend *einheitlich* strukturiert sind (zum Beispiel was die Zuweisung von Mitteln, die Lehrpläne, die Dauer des Schulbesuchs, die Klassengrößen, die Qualität der Lehrerinnen- und Lehrerbildung, die Schulaufsicht oder die Qualität der Lehrmittel anbelangt etc.). Unter diesen Umständen ist zu erwarten, dass ausserschulische Faktoren und individuelle Bedingungen auf Seiten der Schülerinnen und Schüler weit mehr streuen als die schulischen bzw. unterrichtlichen Faktoren. Die Varianz der schulischen Einflüsse dürfte m.a.W. kleiner sein als diejenige der familiären, weil die Differenz zwischen der schlechtesten und der besten Schule geringer ist als diejenige zwischen der «schlechtesten» und der «besten» Familie. Dass dem tatsächlich so ist, zeigt Abbildung 1.1.

Influences on Primary School Science Achievement

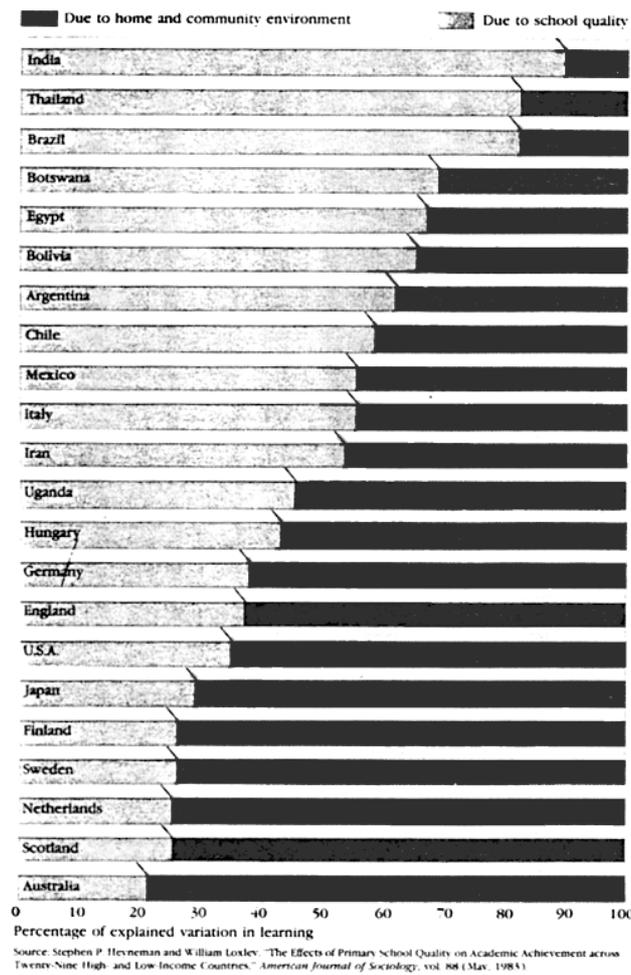


Abbildung 1.1: Familiäre und schulische Bedingungen naturwissenschaftlicher Leistungen in der Primarschule (aus: Fuller & Heyneman 1989, p. 16)

Wenn Studien wie diejenigen von James Coleman (1966) oder Christopher Jencks (1972), die auf Daten der USA basieren, zum Ergebnis kommen, dass die Schule wenig Bedeutung hat¹⁸, dann ist dies möglicherweise ein Effekt der geringen Varianz der untersuchten Schulen.¹⁹ Daraus schliessen zu wollen, die Schule sei praktisch *ohne Bedeutung* für die Erklärung der Leistungsunterschiede zwischen den Schülerinnen und Schülern, wäre demnach genauso absurd, wie wenn man beim Beispiel der Schülerinnen und Schüler mit gleicher Intelligenz schliessen wollte, die Intelligenz sei für die Mathematikleistungen ohne Bedeutung. Aufgrund einer geringen Varianzaufklärung kann nicht zwingend auf Wirkungslosigkeit geschlossen werden. Ein gutes, egalitäres Schulsystem kann hochwirksam sein, was aber nicht sichtbar wird, wenn man mit statistischen Verfahren arbeitet, deren Logik streuende Messwerte voraussetzen.

¹⁸ Wir gehen auf diese beiden Studien im 5. Kapitel kurz ein.

¹⁹ Damit wird nicht gesagt, dass das amerikanische Schulsystem besonders homogen ist, jedoch ist anzunehmen, dass die familiären Verhältnisse, aus denen die Schulkinder in den USA kommen, heterogener sind als die schulischen Verhältnisse. Die Varianz zwischen den Familien ist m.a.W. grösser als die Varianz zwischen den Schulen. Dazu kommt, dass das Zielkriterium, das Jencks ansetzt, im Abbau von sozialer Ungleichheit liegt, was aber nicht das einzige Kriterium ist, das man beiziehen kann, um schulische Wirksamkeit zu beurteilen (vgl. Punkt 1.2.5).

Wenn wir auf die Frage nach der Aufdeckung der Prozesse, die schulische Wirksamkeit bedingen, zurückkommen, dann zeigen diese Ausführungen zur Logik der «Varianzaufklärung», dass internationale Leistungsvergleiche kaum dafür ausgerüstet sind, im echten Sinne Kausalanalysen vorzunehmen. Fast ausnahmslos haben wir es mit Daten zu tun, die aus Querschnittstudien stammen und demnach bestenfalls Vermutungen zulassen, wie und in welcher Richtung die Kausalwirkungen laufen. Darin liegt eine wesentliche Grenze der Aussagekraft dieser Studien, die zu bedenken ist, wenn allzu leichtfertig Schlüsse über die Ursachen der vergleichsweise schlechten Leistungen unseres Schulsystems gezogen werden (vgl. die Medienzitate in der Einleitung). Letztlich müssen wir Mislevy (1995) zustimmen: «We cannot infer the causes of achievement from survey data such as those gathered in international assessment» (Mislevy 1995, p. 428).

1.2.8 Wer soll untersucht werden?

Auch bei dieser Frage geht es um ein Problem der Untersuchungsanlage. Welche Daten muss man erheben, damit Systemvergleiche möglich sind? Zu klären ist vor allem die Frage der *Stichprobenbildung*. Wie wählt man die Schülerinnen und Schüler aus, die untersucht werden? Wählt man sie individuell oder klassenweise aus? Wählt man sie nach dem Alter oder nach der Klassenstufe aus? Die Klassenstufe ist zweifellos der wichtigere Faktor, wenn man die Effektivität von Schule und Unterricht erfassen will, denn die Dauer des Schulbesuchs ist von massgeblichem Einfluss auf den schulischen Output. Wobei für die Bestimmung der Dauer möglicherweise weniger die Anzahl Schuljahre von Bedeutung ist, als die Anzahl *Lektionen*, die ein Fach unterrichtet worden ist. Das Alter stellt demgegenüber das angemessenere Selektionskriterium dar, wenn man entwicklungspsychologische Bedingungen berücksichtigen will oder ganz einfach wissen will, was Jugendliche in einem bestimmten Alter wissen und können.

Auch Fragen wie diejenigen nach der Stichprobengrösse, nach allfälligen Substichproben (z.B. Mädchen vs. Jungen; einheimische vs. ausländische Schülerinnen und Schüler; verschiedene Schultypen; regionale Stichproben) gehören hierher. Besondere Beachtung ist der unterschiedlichen Selektivität von Schulsystemen zu schenken. Während beispielsweise die *High-School* in den USA kaum selektiv ist und auch das Modell der Gesamtschule in Deutschland keine äussere Differenzierung der Schülerpopulation vorsieht, ist das dreigliedrige System der Oberstufe, das die meisten Kantone der Schweiz kennen (Realschule – Sekundarschule – Gymnasium) selektiv, d.h. es finden sich nicht alle Schülerinnen und Schüler *ein und derselben Jahrgangsstufe* im gleichen Schultyp. Wird nun eine internationale Untersuchung auf dem Sekundarniveau durchgeführt, ist die in verschiedenen Ländern unterschiedliche Schulformgliederung zu berücksichtigen, da sonst nicht-vergleichbare Stichproben gebildet würden. Während in Ländern wie den USA, England und Schweden einfache Zufallsstichproben gebildet werden können, muss in Staaten mit gegliederten Systemen eine *stratifizierte* Stichprobe gezogen werden (damit ist eine schultypenbezogene Aufteilung der Stichprobe mit entsprechender Gewichtung der statistischen Kennwerte gemeint). Zu berücksichtigen ist auch, wenn Schulsysteme Sonderschulen bzw. unterschiedliche Formen der externen Sonderbeschulung aufweisen, da sonst Schülerinnen und Schüler entweder nicht berücksichtigt oder entsprechend ihrem Anteil an der Gesamtschülerschaft unter- oder überrepräsentiert werden.

1.2.9 Wie werden die Daten erhoben?

Nachdem eine Stichprobe gebildet worden ist, stellt sich die Frage der Datenerhebung. Dabei werden Probleme tangiert, die in der Forschungsmethodik unter Begriffen wie Operationalisierung, Bildung von Skalen und Indikatoren sowie Methoden der Datenerhebung abgehandelt werden. Es geht um die Frage, welches Verfahren verwendet wird, um die Leistungen der Schülerinnen und Schüler zu messen: Werden die *Noten* erfasst, die von den Lehrkräften gesetzt wurden? Werden standardisierte *Wissenstests* eingesetzt? Werden eigene Erhebungsinstrumente entwickelt? Es geht um Fragen der Objektivität, Reliabilität und Validität von Testaufgaben. Entsprechen die Testfragen dem vermittelten Schulstoff («curriculare Validität»)? Es geht um die Art und Weise, wie die Daten erhoben werden (z.B. mittels schriftlicher Befragung oder mündlicher Interviews, durch Beobachtung, klassenweise oder individuell, vor Ort oder postalisch etc.).

Wie wir schon erwähnt haben, handelt es sich bei internationalen Vergleichsstudien fast ausnahmslos um sogenannte Survey-Forschung, die mit vorgegebenen (standardisierten) Testaufgaben arbeitet und verbale Antworten verlangt («Papier-und-Bleistift-Tests»). Dabei stellt sich die Frage, ob das Format der Testaufgaben der intendierten Messung angemessen ist. So dürften zum Beispiel Multiple-Choice-Aufgaben zur Messung des Sprachvermögens nur bedingt angemessen sein. Die Verwendung schriftlicher Items ist nicht adäquat, wenn die mündliche Sprachkompetenz erfasst werden soll. Des Weiteren stellt sich die Frage, inwiefern die Items der Alltags- bzw. Schulerfahrung der Probandinnen und Probanden entsprechen. Wie vertraut sind die Schülerinnen und Schüler beispielsweise mit «eingekleideten» Aufgaben? Was heisst es, wenn in einem Bildungssystem, in dem Prüfungen normalerweise nicht auf der Basis von Multiple-Choice-Aufgaben durchgeführt werden, die Daten mittels genau dieser Art von Tests erhoben werden?

Mislevy (1995) verweist auf zwei weitere Probleme der Datenerhebung:

(a) **Motivation:** 1991 sind in Korea jene Schüler, die im Rahmen einer Zufallsauswahl in die Erhebungsstichprobe aufgenommen wurden, von der Schuladministration als «Champions» bezeichnet worden. Normalerweise wird eine solche Selektion als das genommen, was sie ist: eine Zufallsauswahl. Die Motivation zur Teilnahme an der Erhebung dürfte jedoch anders ausfallen, je nachdem ob man sich als Champion oder Zufallstreffer fühlt (vgl. Mislevy 1995).

Dazu kommt, dass die bloße Teilnahme an einer Erhebung unter Umständen wenig motiviert erfolgt, weil man – anders als sonst bei Prüfungen – keine promotionsrelevante Benotung erhält. Wie Helmke (2000) bemerkt, ist das Motivationsproblem bei einigen Studien zwar aufgedeckt worden, jedoch sind keine Patentlösungen in Sicht (vgl. Helmke 2000, p. 144).

(b) **Antwortstile:** Mislevy schreibt zu diesem Punkt:

»Previous IEA studies had revealed interesting differences in the amount of guessing among countries. In the mathematics study, for example, students in the United States appeared to engage in a large amount of guessing; i.e., a substantial number of questions were incorrectly answered and relatively few were omitted. In contrast, Belgian students appeared to engage in very little guessing; that is, relatively few items were answered incorrectly but a substantial number of items were omitted. Belgian students, it was learned, are taught from the beginning of their school careers not to attempt to answer a question unless they are almost certain they know the answer. This is not true in the United States, however, where guessing is often encouraged.» (Mislevy 1995, p. 423).

Ein besonderes Problem stellt die Konstruktion von Testitems dar. Das übliche Verfahren der psychometrischen Testkonstruktion ist auf die Maximierung von individuellen Differenzen ausgerichtet. Das zeigen Intelligenztests, die feststellen wollen, inwiefern Menschen in ihrem intelligenten Potential *verschieden* sind. Deshalb verwenden sie Testaufgaben (Items), die

«trennscharf» sind, d.h. zwischen den Probandinnen und Probanden *unterscheiden*.²⁰ Nicht die absolute Leistung bildet den Massstab, sondern die relative Position des einzelnen zur Leistungsposition der anderen. Damit werden möglicherweise die *egalisierenden* Wirkungen von Schule und Unterricht unterschätzt.

1.2.10 Wie werden die Daten ausgewertet?

Grundsätzlich gilt, dass die Ergebnisse von sozial- und erziehungswissenschaftlichen Untersuchungen und damit auch von internationalen Vergleichsstudien nicht unabhängig sind von den angewandten Analysemethoden.

Gängige Methoden der Datenauswertung sind: Häufigkeitsauszählungen, Korrelations- und Regressionsanalysen, Varianzanalysen, Strukturgleichungen, Mehrebenenanalysen, Effektstärkenmasse etc.

Ein besonderer Aspekt der Auswertung bezieht sich auf die Nutzbarkeit der Daten für das praktische Handeln. Werden die Daten so ausgewertet, dass sich Veränderungsmöglichkeiten auf der Ebene des unterrichtlichen Handelns ergeben? Oder verbleiben die Analysen auf der Systemebene? Ist ihre Nützlichkeit damit für die Systemsteuerung gegeben, aber nicht für die Unterrichtspraxis? Es ist davon auszugehen, dass die Ergebnisse internationaler Schulleistungsvergleiche von Lehrerinnen und Lehrern nur dann rezipiert werden, wenn sie sich auf die Handlungsebene beziehen lassen.

1.2.11 Wie werden die Ergebnisse dargestellt?

Hier geht es um die Frage, wie die ausgewerteten Daten präsentiert werden, zum Beispiel in Form von Mittelwerten, Streuungsmassen, Perzentilen oder Quartilen, Tabellen, Abbildungen, Grafiken usw. Zu dieser Frage gehört auch das Ranking von Ländern: Welchen Sinn macht es, verschiedene Bildungssysteme in eine Rangreihe zu bringen?

Helmke (2000) betont, dass Schulleistungsmessungen eingebunden werden sollten in ein Qualitätsmanagement. «Eine isolierte Leistungsmessung hat auf Dauer keine Chance [...]» (Helmke 2000, p. 139). Dies nicht zuletzt deshalb, weil ein blosses Ranking einem Schulsystem oder einer Einzelschule nichts darüber sagt, wie es bzw. sie sich verbessern kann. Wie bereits unter Punkt 10 erwähnt, ist es wichtig, wenn solche Studien etwas bewirken sollen, dass die Daten, die zu Vergleichszwecken erhoben werden, Merkmale von Schule und Unterricht betreffen, die von den Lehrkräften beeinflussbar sind. Das sind im Allgemeinen eher Merkmale im didaktischen und methodischen Bereich, d.h. Merkmale, die auf der Unterrichtsebene liegen. In den bisherigen internationalen Vergleichsstudien sind diese aber eher am Rande oder gar nicht erhoben worden.

²⁰ Aufgaben, die von allen getesteten Personen gelöst werden, sind nicht trennscharf und werden deshalb bei der Konstruktion des Tests ausgeschlossen.

1.3 Analysemodelle

Die elf Fragen, die wir in den vorausgehenden Abschnitten aufgeworfen haben, stellen sich internationalen Schulleistungsvergleichen gewissermassen immanent. Sie betreffen die methodische Qualität der Studien, die Validität (Gültigkeit) ihrer Aussagen und die Brauchbarkeit der Ergebnisse für politische und praktische Entscheidungen.

Es gibt aber zusätzliche Fragen, die man stellen kann. Dazu gehört insbesondere die Frage, wie man das komplexe Gefüge der Bedingungen schulischer Leistungen entwirren kann. Die Frage ist letztlich nur zu beantworten, wenn wir von einem *Modell* ausgehen, das uns hilft, die Vielfalt an denkbaren Einflüssen auf die Leistungen von Schülerinnen und Schülern in eine überschaubare Ordnung zu bringen. Wir wollen im Folgenden zwei solche Modelle darstellen, die in der Forschung zur schulischen Wirksamkeit relativ weit verbreitet sind: das Input-Output-Modell und das Angebot-Nutzungs-Modell.

1.3.1 Das Input-Output-Modell

Die einfachste Art von Modell ist das *Input-Output-Modell*. Man schaut, was geht ins System hinein und was kommt aus dem System heraus. Der Rest (d.h. das System selber) wird als Blackbox behandelt. In der Sprache Heinz von Foersters (1985) stellt sich das Schulsystem als Trivialmaschine dar (vgl. Abbildung 1.2). Man treibt Input-Output-Analyse, ohne in die Maschine hineinzublicken, d.h. ohne die Unterrichtsprozesse zu betrachten.



Abbildung 1.2: Schema einer trivialen Maschine (aus: von Foerster 1985, p. 43)

(a) Eine gute Illustration für die Anwendung des Input-Output-Modells auf die Makroebene von Schulen gibt die *Bildungsökonomie*. Untersucht werden hier Inputvariablen, die in Form von monetären und anderen metrischen Grössen erfasst werden können, wie Lehrerlöhne, Unterrichtserfahrung in Jahren, Klassengrösse bzw. Anzahl Schülerinnen und Schüler pro Lehrperson, Berufsqualifikation (Dauer der Ausbildung), Kosten pro Schülerin bzw. Schüler etc.

Insgesamt muss man zu dieser Forschungstradition sagen, dass sie bisher wenig konsistente Ergebnisse gebracht hat. Des Weiteren sind die Effekte solcher und ähnlicher Inputfaktoren eher gering.

Ein Problem dieser Studien ist allerdings, dass die Varianz der Input-Faktoren unter Umständen zu gering ist, als dass Effekte überhaupt erwartbar wären. Wir haben das Problem in seinen grundsätzlichen Dimensionen unter Punkt 7 im Abschnitt 1.2 erörtert. Wenn wir das Beispiel der Lehrerlöhne nehmen, dann sind diese im Allgemeinen normiert – jedenfalls in der Schweiz und in den meisten europäischen Ländern. Im Falle der öffentlichen Schule ist genau festgelegt, bei welcher Ausbildung, bei welcher Lehrtätigkeit, bei welchem Anstellungsgrad und bei welcher Anstellungsdauer welcher Lohn ausbezahlt wird.²¹ Eine individuelle Varianz ist somit praktisch ausgeschlossen. Das wäre anders, wenn der Lohn nach Leistungskriterien zugeteilt würde – wie ja

²¹ Man schaue sich einmal die Rechtsgrundlagen für die Lehrerbesoldung im Kanton Bern an.

heute durchaus (auch in der Schweiz) gelegentlich gefordert wird. Bei einem leistungsabhängigen Lohn gäbe es mehr Varianz, und man könnte vermuten, dass der Lohnfaktor unter diesen Umständen stärker mit Output-Variablen in Beziehung steht.

(b) Es gibt noch eine andere Form von Input-Output-Forschung, die sich nicht an administrativen oder ökonomischen Kriterien orientiert, sondern an sozialen. Eine traditionelle Fragestellung der *Bildungssoziologie* ist diejenige nach der Ungleichheit der Bildungschancen, wobei im Wesentlichen Unterschiede der sozialen und regionalen Herkunft, der Kultur (inklusive Religion und Rasse) und des Geschlechts untersucht werden. Dieser Ansatz fokussiert die Individuen, d.h. die Schülerinnen und Schüler. Die Ergebnisse sind bekannt: Je höher der soziale Status der Eltern (gemessen am Bildungsniveau, Berufsprestige oder Einkommen), desto besser der schulische Erfolg der Kinder. Je urbaner die Herkunft (Städte und städtische Agglomerationen), desto besser der Schulerfolg. Je besser die Assimilation an die dominante Kultur, desto besser der Schulerfolg. Das Geschlecht spielt in unserer Gesellschaft mittlerweile kaum mehr eine Rolle: Für den Erfolg im obligatorischen Schulbereich ist es kein diskriminierender Faktor mehr.²²

Die bildungssoziologischen Studien geben ein eher pessimistisches Bild, was die Wirksamkeit von Schule – gemessen an der Egalisierung der Bildungschancen – anbelangt. Auf eine berühmte Studie aus den USA – den sogenannten Coleman-Report (vgl. Coleman et al. 1966) – haben wir bereits hingewiesen (vgl. Punkt 5 im Abschnitt 1.2). Diese schätzt, dass gerade einmal 10% der Varianz in den Schulleistungen den *Schulen* zugeschrieben werden kann, der Rest (90%) geht auf andere Einflüsse zurück (Intelligenz, sozialer Status, Familie etc.). Allerdings ist dieser geringe Wert nicht unumstritten geblieben, und es wäre zumindest fahrlässig, allein auf der Basis des Coleman-Reports *allgemeine* Schlüsse über die Wirksamkeit von Schulen ziehen zu wollen. Auf ein methodisches Problem dieser Studie sind wir bereits eingegangen (vgl. Punkt 8 im Abschnitt 1.2). Zu beachten ist auch, dass es sich um eine rein amerikanische Studie handelt, die nicht unbedingt Schlüsse auf die Situation in der Schweiz in oder anderen Ländern zulässt.

1.3.2 Das Angebot-Nutzungs-Modell

Die Vorstellung eines linearen und ungebrochenen Kausalprozesses wird der Realität von Bildungssystemen kaum gerecht. Es gibt daher gute Gründe, das Input-Output-Modell zu überdenken. Allerdings könnten wir dies nur dann tun, wenn uns eine bessere Alternative zur Verfügung stünde. Das ist aber nur beschränkt der Fall. Einen Vorschlag für ein solches alternatives Modell schulischer Wirksamkeit hat Fend (1984) gemacht. Er nennt es ein *Angebot-Nutzungs-Modell*. Dem Modell liegt die Annahme zugrunde, «[...] dass optimale Lernergebnisse dann zu erwarten sind, wenn ein bestmögliches Angebot maximal von Schülern genutzt wird» (Fend 1984, p. 73). Angebotsfaktoren liegen gemäss Fend in der Quantität und Qualität des Unterrichts, Nutzungsfaktoren in der Fähigkeit und Motivation der Schülerinnen und Schüler. Als Stützfaktoren nennt Fend sozialpsychologische Gegebenheiten wie die Klassenumwelt, den Anregungsgehalt des Elternhauses, die Einflüsse der Peergruppe und ein mässiger Medienkonsum. Eine schematische Darstellung des Angebot-Nutzungsmodells findet sich in Abbildung 1.3.

Das Angebot-Nutzungs-Modell verzichtet auf die Vorstellung, pädagogische Wirkungen würden über einfache Kausalprozesse erzeugt. Fend betont, «[...] dass Unterricht und Lernen als interaktive Prozesse anzusehen [sind]» (Fend 1984, p. 73), was eine lineare Darstellung schulischer Wirkungsprozesse ausschliesst. Im Rahmen einer interaktionistischen Betrachtungsweise von Unterricht

²² Auch im postobligatorischen Bereich sind die Differenzen nur mehr gering. Noch nicht abgebaut sind sie bei den Universitäts- und Fachhochschulabschlüssen.

«[...] wird deutlich, dass es sich sowohl auf Lehrer- als auch auf Schülerseiten um *interpretativ- und zielgesteuerte Prozesse handelt*, also um Vorgänge, bei denen die Beteiligten versuchen, eine antizipierte Wirklichkeit im Sinne zu lernender Wissensstrukturen und Fertigkeiten aufzubauen, also eine neue Realität im Sinne dieser Wissensstrukturen und Kompetenzen erst zu schaffen, die ohne diese intellektuell geleiteten Bemühungen nicht zustande kämen» (Fend 1998, p. 323).

Angebote werden gemacht, und man kann sie nutzen. In keinem Fall führt ein Angebot *zwingend* zu seiner Nutzung. Das hat wesentlich damit zu tun, dass Angebote von *Personen* gemacht und von *anderen Personen* genutzt werden. Anders als im Falle von physikalischen Prozessen, ist der Anbietende im Falle von pädagogischen Prozessen nicht die Kausalursache für das Verhalten des Nutzenden. Wir haben es mit zwei verschiedenen Systemen zu tun, die ihrer je eigenen Logik folgen und sich kausal selbst genügen. Insofern löst sich das Angebot-Nutzungs-Modell von der Vorstellung, Erziehung und Unterricht würden einer linearen Kausalität folgen.

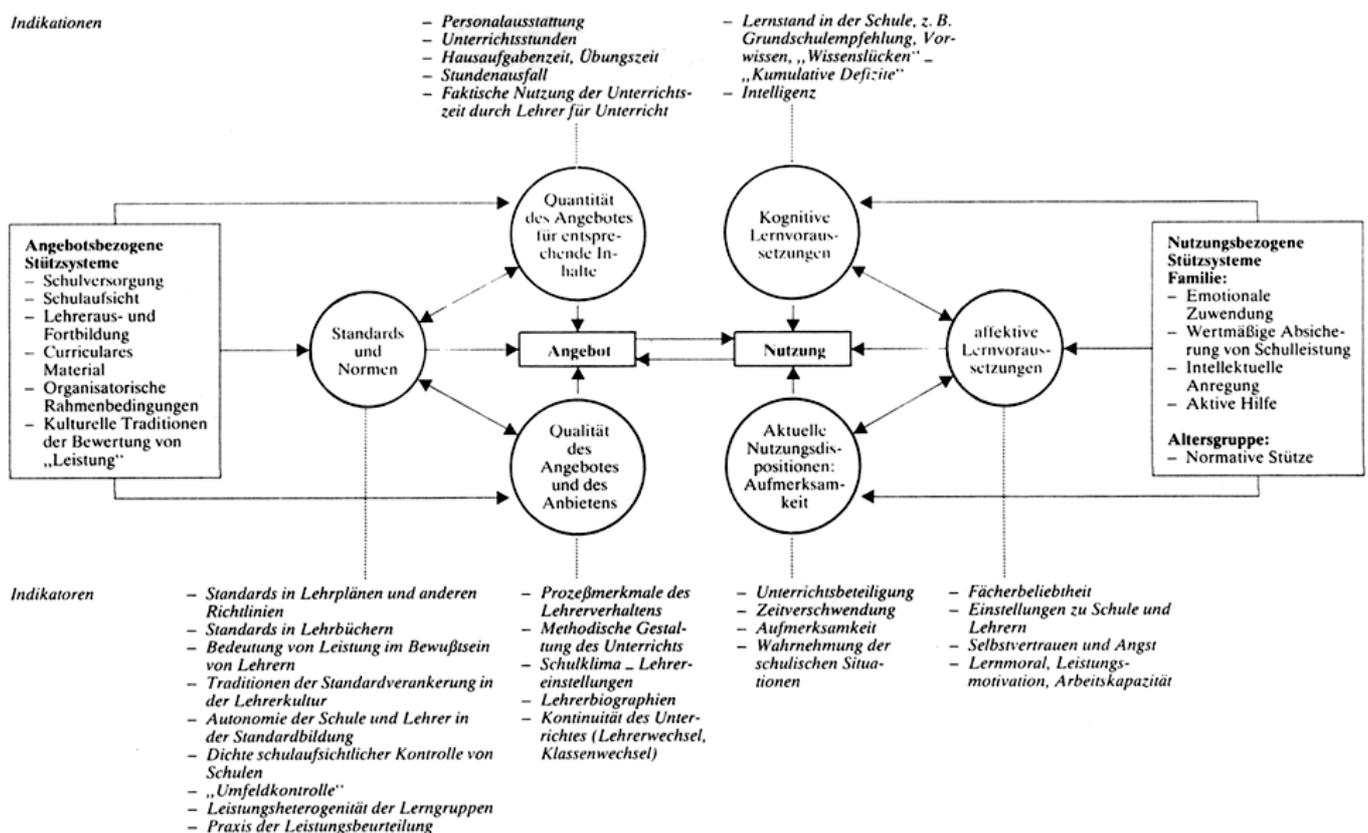


Abbildung 1.3: Angebot-Nutzungsmodell der Schulleistungen (aus: Fend 1984, p. 72)

Ein gravierendes Problem des Input-Output-Modells ist nicht nur, dass es die Blackbox im Dunkeln lässt, d.h. darauf verzichtet, das Prozessgeschehen im Unterricht mit zu erfassen. Ein Problem liegt auch darin, dass es eine naive Vorstellung von Kausalität suggeriert. Was den ersten Punkt angeht, so werden wir uns im 5. Kapitel mit Forschungstraditionen auseinandersetzen, die das Binnengeschehen im Unterricht in die Kausalanalyse schulischer Wirkungen einbeziehen. Das geschieht im Rahmen von Fallstudien und der sogenannten Prozess-Produkt-Forschung (vgl. Abschnitte 5.2 und 5.3). Charakteristisch für diese Art von Forschung ist, dass man gewissermaßen in die Blackbox der Schule hineinblickt. Wenn wir nochmals auf Heinz von Foerster zurückgreifen, dann werden in diesen Studien die Schulen als nicht-triviale Maschinen mit inneren Zuständen behandelt.

Was den zweiten Punkt anbelangt, so bauen auch diese Forschungsparadigmen oftmals auf ein naives Verständnis von Wirksamkeit. Suggestiert wird, dass die Kausalität von links nach rechts fließt und im Sinne einer unkomplizierten Kraftübertragung der Ursachen auf die Wirkungen fungiert. Ausgeblendet werden Feedbackprozesse (Rückkopplungen), kurvilineare Zusammenhänge und Kausalitätsunterbrechungen. Wir können uns hier nicht im Detail mit diesen Problemen auseinandersetzen, wollen aber folgendes festhalten. Es ist wenig plausibel anzunehmen, irgendein Input, der einer Schule gegeben wird, wirke im Sinne eines ursächlichen Kausalprozesses. Vielmehr ist davon auszugehen, dass die Schule ein eigenes System darstellt, das den Input wahrnimmt, rezipiert, verarbeitet, umgestaltet und nach eigenem Ermessen wirksam macht.

Ähnliches dürfte für die Ebene des Unterrichts gelten. Auch hier ist nicht anzunehmen, dass ein Lehrerverhalten ursächlich ein Schülerverhalten bewirkt. Eine Schülerin, die lernt, tut dies aus eigenen Stücken. Die Lehrerin mag das Lernverhalten zwar angeregt haben, ist aber nicht die Kausalursache dessen, was die Schüler bzw. der Schüler tut. Schülerinnen und Schüler vermögen in diesem Sinne Kausalprozesse zu unterbrechen - und im Extremfall das Gegenteil dessen zu machen, was die Lehrperson intendiert hat.

Ähnliches gilt wiederum für die Beziehungen zwischen der einzelnen Lehrkraft und der Schulleitung oder der Schulaufsicht. Was ein Schulleiter mit seiner Schule will, kann er nicht in Kausalprozesse umsetzen, die mit berechenbarer Sicherheit erzeugen liessen, was er beabsichtigt. Von Schulreformen her ist gut bekannt, dass diese unter Umständen mit dem entschiedenen Widerstand der Lehrerinnen und Lehrer rechnen müssen, so dass selten erreicht wird, was sich die Bildungsadministration an Reformzielen gesetzt hat.

Wir wollen nicht weiter auf das Angebot-Nutzungs-Modell eingehen. Es genügt, wenn wir zeigen konnten, dass die Aufklärung von Kausalprozessen im Falle von schulischer und unterrichtlicher Wirksamkeit keine simple Angelegenheit ist. Das bei internationalen Schulleistungsvergleichen vorherrschende Input-Output-Modell stellt daher keineswegs die einzige Möglichkeit dar, wie man sich die Wirksamkeit von Schule vorstellen kann.

2 Internationale Schulleistungsvergleiche am Beispiel der IEA Lesestudie (Study of Reading Literacy)

2.1 Generelles

Dieses zweite Kapitel beschäftigt sich schwergewichtig mit den allgemeinen Voraussetzungen internationaler Schulleistungsvergleiche. Es soll aufgezeigt werden, welche übergeordneten Ziele die Studien verfolgen, wer die Studien durchführt und mit welchen forschungstechnischen Herausforderungen Planung sowie Auswertung solcher Studien konfrontiert sind. Dabei sollen die allgemeinen Aussagen nach Möglichkeit stets auch an einem Beispiel konkretisiert werden.

Den Hintergrund für diese Konkretisierung bildet dabei die *IEA Study of Reading Literacy* (im Folgenden stets IEA Lesestudie genannt). Sie hatte zum Ziel, die Lesekompetenz und die Leseaktivität von 9- und 14-jährigen Schülerinnen und Schülern international zu beschreiben. Die IEA Lesestudie wurde in 27 Ländern durchgeführt. In der Schweiz setzte sich die Stichprobe aus 226 Klassen der 3. Klassenstufe und aus 386 Klassen der 8. Klassenstufe aus allen vier Sprachregionen zusammen. Kinder aus Klassen mit einem besonderen Lehrplan wurden nicht in die Studie einbezogen.

Die Datenerhebung fand im März 1991 statt, der Bericht über die vollständige Auswertung der Schweizer Resultate wurde 1996 vorgelegt (vgl. Notter, Meier-Civelli, Nieuwenboom, Rüesch & Stoll 1996). Damit ist die IEA Lesestudie älter als die anderen in diesem Skript vorgestellten Vergleichsstudien TIMSS (Kapitel 3) und PISA (Kapitel 4).

2.1.1 Die Ziele international vergleichender Studien

International vergleichende Studien sind durch unterschiedliche Zielvorstellungen gekennzeichnet. Einige sind (bildungs-)politisch motiviert, einige sind durch eine wirtschaftliche Ausrichtung geprägt und andere schliesslich dienen Zielen, die auf nationalen, schulorganisatorischen Interessen basieren. In den folgenden Ausführungen werden mögliche Zielvorstellungen aufgezeigt, wobei es sich keineswegs um eine erschöpfende Darstellung handelt.

In Zusammenhang mit der Zielvorstellung von internationalen Studien spricht Döbert vom Wunsch nach einem «*Aufbau eines europäischen Bildungsraumes*» (Döbert 2002, p. 2). Dies sei der aktuelle politische Leitgedanke der Europäischen Union (EU). Laut seinen Aussagen werden internationale Vergleichsuntersuchungen bei solchen Vorhaben als Datenquellen miteinbezogen und sind aus diesem Grund von grosser Wichtigkeit (vgl. ebd.). Ein international vernetzter Bildungs- und Arbeitsraum, wie ihn Döbert vorschlägt, erfordert eine internationale Kenntnis und Anerkennung von Ausbildungen (vgl. ebd.). Verhandlungen um die «*Europakompatibilität*» von Schweizer Ausbildungen (Notter et al. 1996, p. 12), wie beispielsweise die bilateralen Abkommen zwischen der Schweiz und der EU, beruhen auf Daten aus international vergleichenden Studien.

Neben diesen beiden auf Europa ausgerichteten Zielen gibt es weitere, die ihre Verankerung in der Politik und der Wirtschaft finden. So ist heutzutage ein zunehmendes Interesse für den Nutzen, den Bildungsinstitutionen erbringen, und die damit verbundenen Bildungserträge zu verzeichnen (vgl. Weinert 2002, p. 26). In welchem Zusammenhang steht nun diese neue Ausrichtung der international vergleichenden Studien, die man auch als *Outputsteuerung* bezeichnen kann?²³ Das Interesse an Daten, welche die «Produktivität» der Schulen beziffern sollen, steht in Verbindung

²³ Vgl. Abschnitt 1.2.1.

mit (bildungs-)politischen Forderungen, welche nach einer «Überprüfung des Wirkungsgrades der eingesetzten Mittel» verlangen (Döbert 2002, p. 3). Die aus Leistungsvergleichen gewonnenen Daten sollen als *Basis für zukünftige Investitionen* in das Bildungssystem dienen.

In diesen Zusammenhang kann im Übrigen nicht nur die Diskussion um die Qualität und Leistungsfähigkeit des Bildungssystems *generell*, sondern auch die Frage nach der Güte von einzelnen *Teilen* des Bildungssystems gestellt werden (vgl. Weinert 2002, p. 26f.). So erlauben die Resultate von Vergleichsstudien gemäss Baumert auch *nationale Vergleiche* (in der Schweiz beispielsweise zwischen den einzelnen Kantonen). Gerade die ernüchternden Ergebnisse der TIMSS Studie wurden in Deutschland zum Anlass genommen, systematisch über Qualitätsentwicklung und Qualitätssicherung nachzudenken. Dabei wird der Blick laut Baumert auf den Unterricht – die Hauptaufgabe der Schule – (zurück-)gelenkt (vgl. Baumert, Bos & Lehmann 2000, p. 20ff.). Allerdings darf keinesfalls vergessen werden: Nationale Schlüsse aus internationalen Studien dürfen nur dann gezogen werden, wenn die Tests entsprechend konstruiert wurden, das heisst, wenn sie spezifische *nationale* Zusätze enthalten. Sonst schaffen sie, wie Weinert anführt, Orientierungswissen, aus welchem keine direkten Schlüsse auf schulorganisatorischer oder curricularer Ebene gezogen werden dürfen (vgl. Weinert 2002, p. 26ff.).

So wie internationale Untersuchungen nur mit einem spezifisch dafür konstruierten Test Auskunft über Unterricht geben können, so gibt es weitere Ansprüche, welche diese Studien *nicht* erfüllen können. So können sie zwar Unterschiede benennen, erklärt werden sie aber nicht. Aus der IEA Lesestudie erfährt man so beispielsweise, dass die Mädchen generell etwas besser lesen als die Knaben. Nach den Gründen müsste aber mit einer weiteren Untersuchung, einer Folgestudie, gesucht werden. So verstehen sich internationale Studien als eine Art *Planungshilfe* mit Anstosscharakter für weitere Forschungsarbeiten (vgl. Weinert 2002, p. 28).

Im Weiteren vermögen internationale Untersuchungen auch nicht umfassende Zusammenhänge wie beispielsweise denjenigen zwischen der Bildung und der beruflichen *Laufbahn* einer Person aufzuklären. Weinert führt hier an, dass die Zusammenhänge zwischen der Bildung und der Wirtschaft eine hohe Komplexität aufweisen und die Bedeutung der von der Schule erbrachten Leistungen wissenschaftlich umstritten ist. Für all jene, welche öffentliche Verantwortung tragen, sei aber gerade diese Verknüpfung zwischen Schulbildung und beruflicher Laufbahn zentral. So spricht Weinert davon, dass zumindest in den industrialisierten Ländern die Schulbildung einen grossen Einfluss auf die berufliche Entwicklung hat. Dabei gebe es aber neben der Schulbildung weitere Faktoren wie u.a. der soziale Status der Eltern oder persönliche Merkmale, die für die berufliche Laufbahn eine zentrale Rolle spielen. Hier wird deutlich, dass reine Leistungsmessungen, wie sie in der IEA Lesestudie mit der Lesekompetenz durchgeführt wurden, nicht umfassend auf die Frage nach dem Zusammenhang zwischen Schulbildung und beruflicher Laufbahn antworten können. Die OECD hat aus diesem Grund ein Konzept entwickelt, das auf der Frage basiert: «Was braucht ein Mensch, um in einer Gesellschaft und einer beruflichen Welt eine positive Rolle spielen zu können?» Dieses Konzept der Kompetenzen zieht neben kognitiven Fähigkeiten und Fertigkeiten auch motivationale und soziale Aspekte in die Überprüfung mit ein (vgl. Weinert 2002, p. 26ff.).²⁴

Betrachtet man nur die Ziele der IEA Lesestudie allein, so decken sie sich weitgehend mit den bereits erwähnten allgemeinen Zielen international vergleichender Studien. Die Autorinnen und Autoren der IEA Lesestudie stellten ebenfalls ein politisches Interesse an Bewertungsindikatoren für die schweizerische Bildungslandschaft und ein Bedürfnis nach der Angleichung schweizerischer Ausbildungsgänge an internationale Standards fest. Die Lesestudie verfolgte im Weiteren das Ziel,

²⁴ Zum Konzept der Basiskompetenzen vgl. Abschnitt 4.1.

die Anzahl der funktionalen Analphabetinnen und Analphabeten²⁵ in den beiden untersuchten Altersgruppen zu erheben (vgl. Notter et al. 1996, p. 12f.).

Mit den Durchführenden der PISA Studie formuliert können wir zusammenfassen: Internationale Vergleichsstudien sollen den teilnehmenden Nationen Daten über ihre Ressourcen wie auch über die Leistungsfähigkeit ihrer Bildungsinstitutionen zur Verfügung stellen (vgl. Deutsches PISA-Konsortium 2001, p. 15). Da Schulsysteme, Curricula und weitere Kerngrößen von Land zu Land, wie zum Teil zusätzlich auch innerhalb eines Landes, unterschiedlich sind, werden an die Durchführung und die Konstruktion solcher Untersuchungen hohe Ansprüche gestellt (vgl. Baumert et al., 2000, p. 22). Dies führt uns zu den folgenden Teilen, in denen die wichtigsten durchführenden Institutionen internationaler Vergleichsstudien und Probleme der Testkonstruktion vorgestellt werden.

2.1.2 Zur Durchführung international vergleichender Studien

Internationale Vergleichsstudien werden in der Regel von einer übernationalen Institution geleitet, wobei in jedem teilnehmenden Land weitere Institutionen für die Koordination und die Durchführung Verantwortung übernehmen. Bei der IEA Lesestudie, an der sich die Schweiz ebenfalls beteiligte, fungierte die IEA (International Association for the Evaluation of Educational Achievement) mit Sitz in Amsterdam als leitende Instanz. Die IEA ist eine nichtstaatliche Organisation, deren Mitglieder sich aus nationalen Ministerien, Pädagogischen Hochschulen sowie Forschungsinstitutionen zusammensetzen (vgl. Notter et al. 1996, p. 11). Sie wurde 1961 gegründet und momentan sind 62 Mitglieder in der IEA vertreten. Die von der IEA bisher durchgeführten Studien befassten sich in erster Linie mit den Fachrichtungen Mathematik, Naturwissenschaften sowie dem Lesen und dem Schreiben. Die Institution leitete auch die mathematisch und naturwissenschaftlich orientierte TIMSS Studie (vgl. Arnold 1999, p. 71ff.).

Die Koordination der schweizerischen Durchführung der IEA Lesestudie übernahm die Schweizerische Koordinationsstelle für Bildungsforschung (SKBF) in Zusammenarbeit mit dem Psychologischen Institut der Universität Zürich und der beratenden Funktion der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) (vgl. Notter et al. 1996, p. 10f.).

Die SKBF ist 1971 durch die EDK und den Bund gegründet worden. Ihre Aufgabe ist es, den Dialog zwischen Bildungspolitik, -praxis und -forschung zu fördern und zu unterstützen. Im Rahmen der internationalen Bildungsforschung übernimmt sie die Funktion der Dokumentation von Forschungsleistungen, die sich auf das schweizerische Bildungssystem beziehen.

Die Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) ist bestrebt, die bildungspolitischen Interessen der einzelnen Kantone zu vereinigen und konzentriert sich auf nationale sowie auf internationale Projekte rund um die Bildung. So ist sie an den bilateralen Abkommen zwischen der Schweiz und der Europäischen Union beteiligt und dokumentiert bzw. kommentiert auch die nationalen Ergebnisse von PISA.

Speziell erwähnt werden soll schliesslich noch das Institut für Bildungsevaluation (ehemals Kompetenzzentrum für Bildungsevaluation und Leistungsmessung, KBL). Es ist eine Einrichtung der Universität Zürich und versteht sich als Dienstleistungsangebot für das Bildungswesen sowie für Wirtschaft und Verwaltung. Das Institut führt Evaluationen für Auftraggeber durch, setzt sich für eine verstärkte Kooperation von Wirtschaft und Verwaltung ein und strebt eine internationale

²⁵ Funktionale Analphabetinnen und Analphabeten sind Personen, die nicht gut genug lesen können, um den minimalen alltäglichen Anforderungen der Gesellschaft an die Lesekompetenz ihrer Mitglieder zu genügen. Funktionale Analphabeten sind daher von der vollen Teilnahme am gesellschaftlichen Leben ausgeschlossen (vgl. Notter 1996b, p. 109).

Ausrichtung der Bildungsforschung an. Das IBE ist für die Organisation, die Durchführung und die Auswertung der Daten von PISA für den Kanton Zürich und das Fürstentum Liechtenstein zuständig.

Tabelle 2.1: Wichtige nationale und internationale Institutionen mit Beteiligungen an Schulleistungs-
vergleichen

Name/Bezeichnung	Signalement	Aufgabe	Beteiligungen
International Association for the Evaluation of Educational Achievement IEA (iea.nl)	<ul style="list-style-type: none"> international wissenschaftliche Interessen teilw. bildungspolitische Interessen 	Planung, Koordination und Durchführung gross angelegter internationaler Vergleichsstudien mit dem Ziel, ein vertiefteres Verständnis für das Zusammenspiel von Politik und Bildung zu entwickeln.	IEA's SRL FIMS, SIMS, TIMSS PIRLS
Schweizerische Konferenz der kantonalen Erziehungsdirektoren EDK (edk.ch)	<ul style="list-style-type: none"> national bildungspolitische Interessen 	Nationales Bildungsmonitoring etablieren, Harmonisierung der obligatorischen Schule und Sorge um den Berufsstand der Lehrenden	PISA
Schweizerische Koordinationsstelle für Bildungsforschung SKBF (skbf-csre.ch)	<ul style="list-style-type: none"> national bildungspolitische Interessen wissenschaftliche Interessen 	Stärkung der Bildungsforschung, Verbesserung des Dialogs zwischen Bildungspolitik, -praxis, -verwaltung und -forschung, Übernahme von Aufgaben, die im Interesse unterschiedlicher Akteure der Bildungslandschaft liegen.	
Organisation für wirtschaftliche Zusammenarbeit und Entwicklung OECD (oecd.org)	<ul style="list-style-type: none"> international wirtschaftliche Interessen 	Neben anderen Bereichen befasst sich die OECD mit der Bildung in ihren 30 Mitgliedsstaaten und macht es sich mit PISA zur Aufgabe, ihren Mitgliedern Daten über deren Bildungssysteme zur Verfügung zu stellen.	Internationale Leitung von PISA
Consortium of Institutions for Development and Research in Education in Europe CIDREE (cidree.org)	<ul style="list-style-type: none"> international bildungspolitische Interessen 	Setzt sich für die Verstärkung der Zusammenarbeit zwischen den europäischen Bildungssystemen ein.	
Bundesamt für Statistik BfS (bfs.admin.ch)	<ul style="list-style-type: none"> national wissenschaftliche Interessen bildungspolitische Interessen 	Erarbeitet statistische Grundlagen und erstellt Übersichten zu verschiedenen Fachbereichen wie beispielsweise Bildung und Wissenschaft	Arbeitsgruppe des Bundesamtes ist an der schweizerischen Durchführung von PISA mitbeteiligt
Schweizerische Koordinationskonferenz Bildungsforschung CORECHED (www.coreched.ch)	<ul style="list-style-type: none"> national bildungspolitische Interessen 	Koordiniert die Bildungsforschungspolitik zwischen Geldgebern beispielsweise der EDK oder dem Schweizerischen Nationalfonds (SNF) und -nehmern. Formuliert Empfehlungen wie u.a die Wünschbarkeit schweizerischer Beteiligung an internationalen Projekten und valorisiert Ergebnisse aus der Bildungsforschung.	Mitarbeit an der TIMSS Video-Studie
Schweiz. Gesellschaft für Bildungsforschung SGBF (www.sgbf.ch)	<ul style="list-style-type: none"> national bildungspolitische Interessen 	Setzt sich für die Förderung der Bildungsforschung ein.	
Abteilung für Bildungsplanung und Evaluation der Erziehungsdirektion des Kantons Bern BiEv (www.erz.be.ch/bildungsforschung)	<ul style="list-style-type: none"> kantonal bildungspolitische Interessen 	Evaluert Wirkungen des Bildungswesens, entwickelt kantonal neue Lehrpläne und begleitet Schulversuche und Pilotprojekte.	Leitung der TIMSS in der Schweiz sowie Mitarbeit bei Durchführung von PISA im Kt. Bern, Vergleich bernischer mit nationalen und internationalen Ergebnissen

Name/Bezeichnung	Signalement	Aufgabe	Beteiligungen
Institut für Bildungsevaluation der Universität Zürich IBE (www.ibe.uzh.ch)	<ul style="list-style-type: none"> • national und international • wirtschaftliche Interessen • wissenschaftliche Interessen 	Versteht sich als Dienstleistungsangebot für das Bildungswesen sowie für Wirtschaft und Verwaltung. Führt Evaluationen für Auftraggeber durch, setzt sich für eine verstärkte Kooperation von Wirtschaft und Verwaltung ein und strebt internationale Ausrichtung der Bildungsforschung an.	Durchführung IEA Lesestudie in der CH, Mitarbeit bei TIMSS und PISA

2.1.3 Forschungslogik und ihre Probleme

2.1.3.1 Testkonstruktion: Vergleichbarkeit, Mehrdimensionalität und Vergleichsnorm

Internationale Studien müssen sich die *Vergleichbarkeit* der (nationalen) Ergebnisse zum Ziel setzen und sind deshalb bereits in ihrer Konzeption und in der (Test-)Konstruktion auf diese Vergleichbarkeit auszurichten. Die grössten Schwierigkeiten liegen hier in den curricularen Unterschieden zwischen den Ländern. Es stellt sich also die Frage, wie es möglich ist, Schulleistungen einander gegenüberzustellen, wenn sie die Folge unterschiedlicher Schulsysteme sind. Es gilt daher, bei der Testentwicklung gewisse Standards einzuhalten.

In diesem Zusammenhang plädiert Arnold (1999) dafür, einer Untersuchung zunächst die Bildung eines «internationalen Kerncurriculums» vorzuschicken, das heisst ein Curriculum zu generieren, in das die Inhalte aufgenommen werden, die in allen an der Studie teilnehmenden Ländern in den Schulen gelehrt werden (vgl. ebd., p. 92ff.). Der Aspekt der *Fairness* – will heissen: alle beteiligten Länder werden in gleichem Masse in ihren Merkmalen berücksichtigt – ist ein zentrales Kriterium. Je besser den Gemeinsamkeiten der Teilnehmerländer Rechnung getragen werden kann, desto höher ist die internationale Vergleichbarkeit der Ergebnisse und somit die Validität²⁶ einer Untersuchung. Um diese zu erreichen, werden häufig Leistungsvergleiche in der Mathematik, den Naturwissenschaften oder dem Lesen und dem Schreiben durchgeführt, da diese Fächer international bereits eine hohe Homogenität aufweisen. Das heisst, in diesen Fächern wird international ähnliches Wissen vermittelt. Dies erleichtert die Konstruktion von international einsetzbaren Tests erheblich und garantiert auch eine höhere Äquivalenz (Gleichwertigkeit), als wenn beispielsweise die Leistungen in einem Fach wie Geschichte überprüft würden, weil Geschichte im Vergleich zu den naturwissenschaftlichen Fächern, der Mathematik oder dem inhaltsfreien Lesen verständlicherweise stärker länderspezifisch ausgerichtet ist (vgl. Baumert 1998, p. 220f.).

Wie in Abschnitt 2.1.1 bereits angesprochen wurde, können reine Schulleistungsvergleiche nicht auf umfassende Fragen nach dem Zusammenhang von Schulbildung und beruflicher Laufbahn antworten. Um die prognostische Validität der Untersuchungen und ihre Aussagekraft in Bezug auf solche Fragestellungen zu erhöhen, wird daher auch versucht, von reinen *Fächerleistungsvergleichen* weg zu Vergleichen von lebenspraktischen Kompetenzen zu gelangen. Mit ihrem spezifischen Konzept der Kompetenzen zielt die OECD in PISA genau in diese Richtung (vgl. Weinert 2002, p. 27f.).²⁷ Die OECD berücksichtigt damit auch eine zweite Schwierigkeit der Testkonstruktion, welche in der *Mehrdimensionalität* oder auch Vielschichtigkeit des Untersuchungsgegenstandes liegt. So kann eine Studie nie *die* mathematische Kompetenz oder *das* Lesen untersuchen und erfassen – genauso wenig wie die Schule diese Kompetenzen lehren könnte. Deshalb konzentrieren

²⁶ Eine hohe Validität zeigt sich darin, dass die Tests eine möglichst hohe Gültigkeit für alle teilnehmenden Länder aufweisen (vgl. Arnold 1999, p. 92ff.).

²⁷ Das Konzept wird in Abschnitt 4.1 eingeführt.

sich die Forscherinnen und Forscher darauf, Teilrichtungen eines Faches oder gewisse Dimensionen einer Kompetenz zu überprüfen.

Auch bei der IEA Lesestudie, die sich ja dem Namen nach einfach mit *der* Lesekompetenz der Schweizer Schülerinnen und Schüler auseinandersetzte, war dies der Fall. Als Definition gab die IEA noch einen sehr umfassenden Begriff der Lesekompetenz vor. An ihm sollte sich die Testkonstruktion orientieren (vgl. Notter et al. 1996, p. 19): Untersucht werden sollte laut IEA «[...] die Kompetenz und die Praxis der linguistischen und kognitiven Verarbeitung jener Formen schriftlicher Sprache, die von der Gesellschaft verlangt bzw. erwartet und/oder vom Individuum geschätzt werden» (ebd., p. 19). Es liegt auf der Hand, dass eine so verstandene Lesekompetenz nicht direkt beobachtet werden kann, sondern zuerst einmal in ihren verschiedenen Facetten «sichtbar», «beobachtbar» und v.a. auch messbar gemacht werden muss. Wissenschaftlich gesprochen stellt sich hier die Frage, wie Lesekompetenz *operationalisiert* wird. In der Operationalisierungsphase einer Untersuchung geht man also dem Problem nach, wie man aus einem theoretischen Konstrukt wie der obigen Definition der Lesekompetenz einen Test herstellt, dessen Resultate quantifizierbar – also in eindeutigen Messwerten ausdrückbar – sind.

So wurden für die IEA Lesestudie beispielhaft drei Arten von Texten in die Untersuchung einbezogen, welche ein möglichst breites Spektrum der obigen Definition abdecken sollten:

- Dokumente
- Sachtexte
- Erzählungen

Die Kategorie der *Erzählungen* umfasst Texte wie Erzählungen, Romane oder Unterhaltungsliteratur. Texte dieser Art liest man also in erster Linie zur Unterhaltung. Dabei war es aber im Rahmen dieser Untersuchung nicht möglich, Daten zum Textverständnis eines ganzen Romans zu erheben. Mehrseitige Erzählungen fanden aber Eingang in die Testserie. Ein Beispiel für diese Textart aus der Untersuchung der dritten Klasse ist die folgende Geschichte (vgl. Rüesch & Nieuweboom 1996, p. 48ff.):

Der Vogel und der Elefant

Ein grosser Baum wuchs mitten im Urwald. Ganz oben hatte eine kleine Vogelmutter für ihre Familie mit drei Vogelkindern ein Nest gebaut. Eines Tages kam ein Elefant vorbei. Dieser lehnte sich gegen den Stamm und kratzte sich den Rücken. Der Baum knackte und fing an zu schaukeln. Voller Angst kuschelten sich die Vogelkinder an ihre Mutter. Diese streckte ihre Schnabelspitze aus dem Nest heraus und sagte: «He, grosses Tier, es gibt hier überall viele Bäume. Weshalb schüttelst Du gerade diesen Baum? Meine Vogelkinder haben Angst und könnten aus dem Nest fallen.»

Der Elefant sagte nichts, aber er schaute den Vogel mit seinen kleinen Augen an, bewegte seine grossen Ohren im Wind und ging.

Am nächsten Tag kehrte der Elefant wieder zurück und kratzte sich wieder am gleichen Baumstamm. Der Baum begann zu schaukeln. Die erschrockenen Vogelkinder drängten sich wiederum dicht unter den Flügel der Mutter zusammen. Nun war die Vogelmutter verärgert. «Ich befehle Dir, dass Du aufhörst unseren Baum zu schütteln», schrie sie, «oder ich werde es Dir zeigen!»

«Was könntest Du machen gegen einen solchen Riesen wie mich?» lachte der Elefant. «Wenn ich wollte, könnte ich diesem Baum einen solchen Stoss versetzen, dass Dein Nest und Deine Kinder weit weg geschleudert würden.»

Die Vogelmutter sagte nichts. (...)

Die Schülerinnen und Schüler wurden bei diesem Text etwa gefragt, welcher Satz der Geschichte aussage, dass der Elefant *denkt*, er sei der stärkste. Während die richtige Antwort sich auf den Satz «Was könntest Du machen [...]» bezog, nannten viele Schüler/-innen den Satz «He, grosses Tier [...]». Die Schwierigkeit der Frage scheint darin zu liegen, dass nicht danach gefragt wird, ob der Elefant wirklich *gross ist*, sondern ob er dies *denkt* (vgl. ebd., p. 49).

In die Kategorie der *Sachtexte* wurden Texte aus Lehr- und Schulbüchern oder auch Zeitungsartikel eingeteilt. Die Kategorie der Sachtexte zeichnet sich also dadurch aus, dass man sie liest, um etwas zu erfahren und Sachinformationen zu gewinnen. Ein Beispiel für diese Textart ist in Abbildung 2.1 auf der nächsten Seite wiedergegeben.

Die Zukunftsperspektiven des Lasers

Anfangsjahre

Der Laser ist eine der dramatischsten Erfindungen des 20. Jahrhunderts. Er ist ein Gerät, das einen gebündelten Lichtstrahl aussendet. Dies ist aber kein gewöhnliches Licht; es hat besondere Eigenschaften, und Laser können für eine Vielfalt von unterschiedlichen Aufgaben eingesetzt werden. Sie können verwendet werden, um Metalle mit hoher Geschwindigkeit zu schneiden und um maschinelle Bearbeitungen und Bohrarbeiten in harten und spröden Materialien durchzuführen. Präzise geführte Laser sind verwendet worden, um winzige medizinische Operationen und Zahnarztarbeiten auszuführen. Man hat sie dazu benutzt, um die Lichtgeschwindigkeit mit einer unglaublichen Genauigkeit zu messen und um winzige kleine Schadstoffteile aus grosser Entfernung in der Atmosphäre zu entdecken. Die Anwendungsliste wird laufend länger.

Diese und viele andere Entwicklungen haben sich in weniger als 20 Jahren vollzogen. Der erste Laser wurde 1960 von Theodore Maiman gebaut. Er enthielt einen speziellen Stab aus Rubin, der spiralförmig mit einer Leuchtröhre umwickelt war. Wenn der Stab aus Rubin intensiven Lichtblitzen von normalem weissem Licht ausgesetzt wurde, gab er rote Laserstrahlenimpulse ab.

Dieser Durchbruch ereignete sich nicht völlig überraschend. Die theoretischen Grundlagen des Laserprinzips wurden 1917 von Albert Einstein veröffentlicht. Was fehlte, waren die technischen Möglichkeiten, um Laser zu bauen. Erst durch den schnellen Fortschritt der Technologie nach dem Zweiten Weltkrieg konnten die notwendigen technischen Voraussetzungen für die Herstellung entwickelt werden.

Als dann die erste Erfindung gemacht war, waren die Schleusen geöffnet. Innerhalb weniger Jahre wurde eine Unmenge von verschiedenen Lasern produziert, wobei feste, flüssige und gasförmige Materialien verwendet wurden. Die herrlichen Farben des Laserlichts, zusammen mit seinen angeblich phantastischen Eigenschaften erzeugten allgemeines Interesse.

Die Möglichkeit von «Todesstrahlen», bereits aus der Science-Fiction-Literatur bekannt, wurde schnell in Filmen aufgegriffen. Im Film «Goldfinger» drohte James Bond ein schreckliches Ende, als sich ein «Industrielaser» seinen Weg durch einen Goldbarren zu ihm schnitt. Dieses Bild des Lasers hat sich verfestigt, und Laser werden als Waffe der Zukunft betrachtet. Der visuelle Reiz des Lasers wurde ebenfalls genutzt; Popgruppen und Freilichtspektakel bedienten sich bald perfekter Farblaservorführungen.

(...)

Fragen:

1. Wann wurde diesem Text zufolge der erste Laser gebaut?
 - A 1917
 - B Während des 2. Weltkrieges
 - ✓ C 1960
 - D 1970

2. Die Entwicklung der Laser war für die Wissenschaftler keine völlige Überraschung, weil ...
 - A sie nicht schwierig zu bauen sind.
 - B sie für viele Zwecke verwendet werden können.
 - C sie öfters in der Science-Fiction-Literatur beschrieben wurden.
 - ✓ D die Theorie seit einiger Zeit bekannt war.

Abbildung 2.1: Beispiel zur Textsorte «Sachtext» für die achte Klasse
(aus: Rüesch & Nieuweboom 1996, p. 47)

Unter die Kategorie der *Dokumente* fallen in einem weiten Sinn gefasst Texte wie Fahrpläne, Abbildungen oder Tabellen. Ein Textbeispiel aus der Untersuchung in der achten Klasse ist in Abbildung 2.2 wiedergegeben. In diesem Beispiel wurden die Schülerinnen und Schüler etwa gefragt, welchen Bus ein Kind nehmen muss, wenn es zum Bahnhof fahren will oder wann der nächste Bus zum Bahnhof fährt.

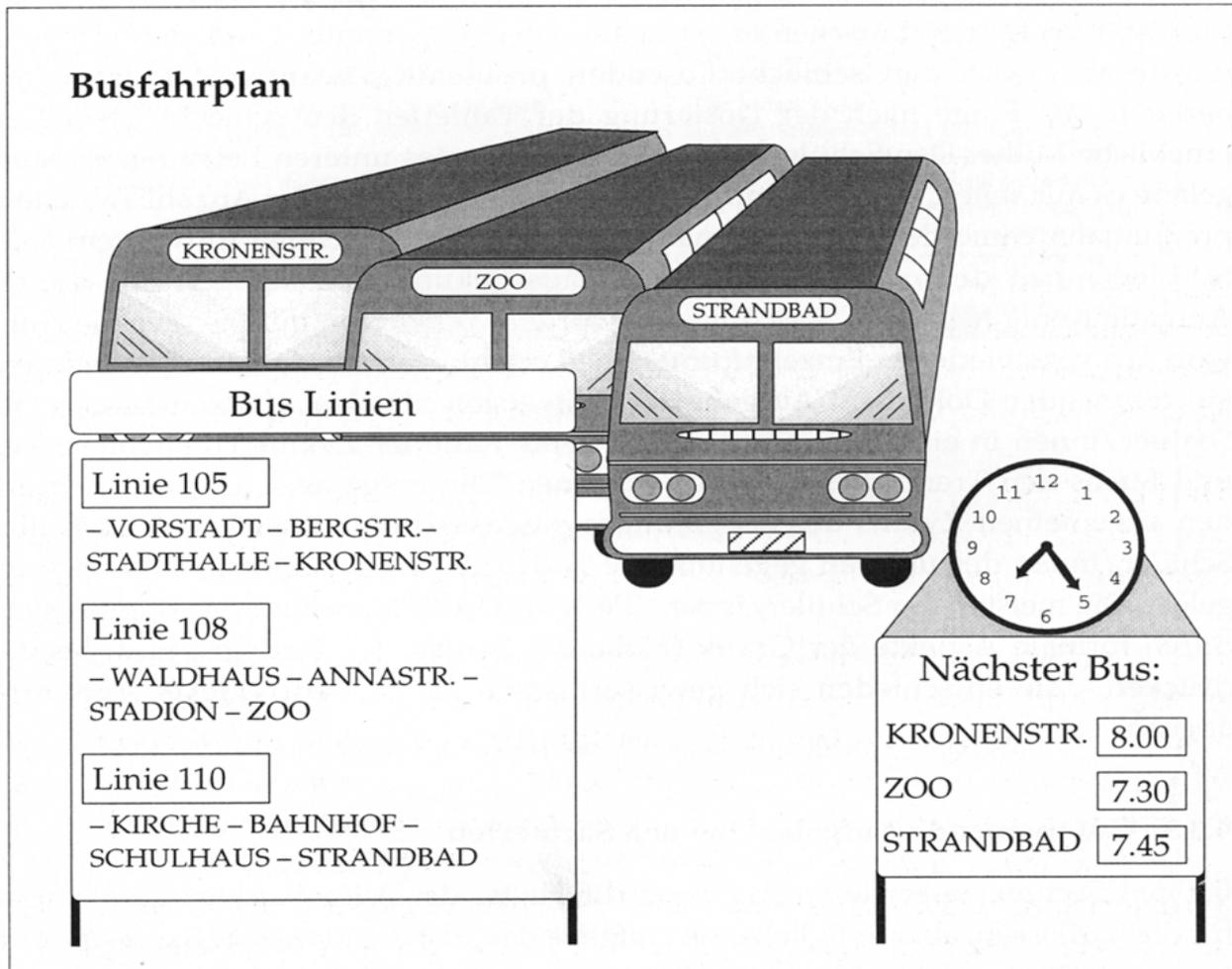


Abbildung 2.2: Beispiel zur Textsorte «Dokument» für die dritte Klasse
(aus: Rüesch & Nieuweboom 1996, p. 45)

Nach der Festlegung dieser Kategorien wurden in einem nächsten Schritt die einzelnen Texte und Aufgaben (Items) für die jeweiligen Textarten ausgewählt. Dieser Prozess ist ein wichtiger Teil der Testkonstruktion, da die ausgewählten Items den drei Kategorien der Fragestellung möglichst genau entsprechen müssen. Jede Kategorie ist für sich eindimensional und auch die Ergebnisse wurden für die Dokumente, die Sachtexte und die Erzählungen einzeln berechnet. Die IEA Lesestudie beschränkte sich auf die Überprüfung *dieser* drei Textsorten der Lesekompetenz, welche nun international in allen beteiligten Ländern erhoben und verglichen wurden. Folglich geben die Ergebnisse auch nur Auskunft über die Lesekompetenz bezogen auf *diese* drei Textformen: Dokumente, Sachtexte und Erzählungen (vgl. ebd., p. 40ff.). Es ist deshalb nicht zulässig, aus den Resultaten Schlüsse über die allgemeine Lesekompetenz von Versuchspersonen zu ziehen – ganz zu schweigen von Schlüssen betreffend die allgemeine *Sprachkompetenz*.²⁸

²⁸ Zum Problem der falschen Folgerungen vgl. auch Abschnitt 2.1.4

Zum Thema der Testkonstruktion soll abschliessend noch die Frage nach der Anwendung der *Bezugsnorm* für die Bewertung von Testresultaten diskutiert werden. Mit anderen Worten: Wie soll festgelegt werden, welche Testresultate als gut und welche als schlecht einzustufen sind? Es lassen sich drei Bezugsnormen unterscheiden (vgl. Rheinberg 2002, p. 59ff.):

- Die *soziale Bezugsnorm*: Die Ergebnisse einer Untersuchung werden miteinander verglichen und es werden Aussagen gemacht wie: Die Versuchsgruppe A schneidet besser ab als die Gruppen B und C.
- Die *individuelle Bezugsnorm*: Die Ergebnisse einer Person werden in einem zeitlichen Längsschnitt miteinander in Beziehung gesetzt. Die Anwendung dieser Bezugsnorm erfordert eine Untersuchung mit mindestens zwei Erhebungszeitpunkten.
- Die *sachliche Bezugsnorm*: Die sachliche Bezugsnorm orientiert sich an einer festgelegten Mindestkompetenz, an einem Standard, der während der Testkonstruktion festgelegt wurde.

Die soziale Bezugsnorm ist dann sinnvoll, wenn es darum geht, die Besten einer Untersuchung herauszufinden. Der individuelle Vergleich dagegen ermöglicht es, den Lernzuwachs einer einzelnen Person festzustellen, während die sachliche Bezugsnorm hilft, herauszufinden, wie viele Personen einem vorgängig festgelegten Standard genügen (vgl. ebd., p. 63ff.). Für die Bedürfnisse heutiger internationaler Schulleistungsvergleiche ist die sachliche Bezugsnorm mit Bestimmtheit am angemessensten. Die Anwendung einer sachlichen Bezugsnorm kommt bei der PISA Studie denn auch am deutlichsten zum Ausdruck. Bei den Veröffentlichungen der Ergebnisse der älteren IEA Lesestudie fällt dagegen auf, dass bei der (graphischen) Darstellung noch oft die soziale Bezugsnorm im Vordergrund steht – möglicherweise Ausdruck dafür, dass zu Beginn der 1990-er Jahre die Bestimmung einer Rangfolge unter den Teilnehmerländern noch wichtiger war als heute.²⁹ Bei der Bestimmung der funktionalen Analphabetinnen und Analphabeten bezieht sich das Forscherteam der IEA Lesestudie aber wieder klar auf eine sachliche Bezugsnorm. Notter umschreibt dies in folgenden Worten: «[...] sie können nicht gut genug lesen, um den minimalen alltäglichen Anforderungen an die Lesekompetenz zu genügen» (Notter et al. 1996, p. 10).

2.1.3.2 Stichprobenauswahl

Bevor man die nach den eben eingeführten Kriterien gebildeten Tests durchführen kann, muss man in einem nächsten Schritt bestimmen, *wen* man befragen bzw. testen will. Da man nur in sehr seltenen Fällen alle Personen befragen kann, welche einen interessieren, muss man fast immer eine Auswahl treffen. Das Festlegen einer solchen Auswahl wird als *Stichprobenbildung* bezeichnet. Die Bildung einer Stichprobe wird, wie auch die Konstruktion eines Tests, gewissen Problemen begegnen. Ein solches Problem ist dasjenige der *Repräsentativität*. Durch die Stichprobenbildung soll ein Teil aus einer Grundgesamtheit ausgewählt werden, der für die Gesamtheit *repräsentativen* Charakter hat. Nun ist es aber so, dass es nicht möglich ist, in einer Untersuchung die Gesamtheit der Merkmale einer Gesellschaft zu repräsentieren. Die meisten Untersuchungen sprechen zwar davon, dass ihre Stichprobe repräsentativ sei, aber sie erklären nicht genauer, was damit gemeint ist bzw. für welche ausgewählten Merkmale ihre Stichprobe repräsentativ ist. So erhoben auch die Schweizer Durchführenden der IEA Lesestudie den Anspruch, eine für die Schweiz repräsentative Stichprobe aller 9- und 14-jährigen Schülerinnen und Schüler zu bilden (vgl. Notter et al. 1996, p. 26). Wie Diekmann in seinem Werk über die Empirische Sozialforschung ausführt, gibt es *die*

²⁹ Selbstverständlich ist hier nur die wissenschaftliche Rezeption angesprochen. Die öffentliche, mediale Verarbeitung der PISA Ergebnisse hat leider deutlich gezeigt, dass die Länder-Rangliste nach wie vor das «attraktivste» Ergebnis darstellt.

ideale Stichprobe aber nicht, da eine Stichprobe nie stellvertretend für sämtliche Merkmalsverteilungen einer Population stehen kann (vgl. Diekmann 1995, p. 368f.). Stichproben sind also immer nur in Bezug auf bestimmte interessierende Dimensionen verkleinerte Abbilder der Grundgesamtheit. Es gilt also, wenigstens eine möglichst genaue *Annäherung* an die Gesamtpopulation zu erzielen. Je besser die Stichprobe die Bevölkerung abbildet, desto aussagekräftiger sind die Interpretationen, die aufgrund der Ergebnisse gemacht werden.

Bei der IEA Lesestudie wurden die Kriterien für die Bildung der Stichprobe durch die IEA vorgegeben. Aufgabe der schweizerischen Gruppe war es dann, anhand dieser Vorgaben die Stichprobe für die Schweiz zu bilden. Um die beste Vergleichbarkeit der Ergebnisse zu erzielen, wurde von der IEA festgelegt, dass alle Versuchspersonen über das gleiche Alter verfügen sollen, gleich lange zur Schule gehen und mit einem ähnlichen Curriculum unterrichtet werden. Für die Stichprobe der Schweiz wurden die Klassen ausgewählt, in denen sich die meisten 9- bzw. 14-Jährigen befinden. Die Schule musste zudem den Anspruch einer vollzeitlichen Schule erfüllen (vgl. Binkley, Rust & Winglee 1995, p. 2f.). Die 9-Jährigen wurden ausgewählt, weil man sich erhoffte, bei ihnen eventuell noch Einflüsse des Erstleseunterrichts festzustellen, die 14-Jährigen weil sie kurz vor Ende der obligatorischen Schulzeit stehen. Obwohl das Einschulungsalter in der Schweiz nicht einheitlich ist, einigte man sich darauf, in allen Regionen der Schweiz die gleiche Klassenstufe zu untersuchen. So wurden die 3. und die 8. Klassen für die Stichprobe ausgewählt.³⁰ Die Schweizer Gruppe musste hier Kompromisse eingehen, da die beiden Anforderungen betreffend Alter und Befragung innerhalb eines Klassenverbandes nicht gleichwertig eingehalten werden konnten. Geht man davon aus, dass alle teilnehmenden Länder gewisse Anpassungen bei der Stichprobenbildung vornehmen mussten, so ergeben sich daraus Einschränkungen bezüglich der internationalen Vergleichbarkeit der Daten. Aus diesem Grund ist es wichtig, genau anzugeben, wie eine Stichprobe tatsächlich gebildet wurde (vgl. Diekmann 1995, p. 368f.). Im Weiteren wurden alle vier Sprachregionen sowie städtische und ländliche Gebiete gleich verteilt ausgewählt. Nur in der rätoromanischen Schweiz beschränkte man sich auf die Untersuchung der 3. Klassen (vgl. Notter et al. 1996, p. 25f.). Die Stichprobe wurde in einer zweistufigen Zufallsauswahl gebildet. In einem ersten Schritt wurden zufällig Gemeinden ausgewählt und danach die Klassen aus diesen Gemeinden gezogen, die in der Untersuchung berücksichtigt werden sollten (3. und 8. Klassen). Nach diesem Verfahren wurde in allen 27 teilnehmenden Ländern die Stichproben gebildet, wobei die IEA die Einhaltung der Vorgaben überprüfte, den nationalen Gruppen in ihrer Arbeit aber auch unterstützend und beratend zur Seite stand (vgl. ebd., p. 26).

2.1.4 Erkenntniswert

Nachdem wir uns mit den Herausforderungen der Testkonstruktion und Testdurchführung auseinandergesetzt haben, soll als nächstes die Frage diskutiert werden, welche Erkenntnisse wir von Schulleistungsvergleichen erwarten dürfen und welche nicht.

In der pädagogischen Forschung kennt man prinzipiell zwei Arten von Studien (Forschungsdesigns): *Querschnittstudien* und *Längsschnittstudien*. Bei den Querschnittstudien wird der Leistungsstand der Schüler zu *einem* bestimmten Zeitpunkt erfasst. Bei den Längsschnittstudien werden die Messungen mehrfach wiederholt, und zwar stets mit genau den gleichen Versuchspersonen. Zusätzlich zu den Aussagen zum Leistungsstand zu einem bestimmten Zeitpunkt lassen sich dadurch auch Lernfortschritte innerhalb bestimmter *Zeiträume* beziffern (vgl. Lange 2001, p. 269).

³⁰ An diesem Beispiel lässt sich eine Eigenart der Stichprobenbildung aufzeigen: Obwohl in den meisten Untersuchungen Individuen im Zentrum des Interesses stehen, werden aus rein praktischen Gründen fast immer (Schul-)Klassen, also *Gruppen* von *Individuen*, befragt. Aus einer rein formalen Sicht werden Stichproben dadurch allerdings stets verfälscht, weil die einzelnen Schülerinnen und Schüler einer Klasse sich nicht unabhängig voneinander verhalten und damit auch stets Effekte dieser Klassen in die Untersuchung einfließen.

Mit anderen Worten: Querschnittstudien können nur sagen, was zu einer bestimmten Zeit *ist*, sie können aber keine Aussagen machen, wie etwas *geworden* ist; dazu braucht es Untersuchungen mit mehreren Erhebungszeitpunkten. Genau dies ist aber bei den allermeisten Vergleichstudien nicht gegeben, da sie reine Querschnittstudien sind. Welche Probleme ergeben sich daraus?

Es sind häufig (zu) grosse Erwartungen an solche Studien gerichtet. Es wird unterstellt, dass die erfassten Lernleistungen die Leistungen in nachschulischer Ausbildung, im Beruf etc. voraussagen können; man erwartet, dass die Verknüpfung von erfassten Leistungen und anderen Variablen auf die Entstehungsbedingungen für diese Leistungen und auf die Handlungsmöglichkeiten zur Leistungsoptimierung schliessen lässt. «Tatsächlich aber können Erwartungen dieser Art von Vergleichsstudien wie TIMSS und PISA kaum erfüllt werden. Angesichts der Untersuchungsanlage solcher Studien sind Aussagen zu prognostischen Beziehungen, Bedingungsbeziehungen und Handlungsmöglichkeiten aus den Daten dieser Studien selber nicht ableitbar: Bei Vergleichsstudien zu Schülerleistungen handelt es sich in der Regel um summative Evaluationen mit querschnittlichem, nicht-experimentellem Design, also einer Designart mit geringer kausaler Aussagekraft» (Pekrun 2002, p. 113). Wir haben damit neben der Komplexität der Zusammenhänge³¹ einen zweiten Grund dafür gefunden, warum Schulleistungsvergleiche keine direkten Aussagen über die Zukunftsaussichten der beteiligten Schülerinnen und Schüler zulassen.

Welche Konsequenzen ergeben sich für zukünftige Forschungen, wenn die querschnittlich-deskriptiven Leistungsstudien nur bedingt zur Aufklärung von Leistungsfolgen, Leistungsbedingungen und Handlungsmöglichkeiten beitragen können? Ziel der Studien sollte in jedem Fall sein, zur Steigerung und Sicherung der Qualität des Bildungswesens beizutragen. Dieses Ziel ist mit Einzelstudien nicht zu erreichen, weshalb es einer systematischen Entwicklung verschiedener Forschungsdisziplinen bedarf, damit einander ergänzende und untereinander vernetzte Erkenntnisse gewonnen werden können. Die zur Zeit leider immer noch häufig anzutreffende Annahme, dass Schulleistungsvergleiche dazu dienen, über die Qualität der Einzelschule Aussagen zu machen und deswegen als direkte Grundlage für Qualitätsverbesserung angesehen werden können, ist unbedingt aufzugeben (vgl. Demmer 2000).

³¹ Vgl. Abschnitt 2.1.1.

2.2 Ergebnisse

2.2.1 Die Lesekompetenz der Schweizer Schülerinnen und Schüler: Nationale Resultate in der dritten und der achten Klasse

Sowohl für die Schülerinnen und Schüler der dritten als auch für diejenigen der achten Klasse wurde untersucht, welche Zusammenhänge zwischen der Lesefähigkeit und einer Reihe von soziodemographischen Daten bestehen. Berücksichtigt wurden im Einzelnen die Faktoren Geschlecht, Alter, Gemeindetyp, sozioökonomischer Hintergrund der Familie, Sprachgewohnheiten und die Zugehörigkeit zu einer Sprachregion. Im Vordergrund steht die Frage nach den Unterschieden zwischen diesen Gruppen sowie nach den möglichen Ursachen für diese Differenzen (vgl. Nieuwenboom 1996, p. 51).

2.2.1.1 Individuelle Merkmale

Bezogen auf das *Geschlecht* lassen sich in beiden untersuchten Altersgruppen nur geringe Differenzen, die zugunsten der Mädchen ausfallen, feststellen: Ein statistisch bedeutsamer Unterschied³² zwischen den Geschlechtern besteht nur bei der Textart «Erzählung». Durchschnittlich haben die Mädchen der achten Klasse in diesem Bereich etwa eine Testfrage mehr richtig gelöst als die Knaben (vgl. Nieuwenboom 1996, p. 52f.). Dieses bessere Resultat kann sich im Ergebnis der Gesamtleseleistung aber nicht niederschlagen; dort bestehen keine signifikanten Unterschiede hinsichtlich Geschlecht.

Die *Alterszusammensetzung* in einer Klasse unterscheidet sich je nach Sprachregion. Dabei spielen neben Repetitionen und Rückstellungen auch die unterschiedlichen Einschulungszeitpunkte in den Sprachregionen der Schweiz eine Rolle. Während im Kanton Tessin die Einschulung mit 6 Jahren erfolgt, vollzieht sie sich in der deutschen und der rätoromanischen Schweiz erst mit 7 Jahren, und in der französischen Schweiz wird mit 6 *oder* 7 Jahren eingeschult (vgl. ebd., p. 53ff.).³³

Tendenziell nehmen die Werte der Leseleistungen aber in allen vier Sprachregionen innerhalb der dritten Klasse mit zunehmendem Alter ab. Dasselbe gilt auch in der achten Klasse: Wer ein Jahr oder mehr älter ist als die regulären Klassenkameradinnen und -kameraden, zeigt eine signifikant schlechtere Leseleistung als diejenigen, die im Regelalter sind. Diese Kinder sind oft Repetenten oder wurden altersversetzt eingeschult, weil sie fremdsprachig und neu zugezogen sind. Diese Einschränkungen erklären den gemessenen Einfluss des Alters auf die Leseleistung weitgehend.

2.2.1.2 Familienmerkmale

Ein Zusammenhang zwischen dem *sozioökonomischen Status* (hier erhoben als persönliches Eigentum an Gütern³⁴ und Ausbildungsstand der Eltern), in dem ein Kind aufwächst, und seiner Lesekompetenz wurde schon lange vor der IEA Lesestudie nachgewiesen (vgl. Thorndike 1973) und konnte durch die Ergebnisse der vorliegenden Studie erneut bestätigt werden. Sowohl für die

³² Als «signifikant» oder eben als «statistisch bedeutsam» gilt das Ergebnis einer Hypothesenprüfung, wenn die Annahme berechtigt ist, dass ein theoretisch angenommener und in den Daten vorgefundener Zusammenhang zwischen Merkmalen oder Unterschied zwischen Gruppen nicht alleine durch die zufällige Unschärfe erklärt werden kann, die mit der Stichprobenziehung verbunden ist.

³³ Diese Angaben beziehen sich auf den Erhebungszeitraum der IEA Lesestudie.

³⁴ Um persönliches Eigentum an Gütern zu ermitteln, wurden die Schüler/innen danach gefragt, wie viel Büchergestelle, Fernsehapparate, Computer sie zu Hause haben und ob sie ein eigenes Zimmer haben. Diese Operationalisierung des sozioökonomischen Status mag hölzern erscheinen, hat sich aber in der Vergangenheit durchaus bewährt.

Dritt- als auch für die Achtklässlerinnen und -klässler wurde übereinstimmend gezeigt, dass mit zunehmend höherem sozioökonomischem Status die Leseleistung des Kindes zunimmt. Generell ergab sich bei sämtlichen (!) Textarten ein statistisch signifikanter Unterschied zwischen den Leseleistungen der Kinder der unteren sozioökonomischen Schicht im Vergleich zu der mittleren Schicht. Die Differenzen zwischen der mittleren und der oberen Schicht sind statistisch gesehen ebenfalls bedeutsam, aber etwas weniger gross (vgl. Nieuwenboom 1996, p. 56f.).

Neben dem sozioökonomischen Status wird das soziale Umfeld des Kindes durch weitere kulturelle Faktoren mitbestimmt. So wurde etwa vermutet, dass die *Sprachgewohnheiten* der Kinder einen Einfluss auf die Leseleistung haben könnten. Die Untersuchung konzentriert sich dabei auf die Frage, «inwiefern sich Unterschiede zwischen der Sprache bzw. den Sprachen, welche die Schüler/innen am häufigsten sprechen, und der Sprache, in der sie in der Schule unterrichtet werden (bezeichnet als Unterrichtssprache) auf die Leseleistung auswirken» (ebd., p.57).

Sowohl die Schülerinnen und Schüler der dritten als auch diejenigen der achten Klasse wurden nach ihren Sprachgewohnheiten in drei Gruppen aufgeteilt. Diejenigen Schülerinnen und Schüler, die nur die Unterrichtssprache sprechen, wurden als *unterrichtssprachig* bezeichnet.³⁵ Die Schülerinnen und Schüler, die neben der Unterrichtssprache auch eine andere Sprache regelmässig sprechen, galten als *unterrichts- und anderssprachig*. Als rein *anderssprachig* wurden diejenigen Kinder bezeichnet, welche neben der Schule nur eine Fremdsprache sprechen.

Die Ergebnisse der unterrichtssprachigen Kinder liegen bei allen drei Textarten leicht höher als diejenigen der unterrichts- und anderssprachigen Schülerinnen und Schüler. Deutlich heben sich die anderssprachigen Schülerinnen und Schüler ab. Sie schneiden am schlechtesten ab, die Unterschiede zu den beiden anderen Gruppen fallen deutlich aus (vgl. ebd., p. 57f.). Dieses generelle Ergebnis gilt für beide untersuchten Klassenstufen.

Es gibt gute Gründe anzunehmen, dass die beiden Dimensionen sozioökonomischer Status und Sprachgewohnheit miteinander verknüpft sind und dass sich daher die Realität komplexer gestaltet, als die Betrachtung der Effekte von sozioökonomischem Status und Sprachgewohnheiten *je für sich alleine* vermuten liesse.

Für die Drittklässlerinnen und -klässler wurde diese Annahme geprüft. Gemäss den Ergebnissen dieser Prüfung sind die *anderssprachigen* Kinder in der unteren sozioökonomischen Schicht am stärksten vertreten. Die meisten Kinder, die zuhause nur die Unterrichtssprache sprechen, gehören dagegen der oberen Schicht an. In diesem Sinne kann man von einem Zusammenhang zwischen den beiden Faktoren sprechen. Es stellt sich nun also die Frage, inwiefern sich Leseleistungsunterschiede mit Sprachgewohnheiten und dem sozioökonomischen Status erklären lassen.

Abbildung 2.3 zeigt die Einflüsse von Sprachgewohnheiten und sozioökonomischem Status auf die Leseleistung. Die Ergebnisse weisen darauf hin, dass in der unteren und in der mittleren Schicht deutliche Unterschiede zwischen den unterrichtssprachigen, den unterrichts- und anderssprachigen und den rein anderssprachigen Schülerinnen und Schüler bestehen. Diese Differenzen sind statistisch bedeutsam. In der oberen sozioökonomischen Schicht lassen sich die kleinsten Unterschiede zwischen den Sprachgewohnheiten feststellen (auf der Abbildung 2.3 mit einem Kreis und einer 1 markiert). So kann davon ausgegangen werden, dass sich in der oberen sozioökonomischen Schicht Fremdsprachigkeit nicht unbedingt als Nachteil in Bezug auf die Lesekompetenz auswirken muss.

³⁵ Streng genommen sprechen die Deutschweizer Kinder zu Hause natürlich nicht die Unterrichtssprache, sondern eine davon abweichende Mundart. Die vorliegende und auch andere Untersuchungen haben aber gezeigt, dass den Deutschschweizer Kindern aus der Tatsache, in der Schule zuerst einmal eine von ihrer Muttersprache abweichende Unterrichtssprache erwerben zu müssen, keine Nachteile hinsichtlich Lesekompetenz erwachsen, weshalb die Kinder aus der Deutschschweiz durchaus als unterrichtssprachig in obigem Sinne bezeichnet werden dürfen.

Dass die Kinder in der unteren Schicht nicht grundsätzlich die schlechtesten Leseleistungen erbringen, zeigt der Vergleich zwischen den unterrichtssprachigen Kindern der unteren Schicht mit den anderssprachigen Schülerinnen und Schülern der mittleren Schicht. Der Unterschied fällt deutlich zugunsten der unterrichtssprachigen Kinder aus und ist statistisch bedeutsam (auf der Abbildung 2.3 mit einer 2 markierte Linie).

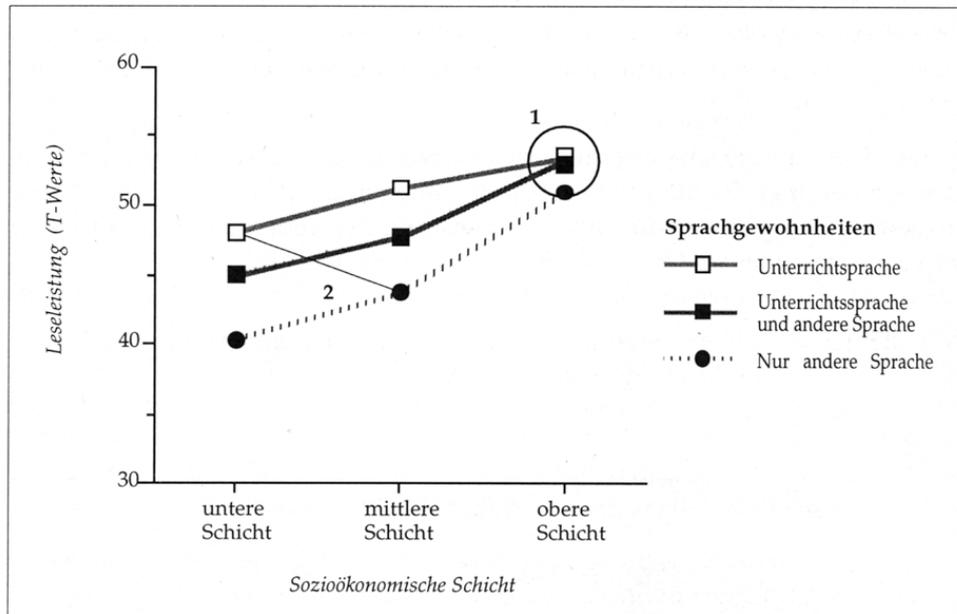


Abbildung 2.3: Der Einfluss von Sprachgewohnheiten und sozioökonomischem Status auf die Gesamtleseleistung (aus: Nieuwenboom 1996, p. 59)

Bei der Auswertung der Daten der Erhebung bei den Achtklässlerinnen und Achtklässlern ging man sogar noch weiter. Hier wurde nicht nur das Zusammenspiel des sozioökonomischen Status mit den Sprachgewohnheiten im Hinblick auf die Leseleistung der Schülerinnen und Schüler untersucht, sondern es wurden weiter auch die Faktoren Schulniveau³⁶ und Anteil Fremdsprachiger pro Klasse einbezogen. Dies ist deshalb sinnvoll, weil (a) fremdsprachige Schülerinnen und Schüler häufiger Schulen mit Grundansprüchen besuchen als nicht-fremdsprachige Schülerinnen und Schüler und (b) weil die fremdsprachigen Schülerinnen und Schüler sich in gewissen Klassen konzentrieren.

Es wurde also überprüft, ob sich die Unterschiede in den Leseleistungen der Schülerinnen und Schüler mit unterschiedlichen Sprachgewohnheiten durch die anderen Merkmale, nämlich Schulniveau, sozioökonomischer Status und Anteil Fremdsprachiger pro Klasse, erklären lassen. Mit anderen Worten wurde die Frage untersucht, ob der Zusammenhang zwischen Sprachgewohnheiten und Leseleistung ein reiner Scheinzusammenhang ist. Dafür wurde mit einem geeigneten Verfahren jener Anteil der Leseleistungsunterschiede zwischen den drei Schülerinnen- und Schüler-Gruppen neutralisiert bzw. kontrolliert, der auf das Schulniveau, den sozioökonomischen Status und auf die Klassen-Zusammensetzung zurückzuführen ist (vgl. Abbildung 2.4).

³⁶ Bei der Variable Schulniveau wurden drei Arten von Schulen unterschieden: 1) Schulen mit Grundansprüchen (Primarschule, Realschule); 2) Schulen ohne Selektion (private Schulen, z.B. «Steiner Schule»); 3) Schulen mit erweiterten Ansprüchen (Sekundar-, Bezirksschule, Gymnasium).

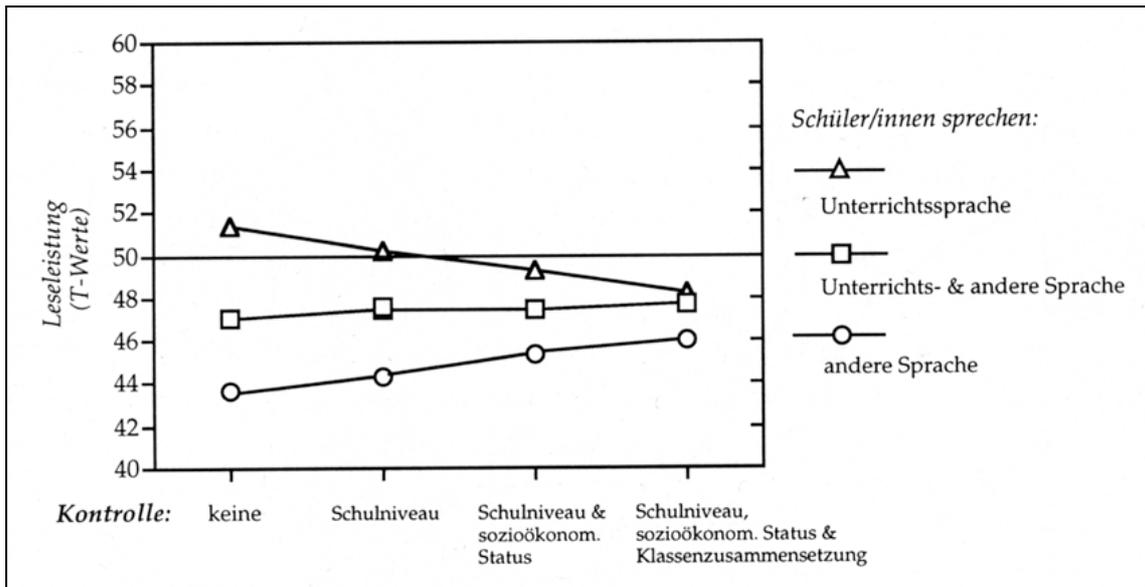


Abbildung 2.4: Unterschiede in den Leseleistungen aufgrund der Sprachgewohnheiten vor und nach Kontrolle von Schulniveau, sozioökonomischem Status und Klassenzusammensetzung (aus: Rüesch 1996, p. 92)

Das Ergebnis war folgendes: Die vergleichsweise schwachen Lesekompetenzen in der Unterrichtssprache von fremdsprachigen Schülerinnen und Schülern lassen sich im Wesentlichen durch das tiefere Schulniveau, den niedrigen sozioökonomischen Status und den Besuch der Klassen mit hohem Anteil fremdsprachiger Schülerinnen und Schüler erklären. Dabei kommt dem Faktor Schulniveau am meisten Erklärungswert zu, gefolgt vom Faktor sozioökonomischer Status und dem Faktor Fremdsprachigenanteil pro Klasse. Die Tatsache, dass diese Schülerinnen und Schüler zuhause eine andere Sprache sprechen, kann also nicht die eigentliche Ursache sein für deren schlechtes Abschneiden bei der Lesekompetenz. Viel grössere Zusammenhänge zeigen sich zu den drei anderen genannten Faktoren.

2.2.1.3 Regionale Merkmale

Die Autoren der IEA Lesestudie gingen davon aus, dass sich durch die unterschiedlichen Bevölkerungszusammensetzungen in verschiedenen *Gemeindetypen* Unterschiede in der Lesekompetenz zwischen den Gemeinden ergeben würden.

Diese These wurde bei den Drittklässlerinnen und -klässlern geprüft. Tendenziell zeigten die ländlichen Gemeinden bessere Ergebnisse als die städtischen. Diese Differenz ist aber nicht sehr gross und es muss berücksichtigt werden, dass die Schweiz im internationalen Vergleich eine Sonderstellung einnimmt. In den meisten anderen Ländern haben sich nämlich für die ländlichen Gemeinden tiefere Resultate ergeben als für die städtischen. Die Besonderheiten der Schweiz liegen vermutlich darin, dass in der Stadt wie auch auf dem Land das Wohlstandsniveau sehr hoch liegt. So umfasst die Kategorie der ländlichen Gemeinden neben den agrarisch orientierten auch touristische und so genannte reiche Gemeinden (Beispiel Gemeinden des rechten Zürichseeufers). Deren Bevölkerung weist meist einen recht hohen sozioökonomischen Status auf. Bezüglich der Ressourcen (z.B. Schulbibliotheken) von städtischen und ländlichen Schulen lassen sich in der Schweiz ebenfalls keine bedeutsamen Unterschiede feststellen (vgl. Nieuwenboom 1996, p. 61f.).

Im Weiteren wurden auch die Ergebnisse der vier *Sprachregionen* der Schweiz miteinander verglichen. Dabei zeigten sich bei den Drittklässlerinnen und -klässlern kaum bedeutende Unterschiede zwischen den Sprachregionen. Einzige die rätoromanisch sprechenden Schülerinnen

und Schüler schnitten deutlich schlechter ab als der Rest der Schweizer Schülerinnen und Schüler. Dieses schlechte Abschneiden lässt sich wesentlich (wenn auch nicht ganz) auf soziokulturelle Merkmale zurückführen. Möglicherweise ist auch die Stellung von Bedeutung, welche die rätoromanische Sprache in der Gesellschaft einnimmt (vgl. Nieuwenboom 1996, p. 76).

Etwas anders sieht die Situation bei den Achtklässlerinnen und -klässlern aus. Die Leseleistung wurde hier nur in drei Sprachregionen – italienische, französische und deutsche Schweiz – verglichen. Laut Gesamttestergebnis haben die Schülerinnen und Schüler aus der deutschen Schweiz die besten Resultate gezeigt, gefolgt von den Schülerinnen und Schülern aus der italienischen Schweiz und am schlechtesten waren die Ergebnisse in der französischen Schweiz (vgl. Ruesch 1996, p. 82).

Was die einzelnen Textarten betrifft, so haben bei den Dokumenten die Schülerinnen und Schüler aus der deutschsprachigen Region am besten abgeschnitten, gefolgt von den Schülerinnen und Schülern aus der französischsprachigen Region und die schlechtesten Ergebnisse zeigten in dieser Kategorie die Schülerinnen und Schüler aus der italienischsprachigen Region. Bei den Sachtexten waren die Schülerinnen und Schüler aus der deutschsprachigen und aus der italienischsprachigen Region gleich gut, die Resultate der Schülerinnen und Schüler aus der französischsprachigen Region lagen aber tiefer. Bei der Textart Erzählungen sind keine bedeutsame Unterschiede zwischen den drei Sprachregionen festgestellt worden (vgl. ebd., p. 83).

2.2.1.4 Der Einfluss des Schultyps auf die Leseleistung

Im Hinblick auf den Einfluss des Schulniveaus auf die Leseleistung der Schülerinnen und Schüler wurden nur die Schülerinnen und Schüler aus der deutschen und aus der französischen Schweiz verglichen, weil in der italienischen Schweiz die Schülerinnen und Schüler aus den Schulen mit Grund- und erweiterten Ansprüchen sehr kleine Gruppe bildeten.

Drei wichtige Ergebnisse sind zu vermerken: Erstens nimmt sowohl in der deutschen als auch in der französischen Schweiz die Leseleistung mit steigendem Schulniveau deutlich zu. Zweitens zeigen die Achtklässlerinnen und -klässler aus der deutschen Schweiz höhere Leseleistungen als diejenigen aus der französischen Schweiz. Dies gilt sowohl bei den Schulen mit Grundansprüchen als auch bei den Schulen mit erweiterten Ansprüchen. Bei den Schulen ohne Selektion haben dagegen die Schülerinnen und Schüler aus der französischen Schweiz besser abgeschnitten. Drittens – und das relativiert die ersten beiden Ergebnisse etwas – ist die Varianz *zwischen* den Sprachregionen deutlich kleiner als die Varianz *innerhalb* der Sprachregionen. Das bedeutet mit anderen Worten: Das Schulniveau ist für die Unterschiede in den Leseleistungen massgebender als die Zugehörigkeit zu einer Sprachregion (vgl. Nieuwenboom 1996, p. 84).

2.2.1.5 Zusammenfassung

Zu den Ergebnissen lässt sich zusammenfassend Folgendes festhalten: Die Untersuchungsgruppen wurden in verschiedene Untergruppen aufgeteilt und nach individuellen, familiären und regionalen Merkmalen miteinander verglichen. Dabei konnten zwar mehrfach Unterschiede festgestellt werden, diese fallen aber allgemein gesehen nicht sehr gross aus. Zwei Ergebnisse stehen allerdings hervor:

Bedeutsame Unterschiede zeigen sich in beiden Alterklassen im Zusammenhang zwischen dem *sozioökonomischen Status* und der Lesekompetenz. Je höher die sozioökonomische Schicht, desto besser fallen die Leseleistungen aus. Diese Unterschiede sind für alle drei Textarten sowie für das Gesamtergebnis bedeutungsvoll und unterstreichen die Bedeutung des «familiären Leseklimas».

Die Ergebnisse verweisen zudem auf einen engen Zusammenhang zwischen dem sozioökonomischen Status, der Lesekompetenz und den *Sprachgewohnheiten* eines Kindes. Sprachgewohnheit

und sozioökonomischer Status haben einen starken Einfluss auf die Leseleistung. Wichtig ist in diesem Zusammenhang die Feststellung, dass beide Faktoren durch die familiäre Umgebung vorgegeben werden und die Schule nur bedingt einen Einfluss auf sie ausüben kann.

In internationalen Vergleich schneiden die Schweizer Schülerinnen und Schüler relativ gut ab. Die Drittklässlerinnen und -klässler fallen zwar noch zurück, holen ihre Defizite später aber offensichtlich problemlos auf. Dennoch sollte sich die Schweiz auf diese Resultate nichts einbilden, denn das Beispiel Finnland zeigt deutlich, dass weitere Verbesserungen möglich sind.

Nachdenklich stimmt zudem der Befund, dass in der Schweiz die Schere zwischen Schulen mit Grundansprüchen und solchen mit erweiterten Ansprüchen offensichtlich besonders weit aufgeht: Während unsere besten Schülerinnen und Schüler auch international zu den besten gehören, stehen die weniger guten Schülerinnen und Schüler sowohl national als auch international schlecht da.

2.3 Folgen und Konsequenzen

Die IEA Lesestudie hat im Vergleich zu den viel beachteten Studien TIMSS und PISA (vgl. nachfolgende Kapitel 3 und 4) kaum Wellen geworfen – um nicht zu sagen: ihre Resultate wurden ignoriert. Eine Recherche zu Zeitungsartikeln aus der Schweiz, welche sich mit der Studie befassen, brachte einen einzigen Treffer (Tages-Anzeiger vom 19. Juli 1996 unter der Überschrift «Lesestudie: Schweiz im vorderen Mittelfeld»). Auch Autoren aus Deutschland kommen zu diesem Schluss. Tillmann (2001, p. 359) etwa schreibt:

«Sofern sich die Bundesrepublik an [...] internationalen Vergleichen beteiligt hat, sind die Ergebnisse über längere Zeit unaufgeregt oder auch gar nicht diskutiert worden: So hat es zum Beispiel noch 1994 eine internationale Vergleichsstudie zur Lesefähigkeit gegeben [...], die in diesem Land niemanden interessiert hat. Wir haben es bei TIMSS also nicht mit einer neuen Sorte von Studie zu tun – sondern mit einer völlig anderen öffentlichen Reaktion auf solche Studien.»

Über die Gründe für dieses fehlende Interesse kann an dieser Stelle nur spekuliert werden. Denkbar sind aber verschiedene Ursachen:

Die Pädagogik im deutschsprachigen Raum war bis vor kurzer Zeit weitgehend normativ besetzt, d.h. vorherrschend war eine geisteswissenschaftliche Auffassung von Pädagogik, in welcher gedacht wurde, was sein sollte, statt dass gemessen wurde, was ist. Erst in jüngerer Zeit setzte sich die Auffassung der Pädagogik als *empirische* Wissenschaft auch im deutschen Sprachraum durch.

In eine ähnliche Richtung zielt das Argument von Baumert und Köller, wonach es in der BRD «keine Tradition der dauerhaften Beobachtung von Erträgen institutionalisierter Bildungsprozesse» gibt, weil «ein Hauptinteresse der Erziehungswissenschaft [...] in den letzten 10 bis 15 Jahren [...] auf der Entwicklung und Erprobung von Modellen zur Optimierung der Arbeit in Einzelschulen und dem Entwurf didaktischer Modelle und deren Einführung in die Unterrichtspraxis [lag]» (Baumert & Köller 1998, p. 12). Mit anderen Worten: Die Erziehungswissenschaft kümmerte sich mehr um die Optimierung von Prozessen und weniger um die Überprüfung eines Outputs.

Das Bildungswesen war (und ist es wohl bis heute teilweise noch) eine rein nationale Domäne. Es bestand daher auch kein Bedarf bzw. Interesse an *internationalen* Vergleichen. Helmke diagnostiziert denn für Deutschland diesbezüglich auch einen «bemerkenswerte[n] Wandel» (Helmke 2001, p. 155) – welcher allerdings erst durch den so genannten TIMSS-Schock ausgelöst wurde.³⁷

Im Gegensatz zu TIMSS und PISA wurde bei der Veröffentlichung und Verbreitung der Resultate im Fall der IEA Lesestudie noch dilettantisch vorgegangen. Auf unsere Frage, wie die Resultate der

³⁷ Vgl. Kapitel 4.

Studie einer interessierten Öffentlichkeit zugänglich gemacht worden seien, haben die kantonalen Geldgeber der Studie zwar unterschiedlich, aber stets etwas hilflos reagiert. Gemäss einer Auskunft der Erziehungsdirektion des Kantons Bern, habe man direktionsintern die Ergebnisse in einem Kolloquium diskutiert. Ein Bericht über die kantonalen Ergebnisse sei hingegen nicht erschienen. Die Ergebnisse der IEA Lesestudie waren möglicherweise auch zu «normal». Die Schweiz schnitt ja nicht so schlecht ab, weshalb sich auch die Medien nicht gross dafür interessierten.

Der Hinweis auf diese möglichen Gründe muss an dieser Stelle genügen. Ausführliche Informationen zum Thema Rezeption internationaler Vergleichsuntersuchungen sind in Kapitel 4 zu finden. Zur Ehrenrettung der IEA Lesestudie aber noch dies: Auch wenn ihr selbst jenseits einer engen Fachöffentlichkeit die Aufmerksamkeit verweigert worden ist, so hat sie doch das Feld für ihre Nachfolgerinnen geöffnet. Sie hatte wohl die Funktion, die (Fach-)Öffentlichkeit auf die nachfolgenden Studien vorzubereiten. Zudem konnte sich an dieser Studie mancher Forscher sein Rüstzeug für spätere Untersuchungen erwerben. Es ist bestimmt kein Zufall, dass mancher Name, der schon eng mit der (schweizerischen) IEA Lesestudie verknüpft ist, auch unter TIMSS und PISA wieder auftaucht.

3 Internationale Schulleistungsvergleiche am Beispiel der TIMSS (Third International Mathematics and Science Study)

3.1 Generelles

1995 haben sich in der Schweiz knapp 20'000 Schülerinnen und Schüler der Sekundarstufen I und II einem Leistungstest in Mathematik und Naturwissenschaften unterzogen. Ziele dieses Tests waren der internationale Vergleich des Wissens in Mathematik und Naturwissenschaften und der Vergleich von Bildungssystemen. Aufgrund von systematisch gewonnenen Daten wurde eine Bewertung von Leistungen einzelner Schülerinnen und Schüler vorgenommen und mit dem Bildungssystem in Zusammenhang gebracht. Gleichzeitig wurden auch Hintergrundinformationen über den Unterricht, die Lehrer, die Schulen sowie Aspekte der ausserschulischen Lebensumwelt der Schülerinnen und Schüler gewonnen. Weltweit nahm mehr als eine halbe Million Schülerinnen und Schüler aus rund 15'000 Schulen in 46 Ländern an der Studie teil.

Die Studie wird seither alle vier Jahre durchgeführt, Deutschland, Österreich und die Schweiz haben sich aber nur an TIMSS 1995 beteiligt.

3.1.1 Einleitung

Ein grosser Teil der jährlichen öffentlichen Staatsausgaben fliesst in das Bildungswesen. Das Interesse an Schulleistungsvergleichen ist also verständlich. Bildung wird in der Schweiz oft als Ersatz für fehlende Rohstoffe bezeichnet, und das Bildungsniveau der Individuen kann als Indikator für das Humankapital unserer Gesellschaft betrachtet werden, das grosse Auswirkungen auf die wirtschaftliche Entwicklung hat. Bildung ist als Voraussetzung für den Wohlstand eines Landes allgemein anerkannt.

Mit der zunehmenden internationalen Verflechtung von Wirtschaft und Politik, aber auch von Bildungssystemen und Bildungsforschung, haben sich internationale Schulleistungsvergleiche etabliert. Sie ermöglichen, verschiedene Variablen zu untersuchen, die sich innerhalb eines Landes weitgehend gleichen und deren Manipulation kaum möglich wäre. Indem man sich die natürliche Variabilität von Faktoren wie «Alter bei Schuleintritt», «Klassengrösse» oder «Curriculum» in den verschiedenen Ländern zunutze macht, können mögliche Einflüsse von Alternativen untersucht werden (vgl. Moser, Ramseier, Keller & Huber 1997; Ramseier, Keller & Moser 1999; IEA 1990).

Auf internationaler Ebene wurde TIMSS von der internationalen Vereinigung IEA (International Association for the Evaluation of Educational Achievement) durchgeführt. In der Schweiz wurde die Erhebung vom Amt für Bildungsforschung des Kantons Bern geleitet.

3.1.2 Forschungslogik und ihre Probleme

Die Forschungslogik und ihre Probleme bei internationalen Schulleistungsvergleichen wurden bereits anhand der IEA Lesestudie (vgl. Kapitel 2) ausführlich diskutiert. Dieser Abschnitt zeigt für die TIMSS auf, wie die Testkonstruktion erfolgte und welche Populationen international und national – am Beispiel der Schweiz – einbezogen wurden.

3.1.2.1 Was wird verglichen?

TIMSS zielt auf den internationalen Vergleich von Bildungssystemen anhand der Leistungen von Schülerinnen und Schülern in den Bereichen «Mathematik» und «Naturwissenschaften» ab. Sie untersuchte dabei drei verschiedene Populationen: Schüler und Schülerinnen der Grundstufe (Population 1), der Sekundarstufe I (Population 2) und der Sekundarstufe II (Population 3). Die Schweiz beteiligte sich an den Vergleichsstudien der Sekundarstufen I und II.

Mit der Fokussierung auf Mathematik und Naturwissenschaften steht TIMSS in einer Reihe mit vielen weiteren nationalen und internationalen Vergleichen, die sich auf diesen Bereich konzentrieren. Das ist verständlich, weil es in diesen Fächern möglich ist, die Schulleistung relativ ökonomisch, d.h. mit «paper and pencil» Tests zu erfassen. In sprachlichen Fächern hingegen würde ein solcher Test nur ein schmales Spektrum des Curriculums abbilden. Die curriculare Validität würde also durch ausschliesslich schriftlich durchgeführte Tests geschmälert. Trotzdem ist die Konzentration auf einen Teilbereich fachlicher Leistung unbefriedigend.

Ein weiterer oft genannter Kritikpunkt ist, dass sich «Erfolg» und «Misserfolg» des Unterrichts nicht anhand einer auf blosser Fachleistung beschränkte Untersuchung bemessen lässt. Neben der Vermittlung einer soliden Wissensbasis zeichnet sich erfolgreicher Unterricht unter anderem auch dadurch aus, dass positive erzieherische Wirkungen erzielt und Schlüsselkompetenzen aufgebaut werden. Aber auch solche Variablen sind durch «paper and pencil»-Verfahren sehr schwer zu erfassen (vgl. Helmke 2000, p. 143).

3.1.2.2 Mehrdimensionalität des Untersuchungsgegenstandes

Sind die Gegenstände des Interesses einmal festgelegt, geht es darum, sie zu operationalisieren, d.h. sie für die Beobachtung zugänglich zu machen. Dabei wird eine gut beobachtbare Variable (z.B. Punktzahl in einem Mathematiktest) ausgewählt, die mit der nicht direkt beobachtbaren Variablen (z.B. Leistungsniveau in der Mathematik) verknüpft ist. Die beobachtbare Variable dient so als «Indikator» für die nicht direkt beobachtbare Variable. Eine Operationalisierung umfasst also immer hypothetische Annahmen über Zusammenhänge.

Nationale Bildungssysteme unterscheiden sich in vielfältiger Weise: im Zeitpunkt der Einschulung, in der Dauer der Ausbildung, in den unterrichteten Fächern, in den Lehrplänen und Lernzielen der Fächer, in materiellen und personellen Ressourcen usw. Um fair zu sein, muss ein internationaler Schulleistungsvergleich einerseits dieser Unterschiedlichkeit gerecht werden und andererseits ein systematisches, einheitliches Verfahren anwenden, das einen aussagekräftigen Vergleich zulässt. Zentrale Elemente dafür sind eine klar definierte Stichprobe und ein einheitliches Instrument zur Messung der Schulleistung (vgl. Moser et al. 1997). Zunächst soll auf die Konstruktion des Messinstruments eingegangen werden.

Bei der Operationalisierung der Leistungsniveaus in Mathematik und Naturwissenschaften für die TIMSS stellte sich also die Frage, welche Test-Items den Leistungsstand der Schülerinnen und Schüler in den verschiedenen Ländern möglichst genau abbilden. Wenn man bedenkt, dass sich bereits Lehrpläne und Lehrmittel in der Schweiz von Kanton zu Kanton unterscheiden, ist leicht einzusehen, dass die Entwicklung von leistungsmessenden Aufgaben, die dem Unterricht von Schülerinnen und Schülern aus beinahe 50 Ländern entsprechen sollen, ein schwieriges Unterfangen ist. Bei der TIMSS hatten alle teilnehmenden Länder die Gelegenheit, Aufgaben für den Test zu entwickeln. Dabei war immer auch zu berücksichtigen, dass die Aufgaben für die einzelnen Länder lösbar und relevant sein mussten; naturwissenschaftliche Aufgaben über Gletscher würden zum Beispiel tropische Länder benachteiligen. In den Jahren 1993 und 1994 wurden anschliessend Pilot-Tests durchgeführt, bei denen es hauptsächlich um die Erprobung und Auswahl von Testaufgaben ging, die einen über die Länder hinweg fairen Vergleich ermöglichten. Trotzdem muss bei allen Interpretationen des Vergleichs klar gemacht werden, dass sie sich ausschliesslich auf ein

«internationales» Curriculum beziehen, das zwar nicht wirklich existiert, welches aber durch den Test implizit umschrieben wird. Weicht das Curriculum eines Landes sehr stark von diesem internationalen Curriculum ab, kann von den Testergebnissen nicht auf mangelhafte Leistungen der Schülerinnen und Schüler geschlossen werden (vgl. Moser et al. 1997, p. 16).

Eine zweite Einschränkung für die Interpretation der Ergebnisse ergibt sich aus dem Fakt, dass der Grossteil des Tests aus Multiple-Choice-Aufgaben bestand. Viele Arten von Wissen und Fertigkeiten, die im Unterricht vermittelt werden, können nicht mit diesem Verfahren geprüft werden. So kann zum Beispiel die Fähigkeit des Experimentierens in den Naturwissenschaften und in der Mathematik auf diese Weise nur ungenügend gemessen werden. Aufgrund der ökonomischen Vorteile eignen sich Multiple-Choice-Aufgaben jedoch besonders für grosse Untersuchungen (vgl. Ramseier et al. 1999, p. 48).

Die Entwicklung der Leistungstests dauerte über drei Jahre. Auf der Sekundarstufe I wurden 286 Aufgaben eingesetzt. In der Mathematik wurden die Bereiche Zahlen und Zahlenverständnis, Geometrie, Messen und Masseinheit, Darstellung und Analyse von Daten, Wahrscheinlichkeitsrechnung, Proportionalität und Algebra geprüft. In den Naturwissenschaften stammten die Aufgaben aus der Erdkunde, der Biologie, der Physik und der Chemie. Die Aufgaben liessen sich den drei kognitiven Anforderungsstufen «Wissen», «Anwenden» und «Problemlösen» zuordnen (vgl. Internetseite des Max-Planck-Instituts: Testaufgaben Mathematik TIMSS 7./8. Klasse³⁸).

Auf der Sekundarstufe II wurde mit 76 Aufgaben mathematisches und naturwissenschaftliches Grundwissen geprüft. Zusätzlich zum Grundwissen wurden in Subgruppen innerhalb der Sekundarstufe II die Leistungen in Mathematik und Physik auf gymnasialem Niveau erhoben (vgl. Moser & Notter 2000).

3.1.2.3 Wen soll man untersuchen?

Bei gross angelegten Studien wird selten die gesamte Population untersucht, über welche man Aussagen machen möchte. Meist wird eine Stichprobe zusammengestellt, von der auf die Population geschlossen werden kann. Damit sich aber aufgrund von Stichprobenergebnissen Schlüsse auf die Population ziehen lassen, muss die Stichprobe «repräsentativ» sein; d.h. sie muss in ihrer Zusammensetzung der Population entsprechen.

Möchte man also beispielsweise das Wissen von Schülerinnen und Schülern auf der Sekundarstufe I vergleichen, muss definiert werden, aus welcher Population die Stichprobe zusammengestellt werden soll. Sollen alle Dreizehnjährigen untersucht werden? Daraus würden sich insofern Verzerrungen ergeben, als in einem Schulsystem die meisten Dreizehnjährigen erst in der 7. Klasse wären, während in einem andern Schulsystem die meisten Dreizehnjährigen bereits die 8. Klasse besuchten. Soll man stattdessen alle Siebtklässler in die Untersuchung einbeziehen? Hier stellt sich das Problem, dass die Siebtklässler je nach Schulsystem ein unterschiedliches Alter aufweisen.

Bei der TIMSS hat man in die Definition der Populationen sowohl den Faktor des Alters als auch den Faktor der Klassenzugehörigkeit mit einbezogen. Hier soll die Festlegung der Populationen 2 und 3 erläutert werden, mit welchen sich die Schweiz an der Studie beteiligte:

Population 2 (Sekundarstufe I)

In die Untersuchung der Sekundarstufe I wurden jene beiden Schuljahre einbezogen, in denen die meisten Dreizehnjährigen unterrichtet werden. Dies betraf in den meisten Ländern das 7. und 8.

³⁸ Internetseite des Max-Planck-Instituts für Bildungsforschung > Testaufgaben Mathematik TIMSS 7./8. Klasse: http://www.mpib-berlin.mpg.de/TIMSSII-Germany/Die_Testaufgaben/TIMSSII-Math.pdf [Februar, 2010].

Schuljahr. Einige wenige Länder haben eine spätere Einschulung, insbesondere Skandinavien sowie die Deutschschweiz, da wurden die Schuljahre 6 und 7 in die Untersuchung einbezogen.

In den drei Sprachregionen der Schweiz werden die Kinder zu verschiedenen Zeitpunkten eingeschult. In der deutschsprachigen Schweiz erfolgt die Einschulung im Alter von sieben Jahren, fast ein Jahr später als in der französisch- und italienischsprachigen Schweiz, wo bereits mit sechs Jahren eingeschult wird. Die unterschiedlichen Einschulungszeitpunkte hatten zur Folge, dass für die italienisch- und französischsprachige Schweiz Schülerinnen und Schüler des 7. und 8. Schuljahres, für die Deutschschweiz aber Schülerinnen und Schüler des 6. und 7. Schuljahrs einbezogen wurden. Die unterschiedlichen Einschulungszeitpunkte erschweren den Vergleich zwischen den Ländern, aber auch zwischen den Regionen der Schweiz.

In der Schweiz bestand die Stichprobe für die Sekundarstufe I aus über 13 000 Schülerinnen und Schülern (vgl. Moser et al. 1997, p. 20).

Population 3 (Sekundarstufe II)

In dieser Population wurde der Leistungsstand in Mathematik und Naturwissenschaften am Ende der Sekundarstufe II erhoben. Auf dieser Stufe stellen sich bei der Definition der Population Probleme, die auf die grossen Unterschiede zwischen den Bildungssystemen aber auch innerhalb des Bildungssystems zurückzuführen sind. Struktur und Inhalte der Ausbildungsrichtungen in Berufsschulen und Gymnasien sind so verschieden, dass TIMSS die Leistungen in Mathematik und Naturwissenschaften mit *zwei grundsätzlich verschiedenen Ansätzen* erfasste:

- In einem *ersten Ansatz* wurden Mathematik und Naturwissenschaften auf dem Niveau einer *Grundbildung* erfasst, die für alle Mitglieder unserer durch Technik und Wissenschaft geprägten Gesellschaft als wesentlich angesehen wird («Scientific Literacy»).
- In einem *zweiten Ansatz* wurden in Mathematik und – exemplarisch – in Physik das Wissen und Können verglichen, welches in Ausbildungsrichtungen vermittelt wird, die auf ein Hochschulstudium bzw. auf weiterführende naturwissenschaftliche Ausbildungen vorbereiten. Wir bezeichnen diese Tests aus schweizerischer Sicht als «*gymnasial*».

Die Population für den Vergleich der Grundbildung wurde gebildet durch alle Schülerinnen und Schüler im letzten Jahr der Sekundarstufe II (vgl. Mullis, Martin, Beaton, Albert, Gonzales, Kelly et al. 1998). Diese Definition ist insofern problematisch, als dass ein Teil der Jugendlichen nie in das Abschlussjahr einer Ausbildung der Sekundarstufe II gelangt. Ebenfalls muss beachtet werden, dass der Anteil Jugendlicher, der die Sekundarstufe II abschliesst, von Land zu Land verschieden gross ist. Die Ergebnisse dürfen also nicht auf die ganze Alterskohorte übertragen werden (vgl. Ramseier et al. 1999, p. 43).

Innerhalb der Population für den Vergleich der Grundbildung wurde eine Teilpopulation für den voruniversitären, gymnasialen Leistungsvergleich gebildet. Diese Definition ist sehr offen formuliert. Es werden Schülerinnen und Schüler im letzten Schuljahr der Sekundarstufe II verglichen, die «physics» besucht haben, bzw. die «advanced mathematics courses» absolviert haben, zu deren Inhalten typischerweise auch Analysis und Trigonometrie gehören (vgl. Mullis et al. 1998). Es bleibt jedem Land überlassen, für sich zu entscheiden, welchen Unterricht es als «fortgeschrittene Mathematik» bzw. «Physik» bezeichnen will. Wenn die Schweiz beispielsweise nur den Maturitätstypus C in die Tests einbezogen hätte, wären die schweizerischen Ergebnisse deutlich besser ausgefallen.

Insgesamt beteiligten sich etwa 6400 Schülerinnen und Schüler an der Studie für die Sekundarstufe II (vgl. Ramseier et al. 1999).

3.1.3 Erkenntniswert

Ziel eines Schulleistungsvergleichs soll nicht nur sein, unterschiedliche Leistungen festzustellen, sondern auch diese Differenzen zu erklären. Erst die Erklärung erlaubt eine objektive Interpretation der Ergebnisse und kann eventuelle Handlungskonsequenzen begründen.

Bildung entsteht nie innerhalb eines geschlossenen Systems. Bedingungsfaktoren der Schulleistung finden sich auf der Ebene des gesamten Schulsystems, der Schulebene, der Unterrichtsebene sowie auf der individuellen Ebene. Das Geflecht der Bedingungsfaktoren kann selbstverständlich nie vollständig erfasst werden. Im Rahmen der TIMSS wurde eine Auswahl von Variablen getroffen, von denen man nach theoretischen Vorstellungen annahm, dass sie eine Auswirkung auf Schulleistung haben könnten.

Die Frage, welche Bedingungen und Merkmale eine optimale Qualifikation der Schülerinnen und Schüler fördert, stand im Zentrum der TIMSS⁺ («TIMSS plus»). Das Plus weist auf den schweizerischen Beitrag zur TIMSS hin, der in erster Linie zum Ziel hatte, diese erklärenden Faktoren, bezogen auf die Schweiz, zusätzlich zu analysieren.

Bei Querschnittuntersuchungen wie der TIMSS und der TIMSS⁺ können jedoch lediglich Aussagen über statistische Zusammenhänge («Korrelationen»), nicht aber über gesicherte Ursachen («Kausalitäten») gemacht werden. Werden Korrelationen festgestellt, können heuristische Aussagen über Ursachen gewagt werden. Vor Fehlschlüssen muss aber gewarnt werden. Auf der Sekundarstufe I wurde beispielsweise ein starker Zusammenhang zwischen sozialer Herkunft und der Mathematikleistung festgestellt. Dies bedeutet aber nicht zwingend, dass sich die soziale Herkunft direkt auf die Schulleistung auswirkt. Vielmehr konnte gezeigt werden, dass sich die soziale Herkunft massiv auf die Zugehörigkeit zu einem Schultyp auswirkte, die dann wiederum direkt die Mathematikleistung bestimmte. Soziale Herkunft übte also lediglich einen *indirekten Effekt* auf die Mathematikleistung aus. Ein nennenswerter direkter Effekt der sozialen Herkunft auf die Mathematikleistung konnte hingegen nicht nachgewiesen werden (vgl. Abbildung 3.1).

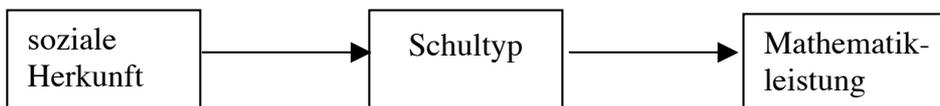


Abbildung 3.1: Indirekter Effekt der sozialen Herkunft auf die Mathematikleistung

3.2 Ergebnisse

Die Schweizer Resultate stehen im Zentrum dieses Ergebnisteils. Zusätzlich wird auf das mässige Abschneiden Deutschlands eingegangen. Der sogenannte «TIMSS-Schock» bewirkte in unserem Nachbarland Aktivitäten auf verschiedenen Ebenen des Schulsystems, welche im Schlussteil des dritten Kapitels vertiefend dargestellt werden.

3.2.1 Zusammenfassung der wichtigsten Ergebnisse der Schweiz

Die Hauptergebnisse der Schweiz sind strukturiert in zentrale und periphere Untersuchungsergebnisse. Zum ersten Bereich werden die Leistungen in Mathematik und Naturwissenschaften gezählt, im zweiten Bereich werden die Themen Curriculumanalyse, Attraktivität des Lehrberufs, Einstellungen zu Schule und Unterricht sowie Geschlechterdifferenzen resümiert.

3.2.1.1 Zentrale Untersuchungsergebnisse

Sekundarstufe I

Im Vergleich mit der Europäischen Union schneidet die Schweiz vor allem im Bereich der Mathematik sehr gut ab. Sie wird nur gerade vom flämischen Teil Belgiens übertroffen. Werden die Ergebnisse nur nach Alter verglichen, so sind die Ergebnisse aller beteiligten Nachbarländer schlechter einzustufen als diejenigen der Schweiz. Auch wenn für die Bewertung nur das achte Schuljahr einbezogen wird, liegt die Schweiz allein an der Spitze. Sehr deutliche Unterschiede zeigen sich zu Deutschland. Die Leistungsunterschiede zu Frankreich und Österreich sind zwar weniger deutlich, die Schülerinnen und Schüler dieser Länder sind aber älter. Die naturwissenschaftlichen Leistungen der Schweizer Schülerinnen und Schüler liegen am unteren Rand eines Mittelfelds, in dem die einzelnen Länder sehr nahe beieinander liegen. An den mittelmässigen Leistungen ändert sich auch nicht viel, wenn man nur die 8. Klasse einbezieht. Insgesamt erbringen Schweizer Schülerinnen und Schüler sehr gute Leistungen in Mathematik und durchschnittliche Leistungen in den Naturwissenschaften (vgl. Moser et al. 1997, p. 29f, 49ff.).

Im interkontinentalen Vergleich liegt die Schweiz deutlich hinter fast allen ostasiatischen Ländern zurück. Singapur als Spitzenreiter bleibt von sämtlichen Ländern in beiden Bereichen unangetastet. Im Vergleich mit den USA schneidet die Schweiz in der Mathematik besser und in den Naturwissenschaften schlechter ab (vgl. Moser et al. 1997, p. 39ff.).

Sekundarstufe II

Wie bereits erwähnt, wurden hier einerseits die Leistungen im Rahmen der Grundbildung und andererseits der gymnasialen Bildung erhoben. Wie auf der Sekundarstufe I liegt die Schweiz bezogen auf die Grundbildung wie auch auf die gymnasiale Bildung in der Mathematik im Vergleich innerhalb Europas auf einem Spitzenrang. In den Naturwissenschaften liegt die Schweiz im Rahmen der Grundbildung im Mittelfeld, bei der gymnasialen Physik nur knapp im Mittelfeld (vgl. Ramseier et al. 1999, p. 50f., 137ff.).

Im internationalen Vergleich schneidet die Schweiz sowohl bei der Grundbildung Mathematik wie auch bei der gymnasialen Mathematik gut ab. Bei der Grundbildung Naturwissenschaften liegt die Schweiz im guten Mittelfeld und ihre relative Position ist im Vergleich zur Sekundarstufe I sogar etwas höher. Im internationalen Vergleich liegt die gymnasiale Physik statistisch signifikant unter dem Mittelwert (vgl. Ramseier et al. 1999, p. 48ff.).

3.2.1.2 Periphere Untersuchungsergebnisse

Neben den beiden Leistungserhebungen in Mathematik und Naturwissenschaften wurden diverse weitere Untersuchungen in verschiedenen Bereichen durchgeführt.

Sekundarstufe I

International wurde eine Curriculumsanalyse durchgeführt. Dabei zeigte sich, dass der Unterricht in der Schweiz die Inhalte des Mathematiktests besser abdeckte als diejenigen des naturwissenschaftlichen Tests. Eine Aufgabeneinschätzung durch Experten zeigte jedoch, dass der internationale Vergleich erstaunlich wenig dadurch beeinflusst wird, ob das Ergebnis auf der Auswertung aller Aufgaben beruht, oder nur auf jenen, welche die Schweiz als ihren Unterrichtsinhalten entsprechend einstuft (vgl. Moser et al. 1997, p. 52ff.).

Lehrpersonen gelten als Schlüsselfaktor für ein qualitativ hochstehendes Bildungssystem. Es ist deshalb wichtig, die Attraktivität des Lehrberufs zu gewährleisten. International gesehen sind die meisten Lehrkräfte der Mathematik unzufrieden mit ihrem Ansehen in der Öffentlichkeit. Eine Ausnahme bildet hier die Schweiz: rund 80 Prozent der befragten Lehrkräfte gaben an, dass sie

glauben, die Gesellschaft schätze ihre Arbeit. Hohe Wertschätzung der Lehrperson durch die Gesellschaft ist jedoch nicht gleichzusetzen mit hoher Berufszufriedenheit. Rund ein Viertel der schweizerischen Mathematiklehrkräfte würde gerne den Beruf wechseln. Damit liegt die Schweiz im internationalen Mittelfeld (vgl. Moser et al. 1997, p. 99ff.).

Positive Einstellungen zu Unterricht und Schule sind aus lernpsychologischer Sicht wichtige Kriterien. Für den internationalen Vergleich wurde der gefühlsbezogene Aspekt des Interesses am schulischen Fach Mathematik mit Hilfe von drei gezielten Fragen erfasst. Das Interesse der Schweizer Schülerinnen und Schüler an der Mathematik liegt im internationalen Mittelfeld. Im Vergleich zu den leistungsmässig besten Ländern Ostasiens ist das Interesse in der Schweiz höher. Auch die Resultate zur Erhebung des «Selbstvertrauens in Mathematik» zeigen, dass Schülerinnen und Schüler aus Ländern mit hohen Leistungen ein eher niedriges Selbstvertrauen haben. Umgekehrt stimmen in der USA mehr als 80% der Frage «Normalerweise bin ich gut in Mathematik» zu, obwohl sie in der Erhebung relativ schlechte Leistungen erbrachten. Das Selbstvertrauen der Schweizer Schülerinnen und Schüler liegt im Mittelfeld (vgl. Moser et al. 1997, p. 84ff.).

Sekundarstufe II

Die Leistungen in Mathematik und Naturwissenschaften in der Schweiz unterscheiden sich sowohl auf gymnasialer Stufe wie auch in Bezug auf die Grundbildung zwischen den Geschlechtern signifikant. In der Grundbildung liegt der Geschlechterunterschied im internationalen Mittel, bei der gymnasialen Ausbildung ist der Unterschied in zwei Dritteln der teilnehmenden Länder tiefer.

In der gymnasialen Mathematik erreichten nur etwa 21 Prozent der Schülerinnen die Durchschnittsleistung der Schüler. Ein Teil dieses Unterschieds ist darauf zurückzuführen, dass mehr junge Männer als Frauen den Maturitätstypus C besuchen. Hinter dieser Typenwahl steht aber bereits der Unterschied in Leistung, Interesse und Selbstvertrauen zwischen den Geschlechtern. Doch selbst innerhalb der einzelnen Maturitätstypen sind die Geschlechterunterschiede bezüglich der Leistung beträchtlich. In der gymnasialen Physik sind die Unterschiede zwischen Schülerinnen und Schülern sogar noch grösser als in der gymnasialen Mathematik. Nur in zwei der an TIMSS beteiligten Länder sind die Unterschiede gleich gross oder grösser (vgl. Ramseier et al. 1999, p. 117ff.)

3.2.2 Zusammenfassung der wichtigsten Ergebnisse Deutschlands

Sekundarstufe I

Die Leistungen von deutschen Schülerinnen und Schülern auf der Sekundarstufe I liegen international sowie innerhalb der Europäischen Union im Mittelfeld. Die Leistungen in den naturwissenschaftlichen Fächern fallen etwas günstiger aus als in Mathematik. In den Naturwissenschaften erbringt Deutschland etwa dieselben Resultate wie die Schweiz, während im Bereich der Mathematik die Schweiz deutlich besser abschneidet. Die deutschen Resultate werden relativiert durch die Tatsache, dass die getesteten Schülerinnen und Schüler aus Deutschland im Durchschnitt um sechs bis zwölf Monate älter sind, als die Teilnehmenden anderer Länder. Die Mehrzahl der Schülerinnen und Schüler der Vergleichsgruppe der Europäischen Union verfügt also über einen deutlichen Leistungsvorsprung. Der Anteil jener Schülerinnen und Schüler des 8. Jahrgangs, deren mathematische und naturwissenschaftliche Fähigkeiten lediglich auf einem erweiterten Grundschulniveau liegen, ist mit etwa 20 Prozent auch im internationalen Vergleich hoch (vgl. Max-Planck-Institut für Bildungsforschung 2003).

Sekundarstufe II

In Bezug auf die mathematisch-naturwissenschaftliche Grundbildung liegt Deutschland im unteren Bereich. Die Abstände zu den leistungsstärkeren Ländern vergrössern sich gar vom Ende der 8. Klasse bis zum Ende der Sekundarstufe II. Der Anteil der Personen, die das Fähigkeitsniveau der

Beherrschung einfacher Routinen nicht überschreiten, liegt bei 70 Prozent. Selbst bei den Gymnasiasten sind bei 40 Prozent Unsicherheiten im Bereich elementarer Rechenfähigkeiten festzustellen. Auch auf gymnasialer Stufe können sich deutsche Schülerinnen und Schüler im Mathematikunterricht nicht mit ihren europäischen Nachbarn messen. Als einzige dieser Gruppe erreichen sie den internationalen Mittelwert nicht. Etwas besser schneidet Deutschland in der gymnasialen Physik ab. In allen gemessenen Leistungsbereichen liegen die Leistungen Deutschlands auf der Sekundarstufe II signifikant unter den Leistungen der Schweiz (vgl. Max-Planck-Institut für Bildungsforschung 2003).

3.2.3 Interpretation der Ergebnisse

Schulleistungen werden von verschiedensten Variablen beeinflusst. Wenn wir das Zustandekommen von Leistungen erklären und Ergebnisse von Schulleistungsmessungen interpretieren wollen, müssen wir diese Einflussfaktoren in unsere Betrachtungen mit einbeziehen. Alle Variablen, die einen Einfluss auf Schulleistung ausüben, erfassen zu wollen, ist selbstverständlich unmöglich. Im Folgenden werden ausgewählte Variablen vorgestellt, deren Betrachtung eine differenziertere Interpretation der Ergebnisse erlauben soll.

3.2.3.1 Geschlechterunterschiede in der Mathematikleistung auf der Sekundarstufe I

Die Schweiz gehört gemäss TIMSS-Ergebnissen zu den wenigen Ländern mit signifikanten Geschlechterunterschieden in der Mathematikleistung. Das Interesse an Mathematik ist bei Schweizer Mädchen geringer als bei den Knaben, und sie verfügen über ein auffällig geringeres Selbstvertrauen in die eigene mathematische Leistungsfähigkeit (vgl. Beaton, Mullis, Martin, Gonzales, Kelly et al. 1996, p. 99ff.). Die Geschlechterunterschiede in der Mathematikleistung sind nicht auf allen Klassenstufen signifikant. So konnten in der 6. Klasse signifikante Geschlechterunterschiede in der Mathematikleistung festgestellt werden, in der 7. Klasse hingegen keine signifikanten Unterschiede (vgl. Beaton, Mullis, Martin, Gonzales, Kelly et al. 1996, p. 33ff.). Wenn wir aber bedenken, dass sie im Sinne einer «Gleichstellung der Geschlechter in der Schule» gar nicht vorhanden sein sollten, gewinnen sie trotzdem an praktischer Bedeutung. Die Geschlechterdifferenzen im Interesse an der Mathematik und im Selbstvertrauen sind sogar schwerwiegend, gehören diese beiden Komponenten doch zu den wichtigsten Voraussetzungen für erfolgreiches Lernen.

Eine Untersuchung der TIMSS erfasste die Geschlechterstereotypisierung der Fächer Mathematik, Biologie, Physik und Sprachen. Bereits im 6. Schuljahr zeigte sich, dass Mathematik und noch stärker Physik als männliche, Sprache und ganz schwach auch Biologie als weibliche Domäne betrachtet werden. Das Ausmass der Geschlechterstereotypisierung nimmt in den folgenden Schuljahren signifikant zu und ist im 8. Schuljahr besonders gross. Auch von den Lehrkräften werden die Schulfächer als männliche oder weibliche Domänen betrachtet. Die Ergebnisse der Befragung der Lehrpersonen zeigten, dass diese die Schulfächer noch stärker stereotypisieren als die Schülerinnen und Schüler. Diese Tatsache erstaunt, da die faktischen Leistungsunterschiede zwischen den Geschlechtern auf der Sekundarstufe I sehr klein sind. Um eine Identifikation der Mädchen mit der Mathematik und der Physik zu ermöglichen und somit gleiche Ausgangsbedingungen für Mädchen und Knaben zu schaffen, ist der Abbau von Stereotypisierung seitens der Schülerinnen und Schüler wie auch der Lehrkraft Voraussetzung (vgl. Keller 1997, p. 138ff.).

3.2.3.2 Klassengrösse

Die Klassengrösse ist von Land zu Land sehr unterschiedlich (vgl. Beaton et al. 1996, p. 152). Während in der Schweiz die Zahl von dreissig Schülerinnen und Schülern pro Klasse als sehr gross bewertet wird, würde sie beispielsweise in Korea wohl eher als klein eingestuft. Auf den ersten

Blick mag erstaunen, dass in den grossen Klassen Ostasiens die besten Leistungen erbracht wurden (vgl. Abschnitt 3.2.1). Die Wirkung der Klassengrösse kann allerdings nicht ungeachtet der Ziele und der kulturellen Voraussetzungen des Unterrichts beurteilt werden. Es zeigte sich zwar, dass sich die Grösse einer Gruppe nicht direkt auf den Lernerfolg auswirkt, diese aber auf Interesse der Lernenden, auf soziale Interaktionen, individuelle Betreuung und auf die Belastung der Lehrkraft einen grossen Einfluss hat. Besonders für schwache Schülerinnen und Schüler wirkt sich eine grosse Klasse negativ aus. Somit unterscheiden sich die Lernbedingungen in kleinen Klassen von denjenigen grosser Klassen und schaffen eine Basis für bessere Leistungen. Tatsächlich zeigte sich, dass TIMSS Klassen mit mehr als 24 Schülerinnen und Schülern sowohl in Mathematik wie auch in den Naturwissenschaften durchschnittlich schlechter abschnitten als kleine Klassen (vgl. Moser 1997, p. 203ff.).

3.2.3.3 Unterrichtszeit in der Mathematik und in den Naturwissenschaften auf der Sekundarstufe I

Die den einzelnen Unterrichtsfächern zugeteilte Unterrichtszeit ist ein grundlegendes Merkmal der Lehrpläne und wichtiger Einflussfaktor dafür, wie schulischer Lernstoff vermittelt werden kann. Im Rahmen der Lehrerbefragung wurde die wöchentliche Unterrichtszeit in Mathematik und Naturwissenschaften erhoben. Die wöchentliche Unterrichtszeit in Mathematik beträgt in der Schweiz durchschnittlich 4.2 Stunden (nicht Lektionen). Damit ist der Zeitaufwand bei uns nach Belgien, der Slowakei und Russland eher hoch. Der Durchschnitt aller Länder liegt bei 3.5 Stunden. Im Gegensatz dazu gehört die Schweiz bei den Naturwissenschaften zu den Ländern mit dem kürzesten Unterricht. Mit 2.6 Stunden liegt die Schweiz deutlich unter dem wöchentlichen Durchschnitt von 2.9 Stunden. Erstaunlich ist, dass Unterrichtsdauer und Leistung im Vergleich der Ländermittelwerte nicht miteinander korrelieren. Offenbar spielen viele andere Faktoren eine ebenso wichtige Rolle. Wenn wir jedoch die Differenz der Unterrichtsdauer von Mathematik und Naturwissenschaften mit der Differenz der jeweiligen Fachleistungen vergleichen, können wir einen Zusammenhang feststellen. Die Schweiz fällt als eines jener Länder auf, deren Naturwissenschaftsunterricht, verglichen mit dem Mathematikunterricht, relativ kurz ist. Entsprechend ergibt sich die Diskrepanz zwischen den Leistungen in Mathematik und denjenigen in den Naturwissenschaften (vgl. Moser 1997, p. 66ff.).

3.2.3.4 Öffentliche Bildungsausgaben

Die Höhe der öffentlichen Bildungsausgaben bestimmt nicht zwingend, wie gut die Schülerinnen und Schüler in einem Land in Schulleistungstests abschneiden, sie ist jedoch ein Mass dafür, wie schwierig es für ein Land ist, die Bildungsaufgabe wahrzunehmen. Wieviel ein Land für Bildung ausgibt, ist einerseits abhängig von seinem Wohlstand, andererseits von politischen Entscheidungen, welche die Bildung betreffen. Die Schweiz ist nach den USA das Land mit dem höchsten Bruttosozialprodukt innerhalb der OECD (vgl. Beaton et al. 1996, p. 13ff). Dementsprechend sind auch die Bildungsausgaben hoch. Innerhalb der OECD haben nur die USA, Norwegen und Dänemark höhere Bildungsausgaben. Im internationalen Vergleich kann aber von der Höhe der Bildungsausgaben nicht direkt auf die Qualität der Leistungen geschlossen werden. Die ostasiatischen Länder und Tschechien beispielsweise erreichen mit weniger Bildungsausgaben deutlich bessere Leistungen als die Schweiz (vgl. Moser 1997, p. 96ff.).

→ Fakultative Lektüre: Rezeption und Konsequenzen der TIMSS³⁹

³⁹ Die Texte zur fakultativen Lektüre finden Sie online in der Lernumgebung des Moduls.

4 Internationale Schulleistungsvergleiche am Beispiel von PISA (Programme for International Student Assessment)

4.1 Generelles

Die PISA-Studie (Programme for International Student Assessment) stellt regelmässig schulpolitisch relevante und international vergleichbare Indikatoren⁴⁰ für Leistungen von Schülerinnen und Schülern bereit. Ausgehend von der ersten Erhebung PISA 2000 werden im Dreijahreszyklus Daten von 15-jährigen Jugendlichen erhoben. Wie in Abbildung 4.1 ersichtlich, lag der Schwerpunkt von *PISA 2000* auf der *Lesekompetenz* und in *PISA 2003* auf der *mathematischen Grundbildung*. In *PISA 2006* hingegen stand die *naturwissenschaftliche Grundbildung* im Zentrum. In den für 2009, 2012 und 2015 geplanten Erhebungen wird sich diese Abfolge wiederholen und eine längerfristige Beobachtung der Entwicklung in den teilnehmenden Ländern ermöglichen.

2000	Lesen	Mathematik	Naturwissenschaften	
2003	Lesen	Mathematik	Naturwissenschaften	Problemlösen
2006	Lesen	Mathematik	Naturwissenschaften	

 = Schwerpunkt

Abbildung 4.1: Das Projekt PISA – eine zyklische Studie (aus: BFS & EDK 2004, p. 7)

PISA ist ein Kooperationsprojekt der Mitgliedstaaten der OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung), wobei auch Staaten teilgenommen haben, die nicht der OECD angehören (vgl. Abbildung 4.2). Die grosse geografische Reichweite macht eine Besonderheit der PISA-Studie aus und erlaubt den Vergleich unter einer Vielzahl von Ländern. In der Schweiz tragen Bund und Kantone, vertreten in der nationalen Steuergruppe, das Projekt PISA. Das Gremium wird von Expertinnen und Experten aus Verwaltung, Wissenschaft und Bildungspraxis beraten. Das Bundesamt für Statistik in Neuenburg ist mit der nationalen Projektleitung betraut (vgl. BFS & EDK 2004, p. 5).

Ziel von PISA ist es, anhand empirisch gesicherter Daten Antworten auf politisch, gesellschaftlich und ökonomisch motivierte Fragen zu erhalten. Es sollen Informationen geliefert werden über die Leistungen von Schülerinnen und Schülern im internationalen Vergleich, über die Zusammenhänge zwischen diesen Leistungen und bestimmten Merkmalen von Jugendlichen und Schulen sowie über die zeitliche Entwicklung dieser Leistungen. Nicht die Einzelleistungen von Schülerinnen und Schülern stehen dabei im Zentrum, sondern Stärken und Schwächen der Bildungssysteme. PISA soll Grundlagen für eine bessere Einschätzung und Kontrolle der Effektivität von Bildungssystemen schaffen.

⁴⁰ Indikatoren sind Merkmale, die als Anzeichen auf etwas anderes hinweisen.

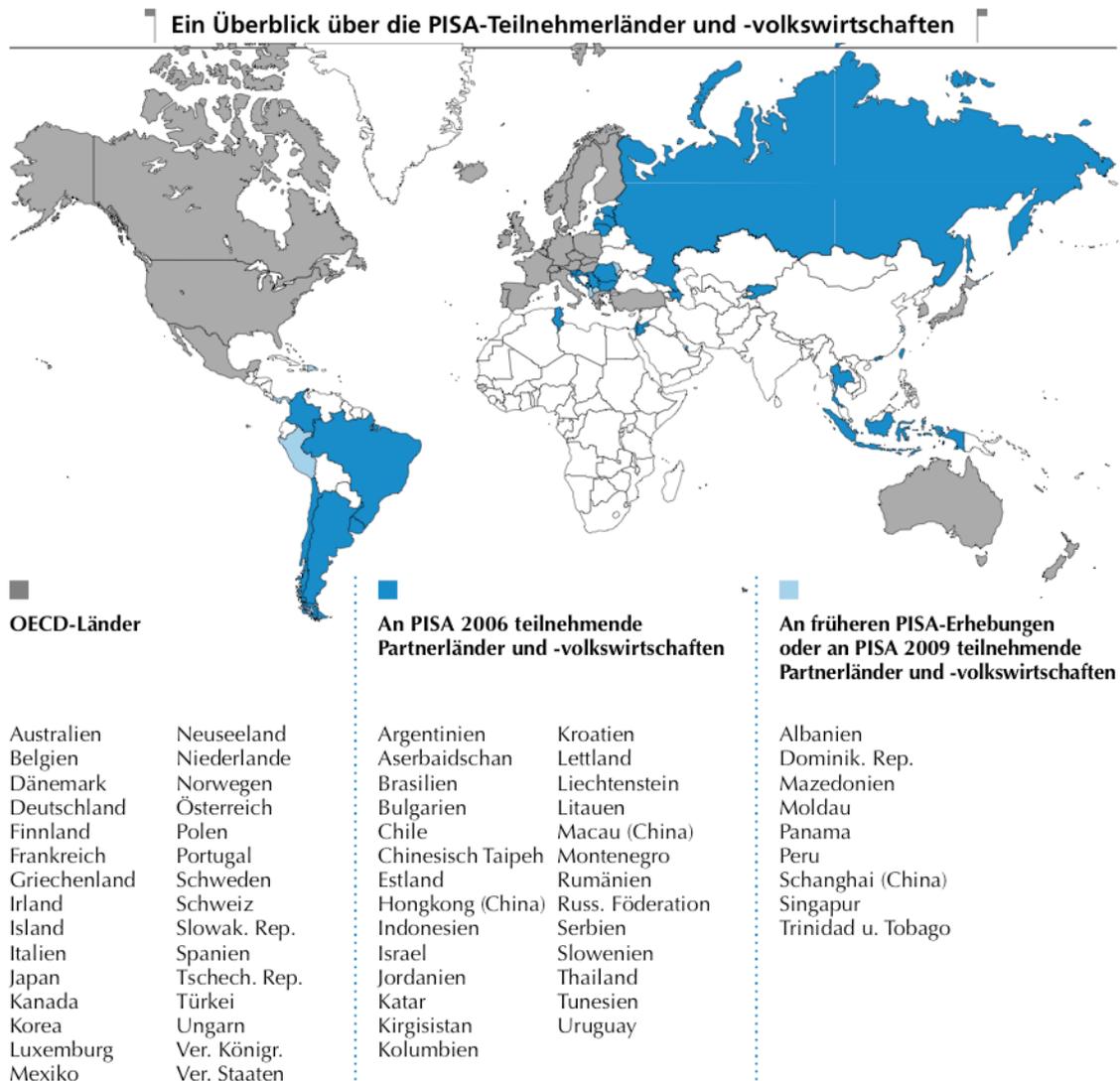


Abbildung 4.2: Überblick über die an PISA teilnehmenden Länder (aus: OECD 2007, p. 20)

Im Unterschied zur Leistungsmessung in der 1991 durchgeführten «Reading Literacy Study» und in der «Third International Mathematics and Science Study»⁴¹ orientiert sich PISA weniger am Schulwissen, wie es in den Lehrplänen aufgeführt ist, sondern konzentriert sich auf die Erfassung von *Basiskompetenzen*.

«Man kann gar nicht nachdrücklich genug betonen, dass PISA keineswegs beabsichtigt, den Horizont moderner Allgemeinbildung zu vermessen, oder auch nur die Umrisse eines internationalen Kerncurriculums nachzuzeichnen. Es ist gerade die Stärke von PISA, sich solchen Allmachtsphantasien zu verweigern und sich stattdessen mit der Lesekompetenz und mathematischen Modellierungsfähigkeit auf Basiskompetenzen zu konzentrieren, die nicht die einzigen, aber wichtige Voraussetzungen für die [...] Generalisierung universeller Prämissen für die Teilhabe an Kommunikation und damit auch für Lernfähigkeit darstellen» (Baumert, Stanat & Demmrich 2001, p. 21).

Der Begriff der *Kompetenz* zielt demzufolge auf die Beherrschung wichtiger Kenntnisse und Fähigkeiten ab, die im Berufs- und Erwachsenenleben benötigt werden, Fähigkeiten und Fertigkeiten

⁴¹ Vgl. dazu die Kapitel 2 und 3.

ten also, die als nützlich erachtet werden, um an der heutigen und künftigen Gesellschaft teilnehmen zu können (vgl. Baumert, Stanat & Demmrich 2001, p. 29). PISA orientiert sich damit an einem Konzept der «Grundbildung» (literacy) und weniger an einer Wiedergabe von angeeignetem Wissen. Im Zentrum steht, wie das erworbene Wissen angewendet werden kann, um in verschiedenen Alltagssituationen kommunizieren sowie Probleme analysieren und lösen zu können (vgl. BFS & EDK 2007, p. 13).

Den Entscheid, was als Basiskompetenz definiert und in variierenden Anwendungssituationen geprüft wird, fällten die Verantwortlichen von PISA pragmatisch, also handlungs- und sachbezogen (vgl. Baumert, Stanat & Demmrich, 2001, p. 21). Es wird dabei eine Auswahl getroffen und gewertet, also normativ vorgegangen. Die von PISA gewählten Kompetenzen sind damit kritisierbar.

Die Leistungen in den Bereichen *Lesekompetenz*, *mathematische Grundbildung*, *naturwissenschaftliche Grundbildung* und ab 2003 neu der *Problemlösefähigkeiten* werden anhand von Leistungstests erhoben. Zusätzlich werden weitere Informationen und Selbsteinschätzungen der schriftlich befragten Jugendlichen ausgewertet. Es handelt sich dabei unter anderem um Fragen zur sozialen Herkunft und zur Chancengleichheit, zu Multikulturalität, zu Bildung für Mädchen und Knaben, zum selbstregulierten Lernen und zur Vertrautheit im Umgang mit dem Computer (vgl. OECD 2001, p. 20; Zahner & Meyer 2002, p. 12, 14).

Die *Stichprobe* von PISA umfasst 15-jährige Schülerinnen und Schüler. In der Schweiz wurden zusätzlich zu den 15-Jährigen auch Neuntklässlerinnen und Neuntklässler untersucht. Die Stichprobe der 15-Jährigen dient v.a. dem internationalen Vergleich. Die nationalen Zusatzstichproben mit den Schülerinnen und Schülern der neunten Klassenstufe erlauben zusätzliche Vergleiche zwischen den Sprachregionen sowie zwischen einigen Kantonen (vgl. Zahner & Meyer 2002, p. 15). Um auch regionale und kantonale Vergleiche zu ermöglichen, wurden in der Schweiz mehr Jugendliche getestet als für PISA vorgegeben. Die erweiterte Stichprobe dient der vertieften Analyse der Daten innerhalb der Schweiz und dem besseren Verständnis der internationalen Resultate. Dies ist umso wichtiger, als die Schweiz über ein sehr heterogenes Bildungssystem verfügt (vgl. ebd., p. 16f.).

In der Schweiz wurden für den internationalen Vergleich repräsentative Stichproben für die drei grossen Sprachregionen auf der Grundlage eines zweistufigen Verfahrens ausgewählt. Nach einer zufälligen Ziehung der Schulen wurden 15-jährige Jugendliche innerhalb dieser Schulen zufällig ausgewählt. Für die nationalen Zusatzstichproben wurden einerseits die 15-Jährigen der internationalen Stichprobe, die sich in der 9. Klasse befanden, berücksichtigt. Andererseits wurden an den Schulen zusätzlich zufällig ausgesuchte 9. Klassen untersucht (vgl. ebd., p. 16).

Die *Ergebnisse* internationaler Vergleichsstudien, auch jene von PISA, werden oft in Ranglisten dargestellt (vgl. Tabelle 4.1). Der Vergleich des Abschneidens der einzelnen Länder findet vor allem in den Medien ein grosses Echo. Ranglisten, die zwischen besseren und schlechteren Leistungen unterscheiden, sind wegen ihrer leichten Lesbarkeit sehr beliebt, können jedoch zu Fehlinterpretationen verleiten. So verdecken sie beispielsweise die Unterschiede *innerhalb* der Länder, die meist grösser sind als die Unterschiede *zwischen* den Ländern.

Tabelle 4.1: Leistungen in den drei Kompetenzen, PISA 2000 (aus: Meyer & Zahner 2002, p. 50)

Lesen	Mathematik	Naturwissenschaften
Finland	<i>Japan</i>	Südkorea
Kanada	Südkorea	<i>Japan</i>
Neuseeland	Neuseeland	Finland
Australien	Finland	Grossbritannien
Irland	Australien	Kanada
Südkorea	Kanada	Neuseeland
Grossbritannien	Schweiz	Australien
<i>Japan</i>	Grossbritannien	<i>Österreich</i>
<i>Schweden</i>	Belgien	Irland
<i>Österreich</i>	<i>Frankreich</i>	<i>Schweden</i>
Belgien	<i>Österreich</i>	Tschechien
Island	Dänemark	<i>Frankreich</i>
Norwegen	Island	Norwegen
<i>Frankreich</i>	Liechtenstein	USA
USA	<i>Schweden</i>	Ungarn
Dänemark	Irland	Island
Schweiz	Norwegen	Schweiz
Spanien	Tschechien	Belgien
Tschechien	USA	Spanien
<i>Italien</i>	<i>Deutschland</i>	<i>Deutschland</i>
<i>Deutschland</i>	Ungarn	Polen
Liechtenstein	Russland	Dänemark
Ungarn	Spanien	<i>Italien</i>
Polen	Polen	Liechtenstein
Griechenland	Lettland	Griechenland
Portugal	<i>Italien</i>	Russland
Russland	Portugal	Lettland
Lettland	Griechenland	Portugal
Luxemburg	Luxemburg	Luxemburg
Mexiko	Mexiko	Mexiko
Brasilien	Brasilien	Brasilien

<p>☐ Jugendliche aus diesen Ländern erreichen statistisch signifikant bessere Leistungen als die Jugendlichen in der Schweiz</p>	<p>☐ Jugendliche aus diesen Ländern erreichen Leistungen, die sich von jenen der Jugendlichen in der Schweiz statistisch nicht signifikant unterscheiden</p>	<p>☐ Jugendliche aus diesen Ländern erreichen signifikant schlechtere Leistungen als die Jugendlichen in der Schweiz</p>
--	--	--

Bezugsländer sind kursiv gedruckt

© BFS/EDK Quelle: OECD - BFS/EDK PISA Datenbank, 2001

Hinzu kommt, dass Ranglisten nichts über die Grösse der Abstände aussagen. Jedes Land belegt einen Rang, ungeachtet des Ausmasses des Unterschiedes. In der Tabelle 4.1 ist deshalb zu beachten, dass sich das Abschneiden der Schweiz nur von jenen Ländern signifikant unterscheidet, die *nicht* im mittelgrau markierten Bereich liegen. So sind streng genommen in der Mathematik nur die Leistungen von Japan und Südkorea in PISA 2000 signifikant besser als jene der Schweiz. Der Unterschied zwischen Neuseeland und der Schweiz besteht zwar anhand der ausgewerteten Daten, allerdings ist die Wahrscheinlichkeit, dass es sich hier um einen zufälligen Befund handeln könnte, statistisch gesehen zu hoch.

Ein Beispiel für die Notwendigkeit, Ranglisten näher zu differenzieren, sind die Unterschiede zwischen den Kantonen Bern, St. Gallen und Zürich. In PISA 2000 hat sich gezeigt, dass die Lesekompetenzen der untersuchten Jugendlichen im Kanton St. Gallen die besten sind (vgl. Abbildung 4.3).

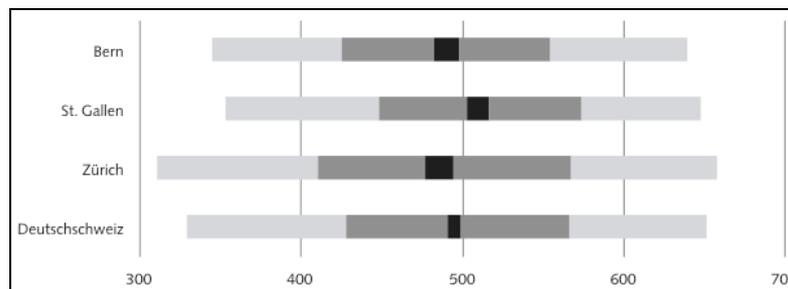


Abbildung 4.3: Lesekompetenzen bei Jugendlichen der 9. Klasse im Vergleich, PISA 2000
(aus: Moser, Ramseier & Berweger 2002, p. 17)

Statistisch gesehen sind aber nur die Differenzen zwischen St. Gallen und Zürich mit 24 Punkten Unterschied und zwischen St. Gallen und Bern mit 19 Punkten Unterschied signifikant. Diese Differenzen müssen als klein beurteilt werden (vgl. Moser, Ramseier & Berweger 2002, p. 17). Dennoch wurden sie in den Medien aufgrund der Rangierung⁴² aufgebauscht.

Für die Schweiz ist es nicht nur von Interesse, sich mit dem Durchschnitt aller an der Studie beteiligten Länder zu vergleichen, sondern insbesondere auch mit den Nachbarstaaten Deutschland, Österreich, Frankreich und Italien sowie mit führenden Industrienationen, zu denen schon in anderen Studien Bezüge hergestellt worden sind und über deren Bildungssystem gute Kenntnisse vorliegen. Dazu gehören Japan, Schweden und die USA. Diese sieben Länder zählen für PISA 2000 zu den Bezugsländern der Schweiz (vgl. Zahner & Meyer 2002, p. 20).

4.2 Ergebnisse

Im folgenden Abschnitt sind die wichtigsten Ergebnisse der PISA-Erhebungen 2000, 2003 und 2006 zusammengefasst. Zunächst werden die Leistungskompetenzen im Lesen, in der Mathematik und in den Naturwissenschaften dargestellt. Es sollen dabei schwerpunktmässig jene Ergebnisse hervorgehoben werden, die zum Erhebungszeitpunkt den Hauptfokus der Studie bildeten, also Lesen im Jahr 2000, Mathematik in der Erhebung 2003 und Naturwissenschaften in PISA 2006. Danach wird auf die wichtigsten Zusammenhänge zwischen den Schülerleistungen und dem sozioökonomischen und dem kulturellen Hintergrund sowie dem Geschlecht eingegangen.

⁴² Vgl. Kapitel 1.

4.2.1 Leistungskompetenzen

Im Zentrum von PISA stehen die Leistungen der Schülerinnen und Schüler im Lesen, in der Mathematik und den Naturwissenschaften. In der ersten Erhebungswelle PISA 2000 lag der Schwerpunkt auf der Lesekompetenz.

4.2.1.1 Lesekompetenz

Die Lesekompetenz wurde mittels *verschiedener Textarten* gemessen.

Dabei wurden folgende Aspekte unterschieden:

- Die *Form des Lesestoffs*, also ob es sich um fortlaufende Texte (Erzählungen, Sachbeschreibungen, Kommentare, Argumentationen) oder nicht fortlaufende Texte (Diagramme, Bilder, Karten, Tabellen, Graphiken) handelt. Der internationale Lesetest umfasst 141 Aufgaben, wobei sich etwa zwei Drittel auf fortlaufende Texte beziehen (vgl. Artelt et al. 2001, p. 80f.).
- Die Verwendung des Textes, also die *Textsorte*. Hier wurde zwischen Lesesituationen unterschieden, die eher einem privaten Zweck (z.B. Auszüge aus Erzählungen), einem öffentlichen Zweck (z.B. amtliche Dokumentationen), der beruflichen Weiterbildung (z.B. Lehrbücher) oder dem allgemeinen Bildungsinteresse (z.B. Sachbücher) dienen (vgl. ebd., p. 81).
- Die *Art der Leseaufgabe*. Es wurden dabei 3 Bereiche (Subskalen) unterschieden (vgl. Meyer & Zahner 2002, p. 24; Artelt et al. 2001, p. 83):
 1. das Heraussuchen und Ermitteln von Informationen
 2. das textbezogene Interpretieren
 3. das Reflektieren und Beurteilen

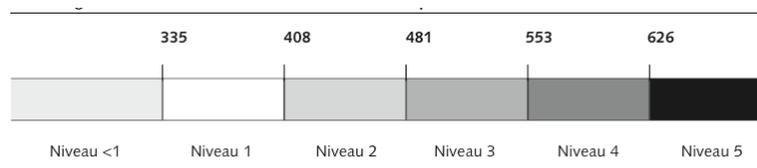


Abbildung 4.4: Skalenbereich und Grenzwerte für die Kompetenzniveaus
(aus: Meyer & Zahner 2002, p. 24)

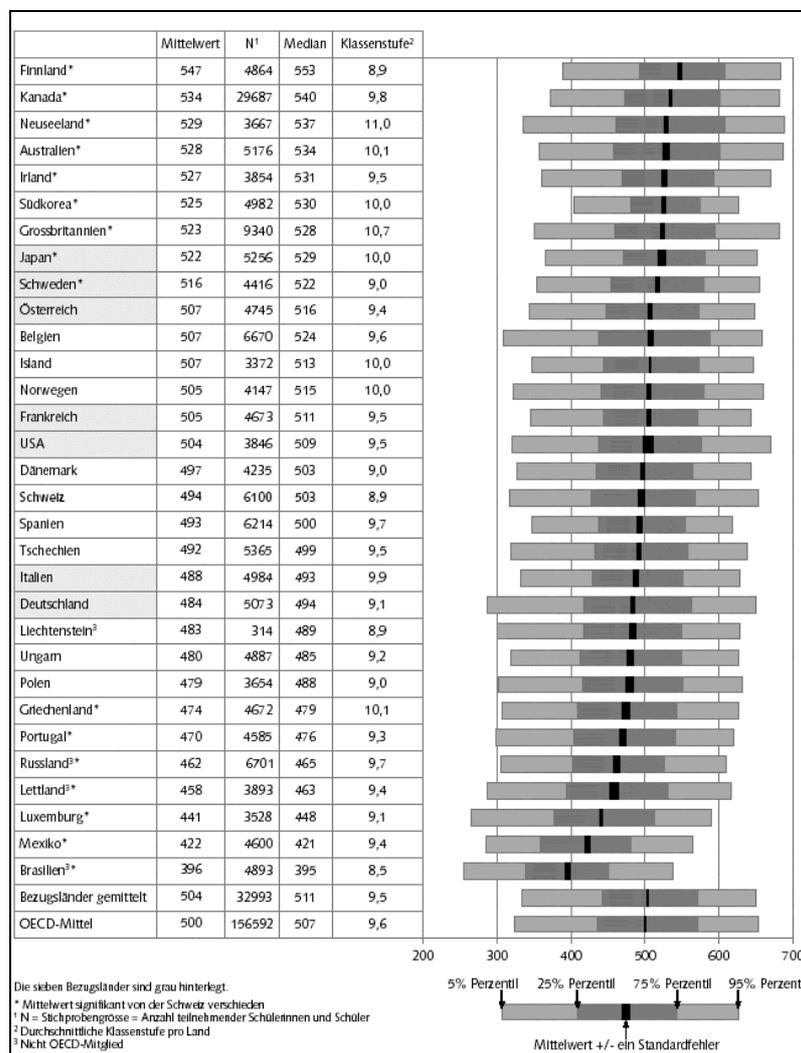
Für die Erfassung der Lesekompetenz wurden fünf Kompetenzniveaus definiert. Dabei entspricht das Niveau 5 einem sehr guten Leistungsbereich, Niveau 1 dagegen einem sehr schlechten. Die Niveaus sind nicht als kontinuierliche Skala von 1 bis 5 zu verstehen. Jedem Kompetenzniveau wurden verschiedene Aufgaben mit entsprechenden Schwierigkeitsgraden zugeordnet. Damit enthalten sie inhaltliche Kriterien, die darüber Auskunft geben, wie gut eine bestimmte Aufgabe gelöst wurde. Jugendliche innerhalb eines Niveaus sind in der Lage, mindestens die Hälfte der Aufgaben des Niveaus richtig zu lösen. Jugendliche, welche die einfacheren Aufgaben innerhalb eines Niveaus mit einer Wahrscheinlichkeit von 62% richtig lösen, die schwierigen Aufgaben hingegen nur zu 42%, werden im unteren Bereich des Niveaus angesiedelt. Wer dagegen 70% aller Aufgaben richtig löst bzw. 62% aller schwierigen und 78% der einfachsten Aufgaben, der gehört in den oberen Bereich (vgl. Meyer & Zahner 2002, p. 24ff.).

Die Skala zur Erfassung der Lesekompetenz wurde für PISA 2000 so normiert, dass sie einen Mittelwert der OECD-Länder von 500 Punkten und eine Standardabweichung von 100 Punkten

aufweist. Dies bedeutet, dass etwa zwei Drittel aller Jugendlichen Leistungen im Bereich 400 bis 600 Punkte erzielen. Die Daten der folgenden Erhebungen werden dann in dieser Skala «verankert», so dass Aussagen über Veränderungen im Kompetenzniveau mit diesem Vergleichsmaßstab möglich werden (vgl. BFS & EDK 2007, p. 41).

Aus der Abbildung 4.4 ist zu entnehmen, dass der durchschnittliche Abstand zwischen zwei Niveaugrenzen knapp über 70 Punkte beträgt. Bildungspolitisch relevant sind vor allem die beiden untersten (bis 480 Punkte) und das oberste Niveau (über 625 Punkte). Wer weniger als 335 Punkte erreicht hat, «ist kaum im Stande, auch nur die einfachsten der in PISA gestellten Aufgaben zu lösen, und hat somit ernsthafte Schwierigkeiten im Umgang mit und im Verständnis von Texten welcher Art auch immer» (Meyer & Zahner 2002, p. 24).

Tabelle 4.2: Leseleistungen im Ländervergleich, PISA 2000
(aus: Meyer & Zahner 2002, p. 29)



Die *Leistungsunterschiede* zwischen den einzelnen Ländern sind erheblich. So haben in der Erhebung 2000 die besten OECD-Länder einen Abstand von mindestens eineinhalb Kompetenzniveaus zu den schlechtesten (vgl. ebd., p. 29). Insgesamt schneidet die Schweiz im Vergleich zum OECD-Mittel in den Jahren 2000 und 2003 in der Lesekompetenz nur mittelmässig ab, was Anlass zu bildungspolitischen Diskussionen gab. In PISA 2006 liegen die Schweizer Schülerinnen und Schüler erstmals signifikant über dem OECD-Durchschnitt (vgl. BFS & EDK 2007, p. 41ff.).

Allerdings ist zu bemerken, dass der Mittelwert der OECD-Länder in der Erhebung 2006 leicht gesunken ist (492 Punkte gegenüber 494 Punkten in PISA 2003 und 500 Punkten in PISA 2000) und damit auch zur signifikanten Verbesserung der Schweiz beigetragen hat (vgl. ebd., p. 43).⁴³

Aus der Tabelle 4.2 wird ersichtlich, dass Finnland, welches im Jahr 2000 die besten Werte erreicht, 53 Punkte (ca. drei Viertel Kompetenzniveaus) über der Schweiz liegt (vgl. Meyer & Zahner, 2002, p. 30).

Vor allem im Bereich «*Reflektieren und Beurteilen*», wo die erzielten Leistungen am niedrigsten sind, zeigt sich das mässige Abschneiden der Schweiz.

«Den 15-jährigen Schülerinnen und Schülern fällt es besonders schwer, Verknüpfungen zwischen einem Text und eigenen Erfahrungen, Kenntnissen und Ideen herzustellen. Wenn es hingegen um das Herausuchen von Informationen und das Lokalisieren einzelner oder mehrerer Informationsteile in einem Text geht (Unterbereich «*Informationen herausuchen*»), stehen die Schweizer Schülerinnen und Schüler ebenso gut da wie der OECD-Durchschnitt» (ebd., p. 30).

Für den internationalen Vergleich wurden so genannte *Bezugsländer*⁴⁴ definiert, mit denen die Schweiz verglichen werden soll: *Deutschland, Österreich, Frankreich, Italien, Japan, Schweden* und die *USA*. Im Vergleich mit eben diesen Bezugsländern weist die Schweiz in PISA 2000 deutlich mehr Jugendliche auf den tiefsten Niveaus auf (<1 und 1). Nur in Deutschland sind auf diesen zwei Niveaus noch mehr Jugendliche zu finden. Auf der anderen Seite hat die Schweiz vergleichsweise weniger Jugendliche auf dem Niveau 4. Auf dem höchsten Kompetenzniveau ist der Anteil der Schweiz nicht kleiner als bei den anderen Bezugsländern (vgl. ebd., p. 32).

Im Durchschnitt weisen sämtliche OECD-Mitgliedstaaten in PISA 2000 im Bereich Lesen 9.5% Jugendliche auf, die anspruchsvolle Aufgaben bewältigen können (Niveau 5). Im Vergleich zu den 6%, die nicht in der Lage sind, einfache Aufgaben zu lösen (unter Niveau 1) bzw. den 11.9%, die nur zur Lösung einfacher Aufgaben fähig sind (Niveau 1) zeichnet sich ein grosser Abstand ab (vgl. Artelt et al. 2001, p. 103).

Die Jugendlichen unterhalb des ersten Kompetenzniveaus müssen als *Risikogruppe* gesehen werden. Diese Jugendlichen sind in Bezug auf ihre Ausbildungsmöglichkeiten und Berufschancen erheblich benachteiligt. Das betrifft in der Schweiz immerhin 7% aller getesteten Jugendlichen (vgl. Abbildung 4.5).

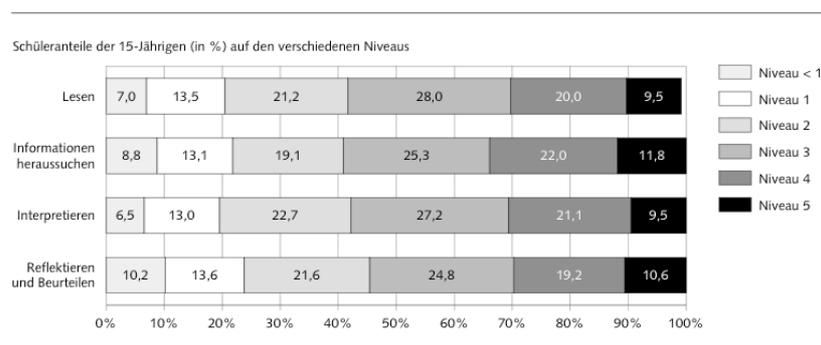


Abbildung 4.5: Kompetenzniveaus im Lesen allgemein und in den drei Unterbereichen des Lesens, PISA 2000 (aus: Meyer & Zahner 2002, p. 33)

⁴³ Das Absinken des Mittelwerts ist zum Teil dadurch zu erklären, dass die Türkei und die Slowakische Republik, die im Vergleich zum OECD-Mittelwert unterdurchschnittlich abschneiden, ab 2003 zum Kreis der teilnehmenden Länder hinzutraten (vgl. OECD 2007, p. 341).

⁴⁴ In der Tabelle 4.2 sind die sieben Bezugsländer grau markiert.

Aber auch den 13.5% der Jugendlichen, die nur das Niveau 1 erreicht haben, dürfte das Lernen in der Schule schwer fallen. Ihre Kompetenzen genügen kaum, um sich aus Texten selbständig Wissen und Kenntnisse anzueignen. Dies kann sowohl zu einer möglichen Benachteiligung im Arbeitsmarkt führen, als auch im Bereich der persönlichen Entfaltung negative Auswirkungen haben. Die Testpersonen unterhalb und auf dem ersten Kompetenzniveau sind keine zufällige Gruppe. In fast allen Ländern sind es in der Mehrheit Knaben, die in ihrer Herkunft oftmals sozial benachteiligt sind. Sei dies, weil die Eltern statusniedrigere Berufe ausüben und/oder weil sie aus dem Ausland stammen (vgl. Meyer & Zahner 2002, p. 32).

Da die getesteten Jugendlichen nicht als ganze Klassen einbezogen wurden, können die Daten nur auf Schulebene ausgewertet werden. Es kann dabei festgestellt werden, dass es einerseits Staaten gibt, in denen zwischen den einzelnen Schulen in Bezug auf die Leseleistungen nur kleine Unterschiede bestehen (z.B. Schweden), andererseits Länder wie Österreich, Deutschland und die Schweiz, die hohe *Schulvarianzen* aufweisen. Ein Teil dieser Varianzen zwischen den Schulen lässt sich durch Unterschiede zwischen städtischen und ländlichen Regionen, zwischen privaten und öffentlichen Schulen sowie zwischen den Schultypen erklären. In der Schweiz reduziert sich diese Varianz beträchtlich, wenn man die verschiedenen Schultypen, die Sprachregionen und die Bildungsstufen (Sekundarstufen I und II) berücksichtigt. Da zwischen den verschiedenen Sprachregionen bezüglich der Lesekompetenzen nur geringe Unterschiede auszumachen sind, ist die verbleibende Varianz grösstenteils auf die Merkmale der jeweiligen Schule zurückzuführen (vgl. ebd., p. 33f.).

Die *Leistungsunterschiede* der Jugendlichen werden durch viele *Faktoren*, deren Wichtigkeit nur schwer auszumachen ist, beeinflusst. Sie können u.a. auf den sozioökonomischen Hintergrund der Jugendlichen, auf die personellen und finanziellen Ressourcen der Schulen und auf unterschiedliche Lehrpläne oder die Unterrichtsorganisation zurückgeführt werden (vgl. ebd., p. 32). So gehört die Schweiz bisher zu jenen Ländern, die ihre Kinder am spätesten einschulen. Dementsprechend haben die getesteten Jugendlichen in der Schweiz durchschnittlich neun Schuljahre besucht, während der OECD-Durchschnitt bei neuneinhalb Jahren liegt. Die spätere Einschulung könnte ein möglicher Grund für die mässige Lesekompetenz sein. Die Beispiele Finnlands und Schwedens, welche trotz später Einschulung hohe Lesekompetenzen aufweisen, zeigen aber, dass andere Faktoren wohl viel entscheidender sind (vgl. ebd., p. 30).

Des Weiteren zeigen sich *Geschlechtsunterschiede*. Bei der Lesekompetenz schneiden die 15-jährigen Mädchen in PISA 2000 in allen Ländern besser ab als die gleichaltrigen Knaben. Dieser Effekt liess sich auch in den Erhebungen 2003 und 2006 bestätigen (vgl. BFS & EDK 2007, p. 41). Der Leistungsunterschied beträgt in der Schweiz in der Erhebung 2006 fast ein halbes Kompetenzniveau (genau genommen 32 Punkte) und ist damit nicht zu vernachlässigen. Unter den Vergleichsländern ist die Geschlechterdifferenz jedoch eher klein, grössere Unterschiede in der Leseleistung von Mädchen und Jungen zeigen sich in der Erhebung 2006 in Finnland, Liechtenstein, Deutschland und Belgien (vgl. ebd., p. 41).

Beim *regionalen Vergleich* der Erhebung 2000 haben die Schülerinnen und Schüler der französischsprachigen Schweiz im Bereich Lesen am besten abgeschnitten, gefolgt von der Deutschschweiz und der italienischen Schweiz. Signifikant ist jedoch nur der Leistungsunterschied zwischen der französischen und italienischen Schweiz. Er beträgt immerhin ein Viertel eines Kompetenzniveaus (also etwas weniger als 20 Punkte) (vgl. Meyer & Zahner 2002, p. 35). Der regionale Bericht für PISA 2006 zeigt, dass die Leistungsunterschiede zwischen den Kantonen der Deutschschweiz eher gering ausfallen (die Mittelwerte der Deutschschweizer Kantone liegen im Lesen innerhalb von 25 Punkten). Die Ergebnisse der Deutschschweizer Kantone fallen zudem etwas besser aus als jene der französischsprachigen Schweiz (vgl. Moser & Angelone 2008, p. 13).

4.2.1.2 Mathematik

Im Gegensatz zum Lesen, das in PISA 2000 im Mittelpunkt stand, war Mathematik Schwerpunktthema im Zyklus 2003. Die Kompetenzen in Mathematik umfassen dabei die Fähigkeiten,

«[...] mathematische Probleme in der Alltagswelt zu erkennen und zu interpretieren, diese Probleme in einen mathematischen Kontext zu übertragen, Kenntnisse und Prozeduren der Mathematik anzuwenden, um die Probleme im mathematischen Kontext zu lösen, die Lösung in der Sprache des ursprünglich formulierten Problems zu interpretieren, über die angewendeten Methoden zu reflektieren und die Ergebnisse zu formulieren und mitzuteilen» (Meyer & Zahner 2002, p. 37).

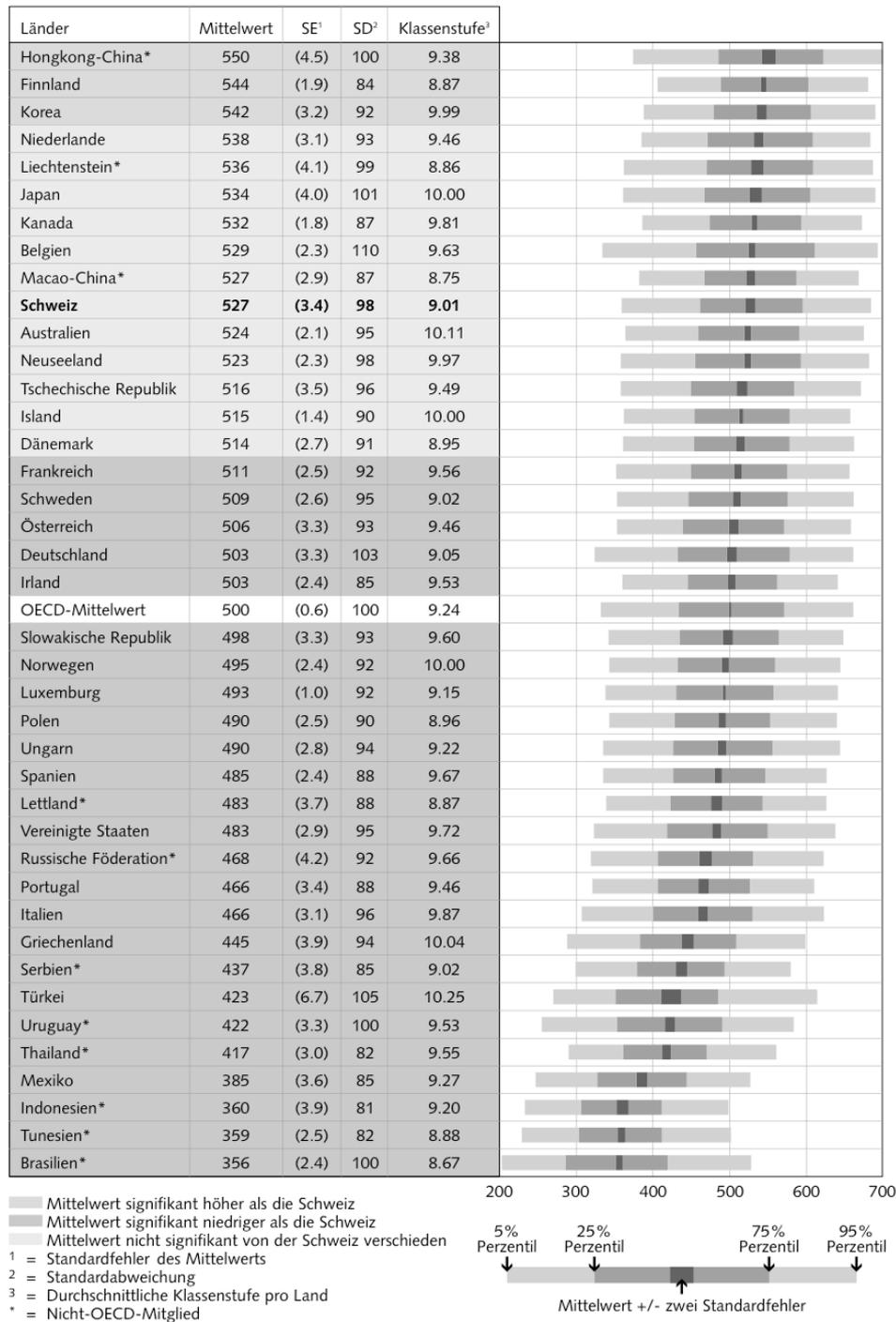
Die Schülerleistungen wurden in vier mathematischen Inhaltsbereichen gemessen (vgl. BFS & EDK 2004, p. 15):

- *Raum und Form*: räumliche und geometrische Phänomene und die Eigenschaften von Objekten (Geometrie)
- *Veränderung und Beziehungen*: Zusammenhänge zwischen Variablen und ein Verständnis der Art und Weise, in der sie dargestellt werden (Algebra)
- *Quantitatives Denken*: Zahlenphänomene sowie quantitative Beziehungen und Muster (Arithmetik)
- *Ungewissheit*: probabilistische und statistische Phänomene (Wahrscheinlichkeitsrechnung und Statistik)

Bezogen auf diese vier Bereiche wurde erfasst, inwieweit die Jugendlichen fähig sind, *mathematisches Wissen wiederzugeben, Verbindungen herzustellen, zu reflektieren und zu beurteilen*. Bei der Aufgabenstellung wurde auf verschiedene typische Lebenskontexte Bezug genommen: das private Leben, die Schule, Arbeit und Sport, die lokale Gemeinschaft und die Gesellschaft sowie die Welt der Wissenschaft (vgl. ebd., p. 15). Im Unterschied zur Lesekompetenz wurden in der Mathematik 6 Kompetenzniveaus definiert, die je eine Spannweite von 62 Punkten aufweisen (vgl. ebd., p. 16). Während Schüler und Schülerinnen auf dem untersten Niveau nur dazu fähig sind, «Fragen [zu] beantworten, die in einem vertrauten Zusammenhang formuliert sind, alle relevanten Informationen beinhalten und klar definiert sind» (ebd., p. 17), weisen jene auf Niveau 6 folgende Fähigkeiten und Fertigkeiten auf: «Konzeptualisieren, Generalisieren und Informationen verwenden, die auf komplexen Problemsituationen basieren. Zwischen verschiedenen Informationsquellen und Darstellungsformen Verbindungen herstellen und sie flexibel aufeinander übertragen. Neue Ansätze und Strategien im Umgang mit unvertrauten Situationen entwickeln» (ebd., p. 17)

Die Schweiz schneidet in PISA 2003 im *internationalen Vergleich* bei der Mathematik sehr gut ab. Ihre Leistung liegt mit 527 Punkten weit über dem OECD-Mittel. Auch PISA 2006 bestätigt die guten Ergebnisse der Schweizer Jugendlichen im Bereich Mathematik aus dem Jahre 2003. Von allen berücksichtigten Ländern schneiden in PISA 2003 nur Hongkong-China, Finnland und Korea signifikant besser ab als die Schweiz. Einen Überblick über die Leistungen in Mathematik in den verschiedenen Ländern bietet die Tabelle 4.3.

Tabelle 4.3: Mathematikleistungen im Ländervergleich, PISA 2003
(aus: BFS & EDK 2004, p. 18)



Auffallend ist, dass die Schweizer Jugendlichen mit dem in PISA 2003 neu eingeführten Gebiet «Ungewissheit» vergleichsweise am meisten Mühe hatten. Während die 15-Jährigen in Finnland, Hongkong-China und Kanada keine Leistungseinbusse auf diesem Bereich in Kauf nehmen mussten, zeigen Schülerinnen und Schüler aus Liechtenstein, Österreich und der Schweiz in diesem Bereich schwächere Leistungen (vgl. ebd., p. 17). Im zweiten nationalen Bericht (vgl. BFS & EDK 2005, p. 32) wird in diesem Zusammenhang darauf hingewiesen, dass in der Schweiz Statistik und

Wahrscheinlichkeitsrechnung – die für den Bereich «Ungewissheit» wichtigen schulischen Themengebiete – im Unterricht der obligatorischen Schule untergeordnete Bedeutung haben.

Betrachtet man die Verteilung der Leistungen auf die verschiedenen Kompetenzniveaus, so erreichen im OECD-Mittel 21 Prozent nur das Niveau 1 oder liegen darunter. Diese Schülerinnen und Schüler sind nicht in der Lage, relevante Informationen aus einer Quelle herauszusuchen und grundlegende Berechnungen und Prozeduren durchzuführen. Dass dies bei den OECD-Ländern auf ein gutes Fünftel der 15-Jährigen zutrifft ist bedenklich. In der Schweiz sind hingegen etwas weniger Jugendliche auf dieser Leistungsstufe, nämlich 15 Prozent. Zudem liegt der Anteil an Jugendlichen, die mindestens Niveau 5 erreichen, in der Schweiz mit 21 Prozent über dem OECD-Durchschnitt (vgl. BFS & EDK 2004, p. 19).

In der Mathematik ist in PISA 2000 ein negativer Zusammenhang zwischen den durchschnittlichen Leistungen eines Landes und der Streuung der Leistungen festgestellt worden. Je höher der Leistungsdurchschnitt, desto kleiner die Streuung und umgekehrt (vgl. Meyer & Zahner 2002, p. 41f.). Dies verdeutlicht Abbildung 4.6.

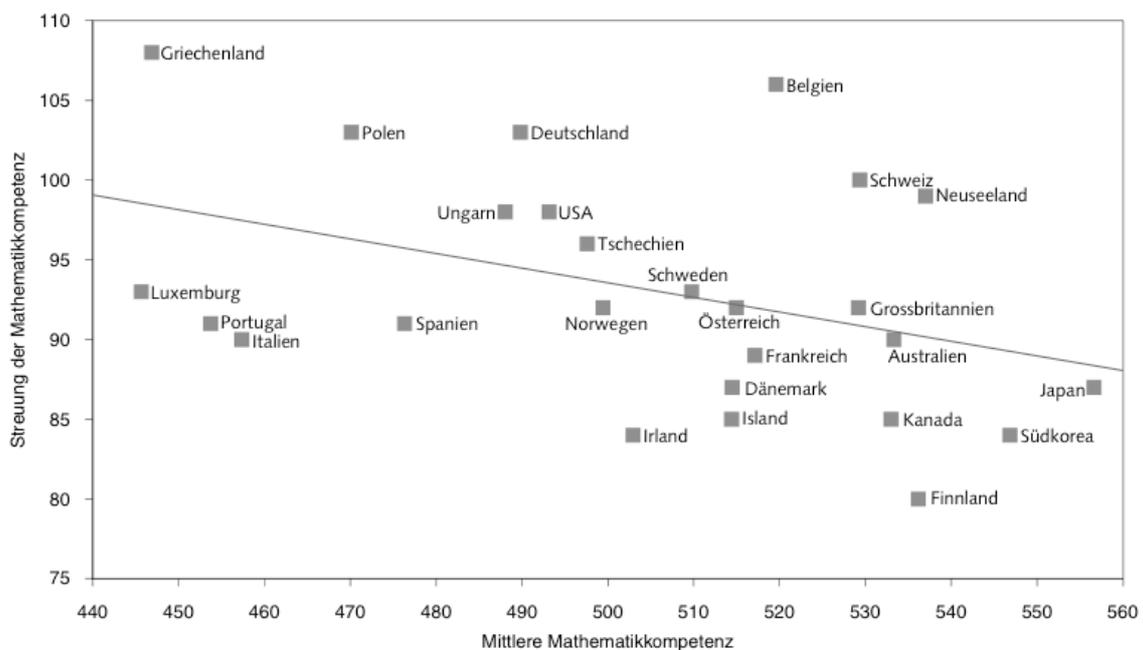


Abbildung 4.6: Streuung der nationalen Mathematikkompetenz im Vergleich zu deren Mittelwert, PISA 2000 (aus: Meyer & Zahner 2002, p. 41)

Als Extrembeispiele gelten Finnland und Griechenland. Während die getesteten Jugendlichen in Griechenland eine geringe mittlere Mathematikkompetenz und eine grosse Streuung aufweisen, ist es in Finnland genau umgekehrt. Hier ist das durchschnittliche Leistungsniveau der Schülerinnen und Schüler gross und die Streuung klein. In der Schweiz ist demgegenüber die Streuung trotz hoher mittlerer Mathematikkompetenz überraschend hoch (vgl. ebd., p. 42).

Meyer und Zahner (2002) erklären sich den eben geschilderten Zusammenhang folgendermassen:

«Eine geringe Leistungsstreuung bei gleichzeitig mittlerem bis hohem Leistungsniveau könnte eine Erfolgsbestätigung des öffentlichen Bildungsauftrages sein, möglichst allen Schülerinnen und Schülern gleich gute Bildungschancen zukommen zu lassen. Je stärker in dieser Hinsicht bildungspolitische Bemühungen unternommen werden und je besser diese greifen, desto mehr Schülerinnen und Schüler werden zumindest eine durchschnittliche Grundbildung erreichen und desto kleiner wird der Anteil an Kin-

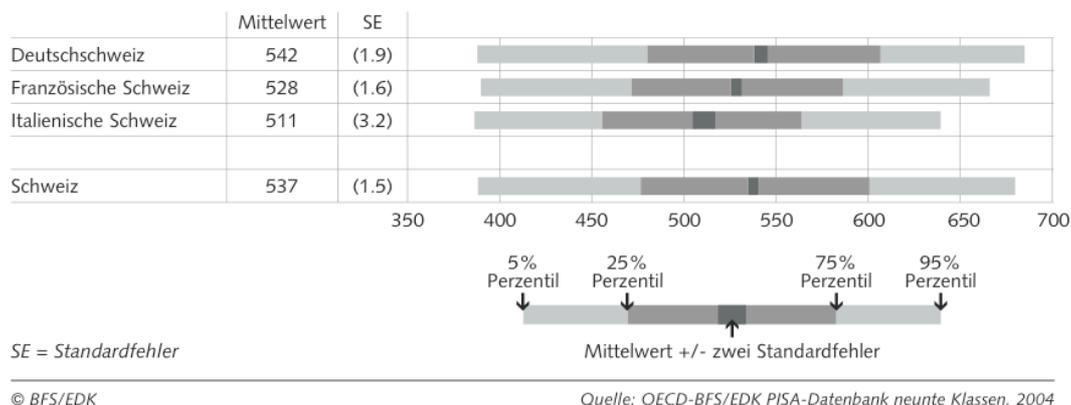
dem mit extrem schwachen Leistungen, mit anderen Worten, es verringert sich die Schere zwischen den Höchst- und den Tiefstleistungen, wobei sich im Idealfall nur die letzteren angleichen sollten» (ebd., p. 42).

Ein *Indikator* für gute mathematische Leistungen scheint die wöchentliche Anzahl Lektionen in diesem Fach zu sein. So erhalten die 15-jährigen Schülerinnen und Schüler in der Schweiz, wie auch in Japan, durchschnittlich ein bis zwei Unterrichtsstunden mehr pro Woche als die Schüler anderer Bezugsländer (vgl. ebd., p. 41).

Wie bereits bei der Lesekompetenz zeigt sich auch bei den Leistungen in Mathematik ein Einfluss des *sozioökonomischen Hintergrunds*. Dieser fällt in der Schweiz für die Erhebung 2003 jedoch im Vergleich mit den anderen OECD-Ländern mittelmässig aus und nicht wie bei der Lesekompetenz in PISA 2000 überdurchschnittlich (vgl. BFS & EDK 2004, p. 21).

Bezüglich eines *Geschlechtereffekts* zeigt sich bei den Mathematikleistungen eine signifikante leichte Überlegenheit der Knaben. Sie schneiden in PISA 2003 in der Schweiz im Durchschnitt 17 Punkte besser ab als die Mädchen. Auch in 20 weiteren OECD-Ländern zeigt sich eine leichte Überlegenheit der Knaben (vgl. ebd., p. 19). Interessant ist auch, dass die Schweizer Mädchen im unteren Bereich der Skala nur leicht übervertreten sind, jedoch bedeutend weniger Mädchen die oberen Kompetenzniveaus erreichen (vgl. ebd., p. 20).

Tabelle 4.4: Mathematikleistungen nach Sprachregionen, PISA 2003
(aus: BFS & EDK 2005, p. 18)



Wie bereits bei der Lesekompetenz schneiden die Neuntklässlerinnen und Neuntklässler der deutschsprachigen Schweiz in PISA 2003 auch in der Mathematik im *regionalen Vergleich* zwischen den Sprachregionen am besten ab, gefolgt von der französisch- und der italienischsprachigen Schweiz (vgl. BFS & EDK 2005, p. 19). Vergleicht man die Ergebnisse der italienischen Schweiz mit den internationalen Daten, dann liegt sie jedoch immer noch im Mittelfeld.

Schon in TIMSS⁴⁵ wurden den Schweizer Schülerinnen und Schülern gute Zeugnisse bezüglich ihrer mathematischen Kompetenzen in der obligatorischen Schule ausgestellt. Diese Ergebnisse lassen sich durch PISA 2000, 2003 und 2006 bestätigen.

⁴⁵ Vgl. Kapitel 3.

4.2.1.3 Naturwissenschaften

Die naturwissenschaftliche Grundbildung wurde in PISA 2000 und 2003 nur auf einer Skala gemessen. Erst die Erhebung 2006, in der Naturwissenschaften im Zentrum standen, erfasste diesen Bereich differenzierter. Dadurch ergibt sich jedoch auch die methodische Einschränkung, dass die Leistungen eines Landes zu den verschiedenen Erhebungszeitpunkten nicht exakt miteinander verglichen werden dürfen (vgl. BFS & EDK 2007, p. 23). Methodisch einwandfreie Trendschätzungen der Kompetenzen in den Naturwissenschaften sind daher erst auf der Grundlage der Daten aus PISA 2006 legitim.

In PISA 2006 wurden drei Prozesse naturwissenschaftlichen Denkens unterschieden und auf unabhängigen Skalen erfasst (vgl. ebd., p. 21):

- Erkennen von naturwissenschaftlichen Fragestellungen
- Erklären von naturwissenschaftlichen Phänomenen
- Anwenden von naturwissenschaftlicher Erkenntnis

Im Unterschied zur Lesekompetenz wurden für die Naturwissenschaften wie bereits für die Mathematik 6 Niveaus definiert. Schüler und Schülerinnen auf Niveau 1 sind nur dazu fähig, naturwissenschaftliches Wissen auf wenige vertraute Situationen anzuwenden. Jugendliche auf Niveau 6 erkennen demgegenüber naturwissenschaftliches Wissen in verschiedenen komplexen Lebenskontexten und können ihr Wissen anwenden, um Entscheide in persönlicher und sozialer Hinsicht zu rechtfertigen (vgl. ebd., p. 17).

Im Bereich Naturwissenschaften unterscheiden sich die gesamthaft aufgewendete Unterrichtszeit sowie der Aufbau der Lehrpläne zwischen den Ländern viel stärker als beim Lesen oder der Mathematik (vgl. Meyer & Zahner 2002, p. 43).⁴⁶ Dieser Umstand sollte bei der Interpretation der Ergebnisse berücksichtigt werden. Wie bereits beim Lesen kann auch bei der naturwissenschaftlichen Grundbildung das Einschulungsalter als Einflussgrösse für die Erklärung der Unterschiede herangezogen werden. Es ist dabei zu bedenken, dass die Schweiz bisher spät einschult und dass ihre Jugendlichen zum Zeitpunkt des Tests weniger Schuljahre absolviert haben als der OECD-Durchschnitt. Es wären unter Umständen bessere Testergebnisse zu erwarten, wenn die mittlere Klassenstufe der Schweiz höher wäre, d.h. auf dem internationalen Durchschnittsniveau (vgl. ebd., p. 48).

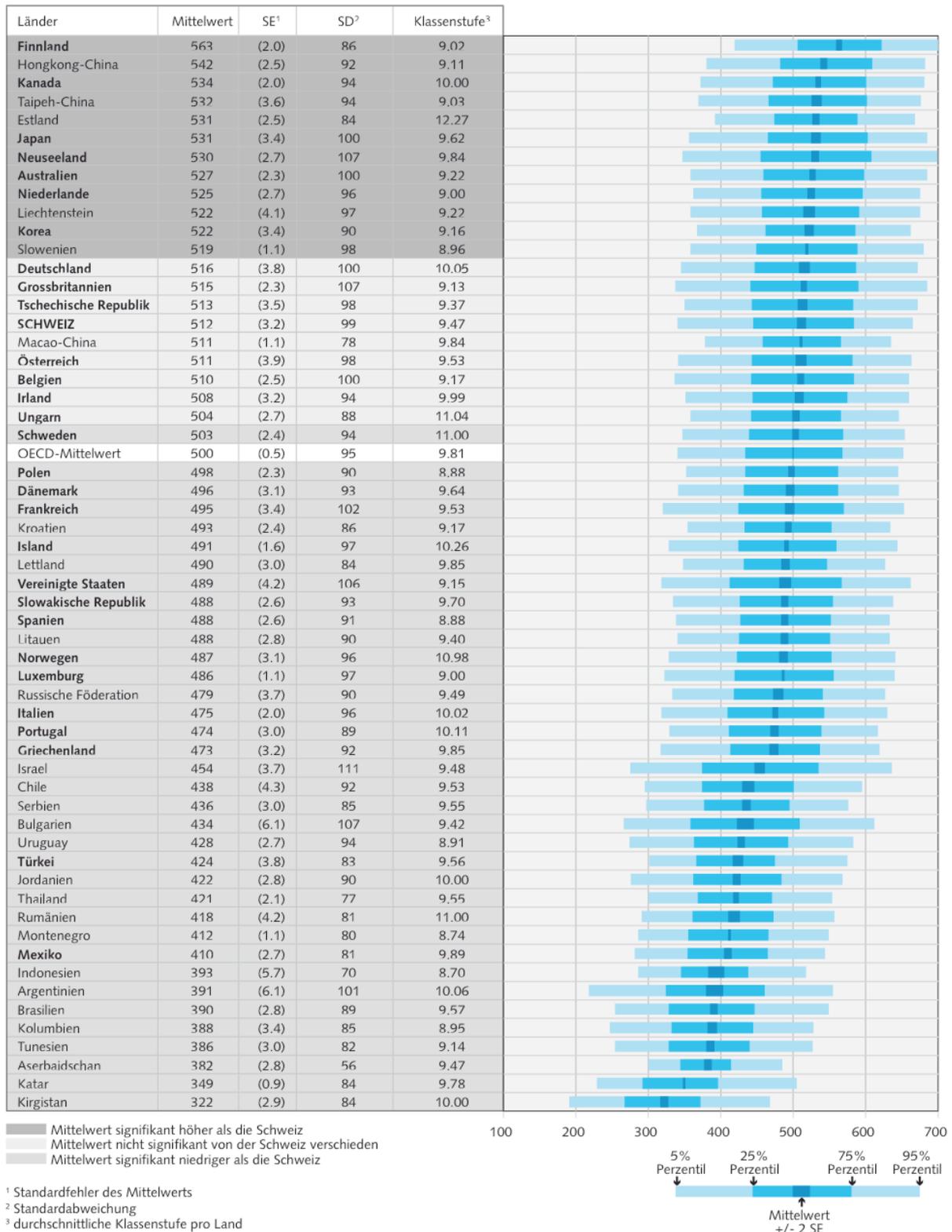
In PISA 2006 liegt der Mittelwert der Schweizer Jugendlichen mit 512 Punkten deutlich über dem OECD-Durchschnitt (vgl. Tabelle 4.5). 12 Länder erzielen bessere Ergebnisse als die Schweiz. Die besten Leistungen erzielen Finnland (563), Hongkong-China (542) und Kanada (534). In PISA 2000 lag die Schweiz in den Naturwissenschaften – wie bereits beim Lesen – hingegen noch im Mittelfeld.

Betrachtet man die Schülerinnen und Schüler mit sehr geringen naturwissenschaftlichen Kompetenzen auf Niveau 1, so ist diese Gruppe in der Schweiz vergleichsweise klein. Während im OECD-Durchschnitt rund 19 Prozent nicht das Niveau 2 erreichen, beträgt dieser Anteil in der Schweiz nur 16 Prozent (vgl. ebd., p. 19). In Finnland, Estland und Hongkong-China ist der Anteil der Jugendlichen auf Niveau 1 noch geringer, hingegen weisen Frankreich (21%) und Italien (25%) mehr Jugendliche mit sehr beschränkten naturwissenschaftlichen Kompetenzen auf (vgl. ebd., p. 20).

⁴⁶ «Der Bereich Naturwissenschaften wird in der Schweiz teilweise in die Fächer Biologie, Chemie, Physik und Erdkunde unterteilt und getrennt unterrichtet, teilweise aber auch integriert vermittelt und entsprechend nur als ein Fach benotet. International gesehen wird mit dem Unterricht der Naturwissenschaften zu verschiedenen Zeitpunkten in der obligatorischen Ausbildung begonnen. Ausserdem wird diesen Fächern auch unterschiedliches Gewicht beigemessen» (Meyer & Zahner 2002, p. 43).

Während in den Bereichen Lesen und Mathematik jeweils klare Kompetenzunterschiede bezüglich des *Geschlechts* auszumachen sind, ist bei den naturwissenschaftlichen Kompetenzen kein bedeutender Unterschied in den Leistungen von Knaben und Mädchen erkennbar (vgl. ebd., p. 23).

Tabelle 4.5: Naturwissenschaftliche Leistungen im Ländervergleich, PISA 2006
(aus: BFS & EDK 2007, p. 20)



Allerdings zeigt sich ein deutlicher Einfluss des *sozioökonomischen Hintergrunds* auf die Leistung in Naturwissenschaften: «Die Leistung einer Schülerin oder eines Schülers mit sozioökonomisch stark benachteiligtem Hintergrund (im untersten Sechstel der Verteilung des sozioökonomischen Hintergrunds) ist um ein Kompetenzniveau niedriger als die eines/einer privilegierten Jugendlichen (im obersten Sechstel)» (ebd., p. 22f.). Dieser Einfluss ist in der Schweiz vergleichbar mit dem OECD-Mittel.

Ebenfalls zeigen *immigrierte Jugendliche* und solche mit einer *Muttersprache*, die nicht der Testsprache entspricht, tiefere Leistungswerte als Einheimische. Dieses Handicap ist in den deutschsprachigen Ländern besonders gross (vgl. ebd., p. 23).

Der bei den Mathematikleistungen gefundene negative Zusammenhang zwischen den durchschnittlichen Leistungsniveaus eines Landes und der Streuung der Leistungen konnte für PISA 2000 auch für die Naturwissenschaften repliziert werden (vgl. Abschnitt 4.2.1.2).

In der Erhebung 2000 zeigte sich in Bezug auf die *sprachregionalen Unterschiede* das gleiche Bild wie in der Mathematik:

«Die Neuntklässlerinnen und Neuntklässler der französischen Schweiz erbringen die besten Leistungen, gefolgt von der Deutschschweiz und der italienischen Schweiz. Wiederum ist v.a. der Abstand der italienischen Schweiz zu den beiden anderen Sprachregionen erwähnenswert [...]. Zwischen der Deutschschweiz und der französischen Schweiz hingegen sind die Unterschiede klein und nicht signifikant» (Meyer & Zahner 2002, p. 48).

Auch die sprachregionalen Auswertungen von PISA 2006 zeigen, dass die naturwissenschaftlichen Kompetenzen je nach Sprachregion variieren. Es sind nunmehr aber die Schülerinnen und Schüler aus der Deutschschweiz ($M = 518$), welche signifikant besser abschneiden als jene der französischen ($M = 502$) und italienischen Schweiz ($M = 501$). Der Unterschied zwischen der französischen Schweiz und der italienischen Schweiz ist hingegen nicht signifikant (vgl. Homepage PISA Schweiz⁴⁷).

4.2.2 Soziale Herkunft und Chancengleichheit

Die Verwirklichung von Chancengleichheit in der Schule ist eines der wichtigsten Ziele der Bildungspolitik in der Schweiz. Schule sollte den Einfluss ungleicher Lernvoraussetzungen auf die Leistung aufheben oder zumindest verringern. Beträchtliche Anstrengungen diesbezüglich haben zu einer gewissen Öffnung und Demokratisierung des Bildungssystems geführt. Allerdings haben davon Kinder aller sozialen Schichten gleichermassen profitiert. Angehörige tiefer sozialer Schichten bleiben so im Gegensatz denjenigen hoher sozialer Schichten in Bezug auf die in der Schule vermittelten Kompetenzen benachteiligt. Besonders nachteilig für das spätere Berufsleben wirkt sich dies beim Übergang von der obligatorischen Schulzeit in eine Berufsausbildung oder eine weiterführende Schule aus.

Mit den PISA-Studien will man den *Zusammenhang zwischen sozialer Herkunft und der Schulleistung* – in PISA 2000 insbesondere der Lesekompetenz – genauer untersuchen. Die statistischen Messwerte basieren dabei nicht auf Beobachtungen, sondern auf Angaben der Schülerinnen und Schüler. Ergänzt werden diese durch Angaben der Schulleitung zum Lernumfeld, zur Schulorganisation und zum sozioökonomischen Kontext, in dem das Lernen stattfindet.

Der *sozioökonomische Status* beeinflusst die berufliche Stellung der Eltern, den Besitz und die Aktivitäten in Verbindung mit «klassischen» Kulturgütern wie Büchern und Kunstgegenständen,

⁴⁷ www.pisa.admin.ch >Wichtige Resultate > PISA 2006 > Naturwissenschaften.

die Bildung der Eltern und ihr Kommunikationsverhalten gegenüber den Kindern, den familiären Wohlstand, die Familienstruktur, den Geburtsort und die zuhause gesprochene Sprache (vgl. Moser 2001, p. 18f.).

Eine hohe *berufliche Stellung der Eltern* kann die Ziele und Einstellungen der Schülerinnen und Schüler beeinflussen. So erreichen im internationalen Durchschnitt Schülerinnen und Schüler, deren Eltern angesehene Berufe ausüben, höhere Leistungen (vgl. Moser 2002, p. 131). Dabei ist der Vorteil in einigen Ländern aber wesentlich grösser als in anderen. Die grössten Zusammenhänge zwischen der beruflichen Stellung der Eltern und der Leistung weisen in PISA 2000 Deutschland, Belgien und die Schweiz auf, die kleinsten Zusammenhänge Korea, Finnland und Island (vgl. Moser 2001, p. 19f.).

In der Schweiz zeigt eine getrennte Analyse der drei *Sprachregionen*, dass der Einfluss der beruflichen Stellung des Vaters bzw. der Mutter auf die Lesekompetenz in der italienischen Schweiz signifikant schwächer ausfällt als in der Deutschschweiz.

Der internationale Vergleich verweist weiter auf die Bedeutung des *ökonomischen Hintergrundes*. So erbringen Jugendliche aus wohlhabenderen Familien in der Regel bessere Leistungen. In den Vereinigten Staaten, in Brasilien, Mexiko und Portugal ist dieser Zusammenhang in PISA 2000 wesentlich ausgeprägter als in den nordeuropäischen Ländern, sowie Österreich, Belgien, Italien, Japan, Lettland und Polen, wo ein relativer Mangel an Wohlstand kein Hindernis für Leistungen über dem internationalen Durchschnitt bedeutet.

Ein hohes *Bildungsniveau* der Eltern beeinflusst international gesehen, besonders aber in Deutschland, Mexiko, Belgien und der Schweiz, die Leistungen der Jugendlichen positiv (vgl. BFS & EDK 2002, p. 97ff.). In Australien, Finnland und Korea schneiden auch Schülerinnen und Schüler mit weniger gebildeten Eltern im internationalen Vergleich gut ab. So stellt sich die Frage, welche anderen Bildungs- und Sozialfaktoren dort wohl die elterlichen Bildungsdefizite zu kompensieren vermögen.

Besonders wichtig für die Lesekompetenzen der Jugendlichen ist der *Besitz von Büchern*, der als Ausdruck einer bildungsnahen Einstellung gewertet werden kann und eng mit der Ausbildung der Eltern zusammenhängt (vgl. ebd., p. 104ff.). Bedeutungsvoll für die Leistungen der Jugendlichen ist auch die gute Ausrüstung zuhause mit Bildungsressourcen wie einem Internetanschluss, einem ruhigen Platz zum Lernen, einem Wörterbuch usw. (vgl. ebd., p. 105ff.). Auch häufiges Diskutieren hängt positiv mit den Leistungen der Jugendlichen zusammen (vgl. ebd., p. 106f.).

Einen Einfluss auf die Schulleistungen scheint auch die *familiäre Lebensform* der Jugendlichen auszuüben (vgl. ebd., p. 102). In den Niederlanden, Grossbritannien, den Vereinigten Staaten, Belgien, Finnland und Frankreich erzielen Jugendliche aus Ein-Eltern-Familien deutlich niedrigere Leistungen gegenüber Jugendlichen aus klassischen Zwei-Eltern-Familien. In der Schweiz wurde kein solcher Unterschied festgestellt. Bei der Auswertung dieser Resultate ist zu beachten, dass es sich um einen sehr komplexen Zusammenhang handelt, bei dem viele soziale Hintergrundfaktoren eine Rolle spielen.

Bei der *Interpretation* der oben dargestellten Befunde stellt sich die Frage, über welche weiteren Kanäle der sozioökonomische Hintergrund auf die Lesekompetenz einwirkt. Hat die berufliche Stellung der Eltern auch Einfluss auf den Wohnort der Jugendlichen und hängt so mit der Qualität der Schule, zu der sie Zugang haben, zusammen? Können bewusste Wohnortsentscheidungen diese Effekte noch akzentuieren? Beeinflusst die mit dem sozioökonomischen Kapital verbundene soziale Stellung der Eltern die Erwartungen an das Kind, steigert so seine Hoffnungen und Bestrebungen und regt es damit zu intensiverem Lernen an? Erweitert die berufliche Stellung der Eltern den Horizont von erreichbaren Möglichkeiten, den ein Kind wahrnimmt?

Die Kompetenzen von Jugendlichen entstehen stets durch das Zusammenwirken vieler Bedingungsfaktoren auf mehreren Ebenen. Eine abschliessende Beurteilung des Zusammenhangs von sozialer Herkunft und fachlicher Kompetenz ist nicht möglich. Wichtig ist die Erkenntnis, dass die soziale

Herkunft trotz jahrzehntelangen Bemühungen um schulische Chancengleichheit ihre Bedeutung als Einflussgrösse behält.

4.2.3 Kulturelle Vielfalt in der Schule: Herausforderung und Chance

Die Schule ist durch die *zunehmende kulturelle Vielfalt* in der Schweiz gefordert. Seit 1980 ist der Anteil Jugendlicher ausländischer Nationalität auf der Sekundarstufe I von 14 Prozent auf 18 Prozent im Jahr 1990 und auf rund 21 Prozent im Jahr 2000 angestiegen. Dies ist für das Bildungssystem vor allem deshalb von Bedeutung, weil ein grosser Teil der Kinder und Jugendlichen aus immigrierten Familien die Unterrichtssprache nur ungenügend beherrscht. Für das erfolgreiche Abschneiden in der Schule ist dies aber eine entscheidende Voraussetzung (vgl. Moser 2002, p. 113).

Kinder und Jugendliche mit mangelnden Kenntnissen der Unterrichtssprache sind oft mehrfach benachteiligt. So sind auch mathematische Probleme oft auf sprachlich hohem Niveau formuliert. Das Leistungspotenzial dieser Schülerinnen und Schüler wird aufgrund der mangelnden Kenntnisse der Unterrichtssprache oft unterschätzt. Damit ist einer der Gründe genannt, die das häufige Schulversagen von Kindern und Jugendlichen aus immigrierten Familien erklären können. Diese Jugendlichen besuchen auf der Sekundarstufe I überdurchschnittlich öfter Schulen mit Grund- und weniger Schulen mit erweiterten Ansprüchen (vgl. ebd., p. 113f.).

Schwierigkeiten mit Schülerinnen und Schülern auf Grund ihrer Muttersprache werden von Behörden, Eltern und Lehrpersonen oft nicht als pädagogische, soziale und didaktische Herausforderung, sondern als ethnisches Problem wahrgenommen. Schülerinnen und Schüler mit mangelnden Kenntnissen der Unterrichtssprache werden zudem für die sinkende Qualität von Schulen mitverantwortlich gemacht. Die Situation wird von betroffenen Schweizer Eltern gelegentlich so dramatisch wahrgenommen, dass getrennte Klassen für fremdsprachige Kinder gefordert werden (vgl. ebd., p. 115).

Die *interkulturelle Pädagogik* hält dieser Position eine grundsätzlich neue Haltung entgegen: Bereicherung durch kulturelle Vielfalt, wobei die Kenntnis der Unterrichtssprache für die erfolgreiche Integration aller Kinder und Jugendlicher eine unabdingbare Voraussetzung darstellt (vgl. ebd., p. 115f.). Massnahmen, zur Behebung des Problems, können nicht nur aufgrund ungenügender Schulleistungen erfolgen. Sie müssen sich auf die Beschreibung der Lernbedingungen der Schülerinnen und Schüler und auf die Überprüfung der Wirkungen der getroffenen Massnahmen abstützen. Die *linguistische Perspektive* betont dabei die Mehrsprachigkeit dieser Kinder. Für die *Bildungssoziologie* ist die Herkunft aus bildungsfernen, sozioökonomisch benachteiligten Familien zentral. Für die Schule ist die Verweildauer der Kinder im Sprachgebiet von Interesse (vgl. ebd., p. 116).

In PISA 2000 wurden für die Analyse des *Einflusses des Geburtsorts* auf die Leistungen drei Gruppen unterschieden und miteinander verglichen (vgl. OECD 2001, p. 180):

- *Gruppe 1*: Im Inland geborene Jugendliche mit mindestens einem im Inland geborenen Elternteil
- *Gruppe 2*: Im Inland geborene Jugendliche mit im Ausland geborenen Eltern
- *Gruppe 3*: Im Ausland geborene Jugendliche mit im Ausland geborenen Eltern

Jugendliche der Gruppe 3 liegen in der Lesekompetenz international hinter den Jugendlichen der Gruppe 1 und 2 zurück. Schülerinnen und Schüler, die im Ausland geboren wurden, erzielten auch nach Berücksichtigung sämtlicher anderer Einflussfaktoren in den meisten Ländern niedrigere Ergebnisse als ihre im Inland geborenen Mitschülerinnen und Mitschüler. Dies trifft allerdings nicht

für alle Länder zu. Während in Österreich, Liechtenstein, Mexiko und in der Schweiz die Leistungsunterschiede in der Lesekompetenz sehr gross sind, fällt der Unterschied in den Vereinigten Staaten beispielsweise gering und statistisch nicht signifikant aus (vgl. ebd., p. 231).

Allerdings liegt auch Gruppe 2 mit einem bedeutsamen Anteil an Immigrantinnen und Immigranten in den meisten Ländern weit unter dem Niveau der Gruppe 1. Dies obwohl beide Gruppen im Erhebungsland geboren wurden und der schulische Werdegang ähnlich sein sollte. In Luxemburg (18%), Frankreich (10%), Belgien (8%) und in der Schweiz (9%) leben vergleichsweise viele Angehörige von Gruppe 2. In diesen Ländern sind zudem signifikante bzw. deutliche Leistungsunterschiede zu den Angehörigen der Gruppe 1 festzustellen. Aus bildungspolitischer Perspektive geben diese Resultate zu denken, wenn das Ziel die Integration von Jugendlichen aus immigrierten Familien sein soll (vgl. ebd., p. 181).

In der Schweiz sprechen die meisten (86,9%) Jugendlichen aus immigrierten Familien (Gruppe 3) zuhause nicht die Unterrichtssprache, so dass die PISA-Ergebnisse bei ihnen Auskunft über das Leseverstehen in der ersten Fremdsprache geben (vgl. Moser 2002, p. 120). Weiter zeigt sich, dass bei gleicher sozialer Herkunft und innerhalb des gleichen Schultyps der Rückstand in der Lesekompetenz von Jugendlichen, deren Muttersprache sich von der Unterrichtssprache unterscheidet, nur gering ist. Die Ergebnisse zeigen also, dass neben der sprachlichen Sozialisation andere Merkmale für die Lesekompetenz von Bedeutung sind. Eine grosse Bedeutung kommt in diesem Zusammenhang dem sozioökonomischen Status zu. So überlagern sich sprachliche Probleme mit dem Einfluss des sozioökonomischen Hintergrunds. Diese Erkenntnis kann Schulen und Lehrpersonen helfen, Probleme fremdsprachiger Kinder auch von einer anderen Seite zu beleuchten und nach neuen Lösungen zu suchen (vgl. ebd., p. 121).

Insgesamt weisen die in der Schweiz getesteten Jugendlichen aus immigrierten Familien gegenüber Jugendlichen aus nicht immigrierten Familien in allen drei Bereichen (Lesen, Mathematik und Naturwissenschaften) etwa gleich grosse Leistungsrückstände auf. Mit zunehmender Verweildauer im Sprachgebiet nehmen die Defizite in allen Bereichen ab. So hat sich gezeigt, dass Jugendliche, die kurz nach der Geburt in die Schweiz kamen, im Lesen, der Mathematik und in den Naturwissenschaften nur noch geringfügig schlechter abschnitten als die einheimischen Schülerinnen und Schüler (vgl. ebd., p. 123, 129).

Jugendliche aus immigrierten Familien sind meist in mehrfacher Weise benachteiligt. In der Schule sind sie in doppelter Weise gefordert: Sie müssen die Unterrichtssprache lernen und zugleich den Unterricht verstehen können. Die sprachlichen Schwierigkeiten stellen sie in sämtlichen Fächern vor besonders hohe Anforderungen (vgl. ebd., p. 129). Weiter sind immigrierte Familien überdurchschnittlich häufig sozioökonomisch benachteiligt. Die Eltern verfügen über tiefere Bildungsabschlüsse und kennen das schweizerische Schulsystem kaum. Somit können sie ihren Kindern weniger Unterstützung für eine erfolgreiche Schullaufbahn bieten (vgl. ebd., p. 130). Im Vergleich zu einheimischen Familien besitzen immigrierte Familien zusätzlich weniger Bücher, seltener einen Computer mit einem Internetanschluss und ihre Kinder haben seltener ein eigenes Zimmer. Ferner zeigt sich die Bildungsnähe der Eltern auch in der Art und Weise der Unterstützung ihrer Kinder. So erhalten Jugendliche aus immigrierten, bildungsfernen Familien weniger Unterstützung beim Erledigen der Hausaufgaben (vgl. ebd., p. 131f.).

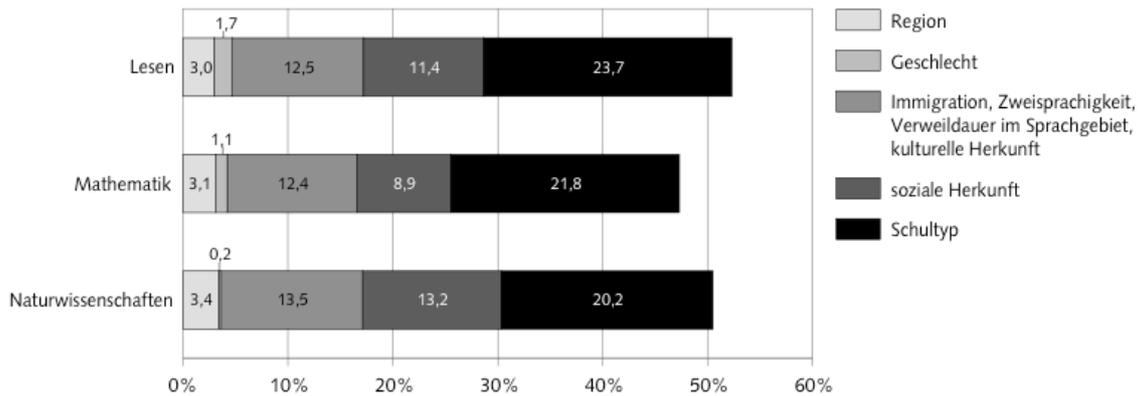


Abbildung 4.7: Erklärungsanteil verschiedener Merkmale für die Leistungsunterschiede in Lesen, Mathematik und Naturwissenschaften, PISA 2000 (aus: Moser 2002, p. 133)

Die Benachteiligung der Jugendlichen aus immigrierten Familien kann durch verschiedene Faktoren erklärt werden (vgl. Abbildung 4.7). Erwähnenswert ist dabei, dass der Immigrationsstatus, die Verweildauer im Sprachgebiet und die kulturelle Herkunft eng miteinander zusammen hängen, so dass sie nicht als einzelne Faktoren, sondern gleichsam als Erklärungspaket herangezogen werden (vgl. ebd., p. 132). Aus der Abbildung 4.7 ist zu entnehmen, dass im Gegensatz zur Herkunftsregion und dem Geschlecht die Merkmale *Immigration, Zweisprachigkeit, Verweildauer im Sprachgebiet, kulturelle Herkunft*, die *soziale Herkunft* sowie der *Schultyp*, den grössten Erklärungsanteil für die Leistungsunterschiede ausmachen.

Häufig sind Schulen und Lehrpersonen nicht ausreichend auf interkulturelle Herausforderungen vorbereitet. Kinder und Jugendliche aus immigrierten Familien sind ein Teil unserer zukünftigen Gesellschaft.

4.2.4 Geschlechterunterschiede

Die Geschlechtszugehörigkeit bestimmt die Identität der Jugendlichen wesentlich und wirkt sich grundlegend auf den persönlichen und schulischen Entwicklungsverlauf aus. Deshalb gehören Leistungsvergleiche zwischen Mädchen und Knaben auch zu den zentralen Vergleichen in der Forschungsliteratur (vgl. Malti 2002, p. 138).

PISA 2000 zeigt, dass die grössten und konsistentesten Geschlechterunterschiede im Bereich Lesen auszumachen sind. In allen Staaten, die an PISA 2000 teilgenommen haben, zeigen Mädchen signifikant bessere Leistungen als die Knaben (vgl. Stanat & Kunter 2001, p. 253). In der Schweiz zeigen die Mädchen in allen drei getesteten Bereichen der Lesekompetenz («Interpretieren», «Informationen herausuchen» und «Reflektieren und Beurteilen») höhere Leistungswerte. Am deutlichsten ist der Rückstand der Jungen im Bereich «Reflektieren und Beurteilen», wo es um die Fähigkeit geht, Bezüge zu den eigenen Erfahrungen herzustellen (vgl. ebd., p. 142f.). Mädchen haben zudem mehr Freude am Lesen. Jungen lesen dagegen tendenziell nur, wenn sie müssen (vgl. ebd., p. 262).

In der Mathematik weisen Mädchen geringere Leistungen auf als Knaben. Die Unterschiede sind geringer als in früheren Untersuchungen (z.B. TIMSS), vor allem aber auch geringer als in der Lesekompetenz, wo die Mädchen vergleichsweise besser abschneiden als die Knaben in der Mathematik. Dies lässt sich möglicherweise auf die eingesetzten Testaufgaben zurückführen, die auch im mathematischen Bereich eine relativ hohe Lesekompetenz voraussetzen. In allen Ländern sind im Lesen prozentual mehr Knaben als Mädchen in den unteren Kompetenzniveaus vertreten. Im Bereich der Mathematik sieht die Verteilung anders aus. Hier erklärt sich der Vorsprung der Knaben

nicht dadurch, dass prozentual mehr Mädchen den unteren Kompetenzniveaus angehören, sondern dass prozentual mehr Knaben als Mädchen die oberen Leistungsniveaus erreicht haben (vgl. Malti 2002, p. 143f.).

In den Naturwissenschaften zeigen Jungen und Mädchen in PISA 2006 ähnliche Kompetenzen. Zwar lassen sich bei einigen Ländern statistisch signifikante Unterschiede feststellen, diese sind aber meist gering. So sind die 6 Punkte Unterschied in der Schweiz inhaltlich nicht interpretierbar (vgl. BFS & EDK 2007, p. 23). Allerdings zeigen sich deutliche Geschlechtsunterschiede im naturwissenschaftlichen Selbstkonzept: Die Schweizer Knaben sind stärker von ihren naturwissenschaftlichen Kompetenzen überzeugt als die Mädchen (vgl. ebd., p. 29). Zudem sind sie besser über die Umweltproblematik informiert und stufen allgemeine und persönliche Bedeutung der Naturwissenschaften höher ein (vgl. ebd., p. 31).

In PISA 2000 wurde der Zusammenhang der Geschlechterunterschiede mit den Schultypen auf der Sekundarstufe I untersucht. Im Bereich der Sekundarstufe I gibt es in der Schweiz Schultypen mit

- höheren Ansprüchen (z.B. Gymnasien),
- mit erweiterten Ansprüchen (z.B. Sekundarschulen),
- mit Grundansprüchen (z.B. Realschulen),
- integrierte Schulformen (kooperative oder integrierte Schulmodelle, bei denen Schülerinnen und Schüler fachweise in Leistungsniveaus mit unterschiedlichen Ansprüchen eingeteilt werden)
- sowie Sonderklassen (Kleinklassen) (vgl. Malti 2002, p. 144).

In den Bereichen Mathematik und Naturwissenschaften zeigen die Knaben in den Schultypen mit hohen, erweiterten und Grundansprüchen bessere Leistungen als die Mädchen. Interessant ist die Tatsache, dass sich Mädchen und Knaben in integrierten Schulformen hinsichtlich ihrer erzielten Leistungen nicht unterscheiden. Im Lesen weisen die Mädchen im Vergleich zu den Knaben in allen Schultypen ausser den Sonderklassen höhere Leistungswerte auf. Am deutlichsten fällt der Leistungsrückstand der Knaben in integrierten Schulformen aus. Einzig in den Sonderklassen weisen die Mädchen im Lesen schlechtere Leistungen als die Knaben auf. Zu beachten ist bei der Interpretation dieses eher nicht zu erwartenden Befundes, dass es sich bei diesem Schultyp um eine sehr kleine Anzahl von untersuchten Jugendlichen handelt (vgl. ebd., p. 154).

Die Untersuchung *individueller Lernvoraussetzungen* ist zentral, wenn Unterschiede in den Schulleistungen von Mädchen und Knaben erklärt werden sollen. Dazu wenden wir uns innerhalb des Bereichs selbstreguliertes Lernen dem Selbstkonzept (Vertrauen in die eigenen Kompetenzen in einem bestimmten Fach) und dem Interesse zu. In praktisch allen PISA-Teilnehmerstaaten weisen die Jungen sowohl ein geringeres Interesse am Lesen als auch ein geringeres Selbstkonzept im verbalen Bereich auf als die Mädchen. Letztere haben deutlich mehr Freude am Lesen und sind in diesem Bereich auch stärker von sich selbst überzeugt. In der Mathematik sieht es genau umgekehrt aus. Hier geben die Knaben öfter als die Mädchen an, von sich selbst überzeugt zu sein. Sie sind auch motivierter, mathematische Inhalte zu lernen. Zusammenfassend lässt sich festhalten, dass im Lesen und der Mathematik deutliche fach- oder bereichsbezogene motivationale Unterschiede zwischen Mädchen und Knaben bestehen (vgl. ebd., p. 149).

Betrachtet man die oben aufgeführten Befunde, lässt sich vermuten, dass demzufolge nicht die Zugehörigkeit zum weiblichen Geschlecht an sich auch automatisch zu besseren Leseleistungen führt, sondern das Vertrauen in die eigene Lesekompetenz und das Interesse am Lesen. Die Ergebnisse aus PISA 2000 belegen diese Vermutung. Es hat sich gezeigt, dass zwischen den Lesekompetenzen von Knaben und Mädchen keine Unterschiede mehr auszumachen sind, wenn das Selbstkonzept im Lesen und auch das Interesse am Lesen ähnlich gross sind. Der Leistungsrückstand der Knaben lässt sich also durch das geringe Selbstkonzept im und das geringe Interesse am Lesen erklären. Wären diese motivationalen Merkmale bei den Knaben durchschnittlich gleich

gross, wie bei den Mädchen, würden auch deren Leistungen vergleichbar gross ausfallen. Ein ähnlicher Effekt in der Mathematik konnte nicht nachgewiesen werden. Hier zeigte sich lediglich für das mathematische Selbstkonzept ein vergleichbarer Effekt. Das Interesse an Mathematik hingegen führt nicht automatisch zu einer Steigerung der Leistungen in Mathematik. (vgl. Stanat & Kunter 2001, p. 265; Malti 2002, p. 151). Malti (2002) vermutet, dass das Interesse das Vertrauen in die eigenen Mathematikkompetenzen mitbedingt. Auf diesem Wege habe das Interesse auch ihren Erklärungsanteil. Sie schreibt zusammenfassend dazu:

«Das bedeutet, dass das Selbstvertrauen in die eigene Lese- bzw. mathematische Kompetenz die entsprechenden Fachleistungen beeinflusst. In Bezug auf das fachspezifische Interesse zeigen die Ergebnisse, dass sich das Interesse im Bereich Lesen signifikant auf die Lesekompetenz auswirkt. Das Interesse an Mathematik wirkt sich hingegen nicht signifikant auf die mathematische Kompetenz aus. Vielmehr bedingt das Interesse an Mathematik das Selbstvertrauen in die eigene mathematische Kompetenz. Je höher dieses ist, desto höher sind auch die mathematischen Leistungen» (Malti 2002, p. 152).

Die festgestellten Unterschiede bei den individuellen Lernvoraussetzungen von Mädchen und Knaben beeinflussen den Lernerfolg und langfristig auch die Berufschancen. Im Hinblick auf eine realistische Umsetzung des Ziels der Chancengleichheit ist die Entwicklung und Umsetzung von Massnahmen auf verschiedenen Ebenen anzustreben (vgl. ebd., p. 137).

Ein auffallendes Ergebnis aus PISA 2000 ist, dass sowohl im Lesen als auch in der Mathematik in den tiefen Kompetenzniveaus besonders viele Knaben zu finden sind. Die schwächeren Leistungen der 15-jährigen Knaben stellen insgesamt eine ernste bildungspolitische Herausforderung dar, mit der sich die Schule in Zukunft auseinandersetzen muss. Im Gegensatz dazu setzen sich die Mädchen mit ihren Leistungen immer mehr an die Spitze.

Es zeigt sich aber, dass in einigen Ländern der Unterschied zwischen den Geschlechtern sehr viel grösser ist als in anderen. Leistungsdifferenzen zwischen den Geschlechtern können also ausgeglichen werden. Zudem verzeichnet ein Teil der Länder mit den grössten geschlechtsspezifischen Unterschieden hohe Gesamtergebnisse. In Finnland z.B. schneiden die Knaben bei der Lesekompetenz keineswegs schlecht ab – in keinem anderen Land erreichen Knaben bessere Resultate (vgl. Meyer & Zahner 2002, p. 34f.).

5 Proximale Bedingungen schulischer Wirksamkeit

Wenn wir uns die in den vorausgehenden drei Kapiteln diskutierten Wirkungsstudien vor Augen führen, dann fallen zwei Dinge auf: *Erstens* sind es vom Charakter her Studien, die auf einem Input-Output-Modell beruhen, auch wenn dies explizit nicht immer so angegeben wird. «Input-Output-Modell» heisst, dass die *Prozessebene* von Schule und Unterricht nicht oder nur marginal beachtet wird. Die Blackbox Schule bleibt m.a.W. was sie ist, nämlich dunkel und unaufgeklärt. Die Tatsache, dass sich internationale Schulleistungsvergleiche im Wesentlichen auf die Erhebung von Eingangs- und Ausgangs-Variablen beschränken, bedeutet, dass der Erklärungsabstand zwischen den Prädiktoren⁴⁸ (unabhängige bzw. Input-Variablen) und dem Kriterium (abhängige bzw. Output-Variablen) vergleichsweise gross ist. Der Einfluss der Unterrichtsebene bleibt weitgehend ausgeklammert, was eine deutliche Schwäche des Ansatzes bedeutet. Es wird zwar zunehmend versucht, Prozessvariablen mit zu berücksichtigen – vor allem in Zusatzstudien wie zum Beispiel der TIMSS-Video-Studie –, doch der Survey-Charakter der internationalen Schulleistungsvergleiche erlaubt nur bedingt, Prozessvariablen in die Erklärung schulischer Leistungen einzubeziehen.

Zweitens fällt auf, dass wir es mit *Querschnittstudien* zu tun haben, die letztlich keine Aussagen über Kausalprozesse zulassen. Wir haben das Problem der Kausalanalyse im 1. Kapitel angesprochen (vgl. Abschnitt 1.2.1). Die Aufklärung von Kausalprozessen setzt ein experimentelles oder Längsschnitt-Design voraus. Letztlich ist es im Falle von pädagogischen Prozessen keineswegs klar, was wir uns unter Kausalität überhaupt vorzustellen haben: Ist die Lehrkraft tatsächlich die Kausalursache des Lernens einer Schülerin oder eines Schülers? Hier liegt vermutlich die schärfste Kritik, die man an den internationalen Leistungsstudien anbringen muss: Sie messen den Output des Bildungssystems, aber sagen wenig dazu, wie der Output *erzeugt* wird.

Allein schon diese zwei Merkmale internationaler Schulleistungsvergleiche reichen aus, um das Ungenügen der bisher diskutierten Analysen schulischer Wirksamkeit zu belegen. Es bedarf weiterer Ansätze, welche die Defizite der internationalen Effektivitätsstudien ausgleichen. Wir wollen in diesem Kapitel einen Einblick in solche alternativen Ansätze geben und beginnen mit einer kurzen Einleitung, in der wir einen Überblick über verschiedene Forschungsansätze liefern (vgl. Abschnitt 5.1). Es folgt eine Auseinandersetzung mit der Forschungsliteratur zu «guten» (effektiven) Schulen (vgl. Abschnitt 5.2). Danach befassen wir uns mit dem dominierenden Paradigma der Unterrichtsforschung, mit dem Prozess-Produkt-Ansatz (vgl. Abschnitt 5.3). Als nächstes geben wir einen Einblick in die experimentelle Unterrichtsforschung (vgl. Abschnitt 5.4). Anschliessend stellen wir die Technik der Metaanalyse vor und referieren Ergebnisse ausgewählter Metaanalysen zur schulischen Wirksamkeit (vgl. Abschnitt 5.5). Des Weiteren folgt eine Diskussion einiger Modelle schulischer und unterrichtlicher Wirksamkeit (vgl. Abschnitt 5.6) und schliesslich halten wir kritisch Rückblick auf die verschiedenen Forschungsansätze (vgl. Abschnitt 5.7).

⁴⁸ Der *Prädiktor* ist in einem statistischen Modell (z. B. in einem Regressionsmodell) diejenige Variable, aus deren Werten die Ausprägung einer anderen Variable (*Kriterium*) erklärt oder vorhergesagt werden soll. Andere häufig gebrauchte Bezeichnungen für *Prädiktor* sind unabhängige Variable (UV), Input-Variable, Kovariate oder erklärende Variable. Andere Bezeichnungen für das Kriterium sind abhängige Variable (AV), Output-Variable oder zu erklärende Variable. Ein Beispielmodell wäre, dass vom Prädiktor Intelligenz (UV) das Kriterium Schulleistungen (AV) vorhergesagt werden soll.

5.1 Einleitung

Unter Einschluss der in den Kapiteln 2 bis 4 diskutierten Ansätze lassen sich insgesamt folgende Formen von Wirksamkeitsstudien im Bereich Schule und Unterricht unterscheiden:

1. Querschnittlich angelegte Input-Output-Studien (Survey-Forschung)
2. Studien zu «effective schools» (Fallstudien; Outlier-Studien⁴⁹)
3. Nicht-interventionistische Unterrichtsforschung (Prozess-Produkt-Paradigma)
4. Interventionsstudien (experimentelle und quasi-experimentelle Forschung; Trainingsstudien; Evaluationsforschung)
5. Metaanalysen und Mehrebenenmodelle⁵⁰

Wie schon dargelegt (v.a. im Abschnitt 1.3), führen Input-Output-Analysen zu grossen Distanzen bei der Rückführung der erfassten Wirkungen (Effekte) auf die Ursachen (Bedingungen). Es fehlen Daten, die den *Unterrichtsprozess* in das Kausalgefüge einbeziehen lassen. Stattdessen stehen ökonomische und organisatorische Aspekte von Schulsystemen und Schulen im Vordergrund. Es ist daher sinnvoll, zwischen proximalen und distalen Bedingungen schulischer Wirksamkeit zu unterscheiden. *Proximal* sind Merkmale, die *unmittelbar* wirksam sind («Nahursachen»), wie Merkmale der Schülerinnen und Schüler (Intelligenz, Vorkenntnisse, motivationale Faktoren wie Interessen oder Ängstlichkeit), Zusammensetzung der Schulklasse (Grösse, Heterogenität, Herkunft etc.), Unterrichtsqualität (didaktische und methodische Merkmale des Unterrichts; Unterrichtsstil und Führungsverhalten der Lehrkraft etc.), elterliche Unterstützung (durch Hilfe bei den Hausaufgaben, intellektuelle Stimulation, Erwartungen der Eltern etc.). *Distal* sind demgegenüber Merkmale, die *mittelbar* wirksam sind («Fernursachen»), wie der Schultyp, die Organisationsform und die Qualität der Schule, familiäre Einflüsse im weiteren Sinne (Nationalität, sozialer Status, Bildungsniveau der Eltern etc.), Peer-Einflüsse, Medienkonsum etc.

Wo genau die Grenze zwischen proximal und distal zu ziehen ist, lässt sich nicht a priori entscheiden. Letztlich ist es eine Frage der Perspektive bzw. der Modelle oder Theorien, an denen man sich orientiert, wie diese beiden Einflussfaktoren unterschieden werden. Ein Beispiel für eine mögliche Unterscheidung gibt das folgende Modell von Andreas Helmke (vgl. Abbildung 5.1). Die proximalen Faktoren sind grau hervorgehoben.

⁴⁹ Von «Outliern» (engl. Ausreisser) spricht man in der Statistik, wenn ein Messwert oder Befund nicht in eine erwartete Messreihe passt oder allgemein nicht den Erwartungen entspricht. Outlierstudien untersuchen «Extremgruppen» (z.B. die besten vs. die schlechtesten Klassen), d.h. zuerst werden die Extremgruppen aufgrund eines vorgängig festgelegten Kriteriums identifiziert (z.B. unterste und oberste 5%) und danach hinsichtlich interessierender Merkmale miteinander verglichen.

⁵⁰ Im Falle von Metaanalysen und Mehrebenenmodellen handelt es sich allerdings nicht um spezifische Forschungstypen, sondern um Verfahren bzw. Familien von Verfahren zur Integration mehrerer Studien (Metaanalyse) bzw. Darstellungen der komplexen Beziehungen zwischen den Determinanten schulischer Wirksamkeit (Mehrebenen-Modelle), wobei im letzteren Fall auch komplexe Verfahren der Datenanalyse gemeint sein können.

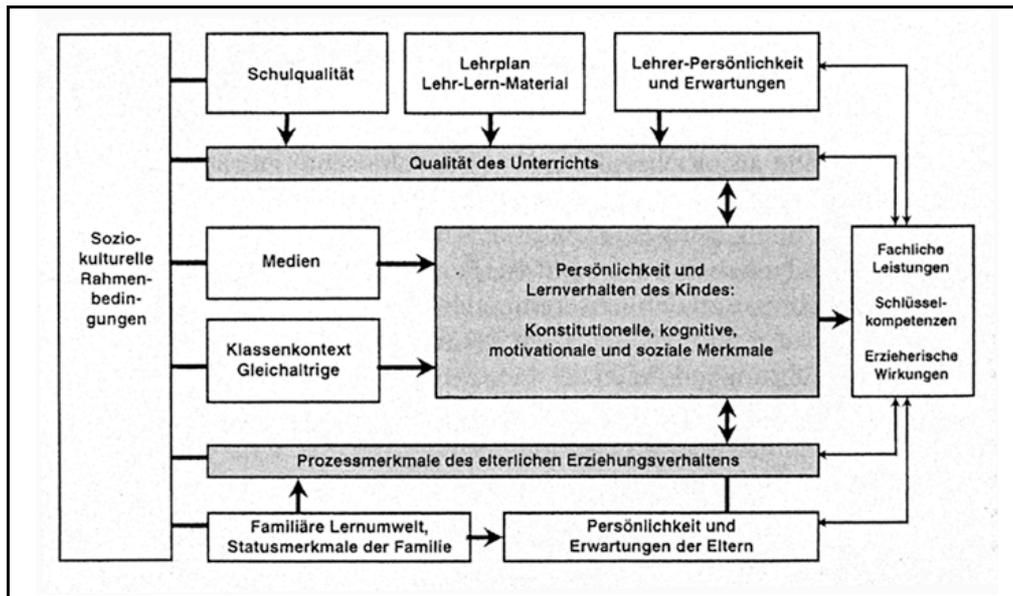


Abbildung 5.1: Makro-Modell der Einflussfaktoren schulischer Leistungen
(aus: Helmke 2000, p. 148)

Während sich die internationalen Wirkungsstudien vorzugsweise mit distalen Bedingungen schulischer Effektivität befassen, wollen wir uns im Folgenden Untersuchungen zuwenden, die eher proximale Faktoren thematisieren. Im Einzelnen befassen wir uns mit den oben unter 2. bis 5. genannten Ansätzen.

5.2 «Gute Schulen» (effective schools)

Die ersten Studien grösseren Ausmasses zur schulischen Wirksamkeit wurden in den späten 1960er und anfangs der 1970er Jahre in den USA durchgeführt. Sie kamen zum Ergebnis, dass Schulen wenig bewirken bzw. kaum ein Unterschied zwischen ihnen ausgemacht werden kann. Dieses pessimistische Urteil basierte im Wesentlichen auf Arbeiten des Soziologen James Coleman (1966) und des Psychologen Christopher Jencks (1972). Dazu gesellte sich ein Aufsatz von Arthur Jensen (1969, dt. 1973), der bezweifelte, dass Intelligenz und schulische Leistungsfähigkeit durch Umweltfaktoren wesentlich beeinflusst werden.

Der *Coleman*-Report gilt gemeinhin als der eigentliche Startpunkt der Schulwirkungsforschung. Coleman und seine Mitarbeiter kamen zum Schluss, dass die Schule lediglich rund 10 % der Varianz der Schülerleistungen zu erklären vermag. Anders formuliert: Alles, was schulische Faktoren (inklusive didaktische und andere Variablen der Unterrichtsebene) zu bewirken vermögen, ist eine Reduktion der Streuung der schulischen Leistungen um 10%. Der Rest der Varianz ist durch ausserschulische Faktoren (inklusive individuellen Bedingungen auf Seiten der Schülerinnen und Schüler) bedingt.

Zu ähnlichen Ergebnissen kam *Christopher Jencks*. Er und seine Mitarbeiter behaupteten eine Varianzaufklärung schulischer Leistungen durch schulische Variablen von gerade einmal rund 3%, während sie den Einfluss des familialen Hintergrundes auf ca. 50% schätzten. Im Einzelnen lassen sich die Analysen von Jencks et al. wie folgt zusammenfassen (vgl. Scheerens & Bosker 1997, p. 141f.):

- Die Schulen tragen wenig dazu bei, den Graben zwischen reichen und armen, begabten und weniger begabten Schülerinnen und Schülern zu überbrücken.
- Die Qualität des Unterrichts hat kaum Effekte auf die nachschulische Karriere von Schülerinnen und Schülern, insbesondere auf deren Einkommen.
- Schulische Leistungen werden zum grössten Teil bestimmt durch einen einzigen Faktor, und dies sind die familiären Bedingungen eines Schülers oder einer Schülerin. Alle anderen Faktoren sind von sekundärer Bedeutung, wenn nicht sogar irrelevant.
- Es gibt wenig Hinweise dahingehend, dass Schulreformen oder Programme der kompensatorischen Erziehung die kognitiven Ungleichheiten zwischen Schülerinnen und Schülern ausgleichen könnten.
- All dies in Rechnung stellend, kann ökonomische Gleichheit kaum über schulische Massnahmen erzielt werden, jedoch über eine (politisch begründete) Umverteilung der Einkommen.

Der Pädagoge *Arthur Jensen* stützte diese soziologisch und psychologisch ausgerichteten Analysen insofern, als er behauptete, dass rund 80% der Varianz in den Intelligenztestleistungen auf *genetische* Faktoren (Vererbung) zurückzuführen seien. Zusätzlich postulierte er eine sogenannte *Schwellenhypothese*, wonach die Intelligenz lediglich bis zu einer bestimmten Ausprägung durch Umweltfaktoren beeinflusst werden könne. Darüber hinaus, d.h. bei einer durchschnittlich anregungsreichen Umweltsituation, seien ausschliesslich genetische Faktoren für Variationen im Bereich der Intelligenz verantwortlich. Das bedeutet im Konkreten, dass bei Umweltbedingungen, wie sie in den homogenisierten Schulsystemen westlicher Gesellschaften vorhanden sind, die Schule selber keinen Einfluss auf die Ausprägung der Intelligenz nehmen kann.

Die Arbeiten von Coleman, Jencks und Jensen haben – aus unterschiedlichen Gründen – eine breite Diskussion ausgelöst. In deren Folge sind, vor allem im Bereich der Wirkungsstudien, detaillierte Untersuchungen von *Einzelschulen* vorgenommen worden, die nachweisen konnten, dass sehr wohl bestimmte schulische Merkmale dafür verantwortlich gemacht werden können, dass Schülerinnen und Schüler unterschiedlichen Schulerfolg haben. Es wurden Studien publiziert, die schon im Titel versuchten, Gegensteuer zu geben: 1979 erschien von W. B. Brookover, C. Beady, P. Flood, J. Schweitzer und J. Wisenbaker ein Buch mit dem Titel: «School Social Systems and Student Achievement. *Schools Make a Difference*»; 1988 erschien von Peter Mortimore, Pamela Sammons, Louise Stoll, David Lewis und Russell Ecob: «*School Matters. The Junior Years*».⁵¹ Erwähnenswert ist eine weitere Studie, die wesentlich zur Neubeurteilung der Bedeutung von Schule und Unterricht beigetragen hat: Michael Rutter, Barbara Maughan, Peter Mortimore und Janet Ouston: «Fifteen Thousand Hours. Secondary Schools and Their Effects on Children» (1979, dt. 1980). In dieser Studie ist vom «Ethos» der Schule die Rede, womit globale Faktoren gemeint sind, die heute unter den Begriff «Schulklima», «Unterrichtsklima» oder «Schulkultur» zusammengefasst werden.

Diese Art von Forschung hat im deutschen Sprachraum eine Bewegung ausgelöst, die unter dem Begriff «Schulqualität» und «gute Schulen» bekannt geworden ist. Im Englischen ist zumeist von «effective schools» die Rede, was nochmals zeigt, dass der Qualitätsdiskurs, importiert in den deutschen Sprachraum, andere Assoziationen weckt als die Diskussion in England und den USA beinhaltet. «Gute» Schulen sind *wirksame* Schulen – womit wir zum Thema der schulischen Wirksamkeit zurückgekehrt sind.

Insofern die Einzelschule im Vordergrund steht, sind in der Forschungstradition zu «effective schools» häufig *Fallstudien* durchgeführt worden – das gilt schon für die Rutter-Studie, die zwölf Sekundarschulen im Inner Circle von London untersucht hat. Fallstudien sind – im Vergleich zur

⁵¹ Auf das letztere Buch werden wir im Abschnitt 5.2.2 ausführlicher eingehen.

Survey-Forschung – in methodischer Hinsicht sehr aufwendig, vor allem was die Zeit anbelangt, die benötigt wird, um die Daten zu erheben. Deshalb werden in den internationalen Schulleistungsvergleichen solche Daten kaum je erhoben. Wie inzwischen bereits mehrfach betont, beschränken sich die internationalen Studien zur schulischen Wirksamkeit im Wesentlichen auf den Input- und den Output-Bereich. Diesbezüglich sind am ehesten Survey-Erhebungen möglich, während die Erfassung von Prozessvariablen aufwendig ist und vorzugsweise Unterrichtsbeobachtungen voraussetzt.

Allerdings zeigen die internationalen Wirksamkeitsstudien durchaus, dass es sinnvoll sein kann, die Einzelschule in den Forschungsfokus zu nehmen. Denn ein wiederkehrendes Ergebnis dieser vergleichenden Studien ist, dass nicht nur das Leistungsniveau im Vergleich verschiedener *Länder* unterschiedlich ist und auch nicht nur innerhalb eines Landes *regionale* Unterschiede bestehen, sondern auch Schülerinnen und Schüler, die *denselben Schultyp* besuchen, heterogene Leistungen aufweisen. Wenn man diese Heterogenität nicht einfach den genetischen Bedingungen zuschlagen will, dann muss man folgern, dass in ein und demselben Schultyp unterschiedliche Leistungen erbracht werden, was zumindest die Frage aufwirft, ob diese Unterschiede nicht durch die Qualität der Einzelschule oder des Unterrichts, den eine Lehrkraft erteilt, bedingt sein könnten.

Dieser Frage ist vor allem im Rahmen von Studien zur Schulqualität nachgegangen worden, die nach Kriterien «effektiver» bzw. «guter» Schulen suchen. Die Literatur zu diesem Thema ist mittlerweile umfangreich. Wir wollen uns im Folgenden kurz mit einer einflussreichen englischen Studie im Sekundarschulbereich beschäftigen (vgl. Rutter, Maughan, Mortimore & Ouston 1980), um uns dann etwas eingehender mit einer ebenfalls englischen Studie im Primarschulbereich auseinanderzusetzen (vgl. Mortimore, Sammons, Stoll, Lewis & Ecob 1988). Anschliessend folgt eine Zusammenfassung einer viel zitierten Reviewarbeit von Purkey und Smith (1991). Schliesslich versuchen wir eine abschliessende Interpretation der Forschung über «gute Schulen».

5.2.1 Eine Studie im Sekundarschulbereich

Die Studie von Rutter, Maughan, Mortimore und Ouston (1980) umfasst zwölf Schulen des Sekundarniveaus in der Innenstadt von London. Es handelt sich – und das ist wichtig – um *dieselbe* Schulform. Es sind nicht Schulen mit unterschiedlichem Niveau – wie zum Beispiel Gymnasium und Sekundarschule – miteinander verglichen worden, sondern Schulen *desselben Typs*. Zur Datenerhebung sind verschiedene Verfahren eingesetzt worden, zum Beispiel sind Dokumente der einzelnen Schulen analysiert worden (vor allem Dokumente über Noten und Prüfungsergebnisse der Schülerinnen und Schüler), es sind auch Beobachtungen in den Klassen durchgeführt worden, und es wurden mittels Fragebogen Daten erhoben, sowohl bei den Lehrkräften wie auch bei den Schülerinnen und Schülern.

Die Auswertung der Daten wurde so angelegt, dass man verschiedene Merkmale der Schulen als unabhängige Variablen (UV) behandelte. Auf der anderen Seite sind fünf Gruppen von Variablen als abhängige bzw. »Effektvariablen« (AV) bezeichnet worden. Diese umfassten das 1. Verhalten der Schülerinnen und Schüler in der Schule, die Anwesenheit im Unterricht, Prüfungsleistungen, die Beschäftigungssituation nach der Schulentlassung und die Schülerdelinquenz⁵². Die Daten wurden sehr sorgfältig erhoben, was allein schon die Tatsache bezeugt, dass die Erhebungsphase zweieinhalb Jahre gedauert hat.

Wir gehen nicht im Detail auf die Ergebnisse der Rutter-Studie (1980) ein, sondern beschränken uns auf eine Zusammenstellung der wichtigsten Resultate (vgl. Rutter et al. 1980, p. 209ff.):

⁵² Mit Delinquenz ist allgemein die Tendenz, rechtliche Grenzen zu überschreiten und eventuell sogar straffällig zu werden, gemeint.

- In Bezug auf Verhalten und Lernerfolg der Schülerinnen und Schüler bestanden zwischen den untersuchten Londoner Sekundarschulen deutliche Unterschiede. Dies gilt sowohl für das Verhalten der Schüler während der Schulzeit als auch für die Regelmässigkeit der Teilnahme am Unterricht, für den Prozentsatz der Schülerinnen und Schüler, die nach Ableistung der Pflichtschuljahre einen weiterführenden Abschluss anstrebten, für die Prüfungsleistungen und für die Delinquenzrate.
- Angemessenes Verhalten und gute Leistungen waren (unabhängig von der Zusammensetzung der Schülerschaft) in einigen Schulen eher zu erwarten als in anderen.
- Schulen, deren Schüler während der Schulzeit ein überdurchschnittlich positives Verhalten zeigten, boten in der Regel auch in Bezug auf Prüfungsleistungen und Delinquenz ein eher günstiges Bild. Abgesehen von einigen Ausnahmen lässt sich feststellen, dass die verschiedenen Erfolgskriterien eng miteinander zusammenhingen.
- Die schulspezifischen Ergebnisunterschiede waren *nicht* auf strukturelle Bedingungen - Gesamt-Schülerzahl, Alter der Gebäude, räumliche Kapazität, Art der Trägerschaft, Unterschiede im organisatorischen Bereich usw. - zurückzuführen.
- Die leistungsmässige Zusammensetzung der Schülerschaft spielte für die Ergebnisunterschiede zwischen den Schulen eine wichtige Rolle: Die Prüfungsleistungen fielen tendenziell besser aus an Schulen, die sich auf einen soliden Stamm von Schülerinnen und Schülern mit mindestens durchschnittlicher Leistungsfähigkeit stützen konnten; Schulen, in denen der Anteil der leistungsschwächsten Schüler stark überwog, verzeichneten höhere Delinquenzraten.
- Das Ausmass der Leistungsorientierung, das Lehrerverhalten im Unterricht, die Verteilung von Belohnungen und Strafen, günstige Lernbedingungen oder die Bereitschaft zur Übertragung von Verantwortlichkeit an Schülerinnen und Schüler liessen sich hingegen sämtlich in signifikanter Weise mit den schulspezifischen Ergebnisunterschieden in Verbindung bringen.
- Der *kumulative Effekt* der verschiedenen Aspekte der Schulsituation scheint erheblich grösser gewesen zu sein als der Einfluss irgendeines einzelnen Faktors. Vermutlich entsteht aus dem Zusammenwirken der verschiedenen Situationselemente ein gewisses »*Ethos*«, eine Grundstruktur bestimmter Wertorientierungen, Einstellungen und Verhaltensmuster, die für die Schule insgesamt charakteristisch wird.

Das Gesamtbild der Befunde lässt den Autoren zufolge mit hoher Wahrscheinlichkeit vermuten, dass den Zusammenhängen zwischen Schulsituation und Effektvariablen zum Teil ein kausaler Prozess zugrunde liegt. Mit anderen Worten: Durch ihre Erfahrungen mit der Schule, insbesondere mit der Schule als einer sozialen Institution, werden Kinder in ihren Verhaltensweisen und Einstellungen nachhaltig beeinflusst (vgl. Rutter et al. 1980, p. 209ff.).

5.2.2 Eine Studie im Primarschulbereich

Eine breit angelegte Studie zur Schulqualität wurde von Peter Mortimore (der bereits bei der Rutter-Studie mitgearbeitet hatte), Pamela Sammons, Louise Stoll, David Lewis und Russel Ecob (1988) ebenfalls im Inner Circle von London durchgeführt. Untersucht wurden Primarschulen (junior-only-schools), wobei in einigen Fällen auch Schulen mit Vorschulklassen (junior and infant schools) miteinbezogen wurden. Die Studie umfasste insgesamt 50 Schulen mit total rund 2'000 Schülerinnen und Schülern. Die Schulen wurden während drei bis vier Jahren begleitet, was es nicht nur erlaubte, eine Vielzahl von Daten zu erheben, sondern auch eine *längsschnittliche*

Analyse vorzunehmen. Die Hauptfrage der Studie war: «Are some schools more effective than others?» (Mortimore et al. 1988, p. 1).

Wir beschränken uns im Folgenden auf die wichtigsten Ergebnisse. Die Autorinnen und Autoren der Studie unterscheiden in Bedingungen, die nur schwer beeinflussbar sind («givens» genannt), und solche, die durch das Handeln der Schulleitung und der Lehrkräfte beeinflussbar sind.

Schwer beeinflussbare Faktoren («givens»), die mit höherer Effektivität von Schulen in Beziehung stehen sind das Vorhandensein einer «Basisstufe», kleinere bis mittlere Schulgrößen sowie Klassengrößen kleiner als 24 Schülerinnen und Schüler, eine gute Infrastruktur und materielle Ausstattung sowie ein gut unterhaltenes Schulgebäude, eine ansprechende Umgebung und Stabilität der Lehrerschaft (vgl. Mortimore et al. 1988, p. 250). Den Autoren folgend stehen diese Faktoren mit der Wirksamkeit einer Schule durchaus in positivem Zusammenhang, sie garantieren sie aber noch nicht. Vielmehr, so die Autoren weiter, sind es diejenigen Faktoren, welche von der Schulleitung und den Lehrpersonen kontrolliert bzw. beeinflusst werden können, welche ausschlaggebend sind. Aufgrund ihrer Analysen identifizieren die Autoren schliesslich 12 Schlüsselfaktoren, welche von Lehrpersonen und Schulleitungen beeinflusst werden können und welche mit der Wirksamkeit von Schule und Unterricht in Zusammenhang stehen (vgl. ebd., p. 250ff.).

Ohne an dieser Stelle einzeln auf die 12 Faktoren einzugehen, seien die Empfehlungen, welche die Autoren zuhanden von Lehrpersonen formulieren, kurz zusammengefasst (vgl. ebd., p. 268ff.):

1. *Strukturierte Lektionen*: Gemeint ist ein Unterricht, der weder rein lehrerzentriert noch rein schülerzentriert ist. »In the structured sessions that we observed, teachers sought to provide a considerable degree of freedom for pupils, but did so within a well-defined framework« (ebd., p. 268).
2. *Intellektuell herausfordernder Unterricht*: Die Lektionen sind stimulierend und intellektuell herausfordernd. Hohe Anforderungen vermitteln hohe Erwartungen, die auch darin bestehen, dass Lehrkräfte ihre Schülerinnen und Schüler zur unabhängigen Kontrolle ihrer Arbeit ermutigen (vgl. ebd., p. 269).
3. *Arbeitsförderliche Unterrichtsbedingungen*: Der Unterricht ist arbeitsorientiert, d.h. frei von Störungen und sachfremden Aktivitäten (vgl. ebd., p. 287).
4. *Limitierter thematischer Fokus des Unterrichts*: Gemeint ist, dass die Schülerinnen und Schüler mehr oder weniger an einem Thema ausgerichtet sind, und nicht zur gleichen Zeit verschiedene Schülergruppen verschiedene Stoffe bearbeiten (vgl. ebd., p. 287).
5. *Ein Maximum an Interaktion mit den Schülerinnen und Schülern*: Lehrkräfte sollen so viel wie möglich mit ihren Schülerinnen und Schülern kommunizieren. »We were surprised at how, in a few classrooms, the amount of communication was very limited with teachers busying themselves with administration and pupils being expected to work by themselves. This tended not to be the pattern in the most effective schools where much teacher-pupil communication took place« (ebd., p. 288 – unsere Hervorhebungen). Ebenso von Bedeutung ist, wieviel Zeit eine Lehrperson für die Kommunikation mit der ganzen Klasse aufwendet. Eine Balance zwischen Kontakten mit einzelnen Schülerinnen und Schülern und der Klasse im Ganzen erweist sich als zuträglicher als die absolute Betonung der Kommunikation mit einzelnen Schülerinnen und Schülern oder mit Schülergruppen.
6. *Häufige Diagnose der Lernleistungen sowie deren Dokumentation*: In der kontinuierlichen individuellen Beurteilung der Schülerinnen und Schüler sowie dem Führen von Schülerberichten liegt eine bedeutsame Methode eines guten Unterrichts. »Our data show quite clearly the

need for the careful recording of pupil progress» (ebd., p. 270).

7. *Einbezug der Eltern*: Das Engagement der Eltern beinhaltet Mithilfe im Unterricht und Kontakte mit der Lehrkraft. Ebenso erweist sich die Unterstützung zuhause, etwa bei der Erledigung der Hausaufgaben oder die Vermittlung intellektueller Anregungen, als sehr förderlich.

Als wichtige Faktoren nennen Mortimore et al. (1988) des Weiteren die Kontrolle der eigenen Erwartungen und ein Merkmal, das man »Feedback-Kultur« nennen könnte. Erwartungen gegenüber bestimmten Schülerinnen und Schülern oder Schülergruppen (wie Mädchen versus Jungen, Ausländer versus Einheimische etc.) können die Qualität des Unterrichts stark negativ bestimmen, vor allem, wenn es sich um Erwartungen geringer Leistungsfähigkeit handelt. Lehrkräfte sollten sich ihrer allfälligen Erwartungen bewusst sein. Was die »Feedback-Kultur« anbelangt, so stellen Mortimore et al. (1988) mit einem gewissen Erstaunen fest, dass das Ausmass an Schülerkritik, das sie in den 50 untersuchten Schulen beobachtet haben, das Ausmass an Lob deutlich übersteigt. Dazu kommt, dass Lob geringer wird, wenn die Schülerinnen und Schüler älter werden.

5.2.3 Eine zusammenfassende Darstellung der Forschungsliteratur

Die in den vorausgehenden Abschnitten diskutierten Einzelstudien zeigen eine Reihe von Entsprechungen. Trotzdem decken sie sich in ihren Ergebnissen nicht in jedem Fall. Es ist daher wichtig, möglichst viele Studien beizuziehen, um zu einem verlässlichen Eindruck von der Wirksamkeit «guter» bzw. «schlechter» Schulen zu kommen. Eine viel beachtete Studie, die sich diese Aufgabe der Integration von Teilstudien vorgenommen hat, stammt von Stewart Purkey und Marshall Smith (1991). Ihrer Arbeit liegen acht Outlier-Studien, acht Fallstudien, sechs Evaluationsstudien sowie zwei weitere Studien, die alle im Wesentlichen in den 1970er Jahren durchgeführt wurden, zu Grunde, darunter auch diejenige von Rutter et al. (1980).

Purkey und Smith gehen davon aus, dass eine Schule als ein Gefüge von verschiedenen Ebenen, die ineinander verschachtelt sind, verstanden werden kann. Die beiden bedeutsamsten Ebenen in Bezug auf die Wirksamkeit einer Schule sind diejenigen des *Unterrichts* und der *Einzelstufe*. Obwohl die pädagogischen Prozesse in der Klasse strukturell das unterste Niveau schulischer Wirksamkeit darstellen, bestimmt nichtsdestoweniger die angrenzende Ebene, nämlich die organisatorische bzw. institutionelle Ebene der Schule, das Feld, in dem sich der Unterricht vollzieht. «Die Qualität des Erziehungsprozesses auf Klassenebene wird positiv oder negativ von der Qualität der Handlungen auf der nächst höheren Ebene beeinflusst» (Purkey & Smith 1991, p. 15). Dementsprechend unterscheiden Purkey und Smith Kriterien, die den *Rahmen* einer Schule bestimmen, und andere, die die *Prozesse* betreffen, die sich innerhalb des schulischen Rahmens abspielen. Der Rahmen besteht, wie angedeutet, aus organisatorischen und institutionellen Variablen, «[...] die durch administrative oder bürokratische Mittel hergestellt und gesteuert werden können» (ebd., p. 36).

Purkey und Smith (1991) bezeichnen die folgenden *Strukturmerkmale* als die wichtigsten Elemente einer wirksamen schulischen Organisation:

- eine weitgehend autonome Schulleitung
- eine unterrichtsbezogene Führung der Schule
- Stabilität des Kollegiums

- ein Kerncurriculum, d.h. ein zielorientiertes Unterrichtsangebot (statt einer Vielzahl von Wahlfächern)⁵³
- eine bedarfsorientierte, curriculumbezogene Weiterbildung des Kollegiums
- Unterstützung und Mitarbeit der Eltern
- schulöffentliche Anerkennung fachunterrichtlicher Leistungserfolge (öffentliche Ehrung von guten Leistungen)⁵⁴
- wirksam genutzte Unterrichtszeit («time on task»)
- Unterstützung durch die regionale Schulbehörde

Wie erwähnt bilden diese Strukturvariablen den *Rahmen*, innerhalb dessen die Prozessvariablen für die Qualität der Schule garantieren. Purkey und Smith nennen die Prozessvariablen die «treibende Kraft einer Schule» (Purkey & Smith 1991, p. 40), die zu erhöhten Schülerleistungen führt. Sie erwähnen vier *Prozessvariablen*, die sie aufgrund ihrer Sichtung der Forschungsliteratur für wesentlich erachten:

- gemeinsame Planungen und kollegiale Beziehungen
- Zusammengehörigkeitsgefühl (bei Lehrern wie bei Schülern)
- klare Ziele und hohe Erwartungen, die von allen geteilt werden (inkl. kontinuierliche Kontrolle der Zielerreichung)
- Ordnung und Disziplin

In methodischer Hinsicht liegt eine Beschränkung der Integration von Einzelstudien, wie sie von Purkey und Smith vorgenommen wurde, darin, dass die Vielzahl an Untersuchungen lediglich *qualitativ* integriert wird. Es handelt sich um sogenannte «narrative reviews», die – weil sie keine *quantitativen* Vergleiche vornehmen – über die *Gewichtigkeit* der einzelnen determinierenden Faktoren keine präzisen Aussagen machen können, d.h. sie können nichts darüber aussagen, wie bedeutsam jeder einzelne Faktor ist oder in welchem Mass er das Kriterium Schulleistungen zu erklären vermag.⁵⁵ Das ist anders bei den sogenannten *Metaanalysen*, denen wir uns im Abschnitt 5.5 zuwenden werden.

5.2.4 Ausblick

Die Besonderheit der Merkmale, die in der Literatur zu »guten Schulen« erörtert werden, liegt in ihrem globalen Charakter. Das heisst, es handelt sich im Allgemeinen um Merkmale *ganzheitlicher* Art, die einen grossen Teil der Bereiche und Handlungsebenen einer Schule betreffen. Eine handliche Zusammenstellung dieser Merkmale ist nicht leicht vorzunehmen. Trotzdem hat sich in

⁵³ Dieses Kriterium nimmt Bezug auf die spezifische Situation in den USA, wo an den meisten Schulen eine relativ grosse Freiheit in der Fächerwahl besteht.

⁵⁴ Auch dies dürfte auf die USA bezogen sein: Im öffentlichen Leben der USA ist es üblich, Exzellenz auszuzeichnen, d.h. die Entfaltung von Bestleistungen durch öffentliche Anerkennung und Zeremonien (wie Preisverleihung und ähnlichem) zu unterstützen.

⁵⁵ Nach Purkey und Smith (1991, p. 40) haben *alle* vier Prozessvariablen einen Einfluss auf die Leistungen der Schülerinnen und Schüler; wir wissen aber nicht, *wie gross* dieser Einfluss ist. Vielleicht ist es so, dass «Ordnung und Disziplin» oder «klare Ziele und hohe Erwartungen» in einem engeren Zusammenhang mit den Leistungen stehen und für diese bedeutsamer sind als die «kollegialen Beziehungen». Solche «gewichtete» Aussagen sind aber aufgrund der qualitativen Vorgehensweise der Studie gerade nicht möglich.

jüngster Zeit ein gewisser Konsens über die Merkmale »guter Schulen« herausgebildet. Die Merkmale entsprechen weitgehend den beiden Listen, die wir bei Mortimore et al. (1988) sowie Purkey und Smith (1991) angetroffen haben. Danach gehören zu einer »guten« Schule (vgl. Aurin 1991; Fend 1998; Scheerens & Bosker 1997; Weinert 2001a):

- eine differenzierte Schulkultur im Sinne gemeinsamer Ziele, Wertorientierungen und sozialer Verhaltensregeln (insbesondere gegenseitiger Respekt und Vertrauen)
- ein schulspezifisches Profil von curricularen und extracurricularen Ansprüchen und Aktivitäten (ein pädagogisches Leitbild)
- eine effiziente Führung (Schulleitung), die auch für Fragen des Unterrichts und der Unterrichtspraxis Verantwortung trägt («educational leadership»)
- ein positives soziales Klima und ein Verantwortungsstil, der dem Schulleiter, dem Lehrerkollegium, den Lehrerinnen und Lehrern, den Eltern sowie den Schülerinnen und Schülern bestimmte soziale Rollen mit individuellen Spielräumen zuweist
- eine sichere, ordentliche und ästhetische Schul- und Lernumwelt
- strukturierte Lernangebote (Handlungssicherheit für die curricularen und aussercurricularen Aktivitäten der Schülerinnen und Schüler)
- hohe Erwartungen: Gute Schulen erwarten gute Leistungen, sie unterstellen, dass alle etwas lernen können; Fördern und Fordern stellen Prinzipien dar, die sich gegenseitig ergänzen.
- eine ausreichende schulische Lerneffektivität und deren regelmässige interne und externe Evaluation (inklusive optimale Zeitnutzung und häufige Erfassung der Lernfortschritte der Schülerinnen und Schüler)
- eine prospektive Schulentwicklungsplanung, welche die pädagogischen Qualitätsansprüche definiert und zu Verbesserungen auf der Ebene der Schule, der Schulklassen, des Unterrichts und zusätzlicher Lerngelegenheiten in Schulen führt (inkl. schulinterne Weiterbildung der Lehrkräfte)
- gute, effektive und vertrauensvolle Beziehungen zwischen Schule und Eltern

5.3 Prozess-Produkt-Forschung

Der »Gute-Schule«-Ansatz ist stark an organisatorischen Merkmalen von Schule orientiert, d.h. er thematisiert Bedingungen von Schülerleistungen, die nach wie vor eher *distalen* Charakter haben. Das ändert sich mit dem Forschungsparadigma, das wir nun erörtern wollen. Hier geht es eindeutig um *proximale* Bedingungen schulischer Leistungen.

5.3.1 Die Persönlichkeit als Wirkkraft

Als eine solche proximale Bedingung gilt bei Praktikerinnen und Praktikern häufig die *Lehrerpersönlichkeit*. Allerdings ist nicht immer klar, was mit der »Persönlichkeit« des Lehrers bzw. der Lehrerin gemeint ist. Oft stehen statische und unbeeinflussbare Merkmale im Vordergrund, wie das Temperament, der »Charakter« der Lehrperson oder die Vorbildwirkung, die von ihr ausgeht.

Wie das alltägliche Selbstverständnis der Lehrerinnen und Lehrer misst auch die geisteswissenschaftliche Pädagogik der Lehrerpersönlichkeit unter den pädagogischen Wirkkräften grosse Bedeutung zu. Oft steht sie gar an erster Stelle des pädagogischen Wirkungsgefüges. So ist für Kerschensteiner ein Lehrer, »[...] der eine volle Persönlichkeit geworden ist, [...] das wertvollste Bildungsgut, das [...] die grösste Wirkung auf die Schüler auszuüben vermag« (Kerschensteiner

1949, p. 110). Zwar könne man das Lehrersein bis zu einem gewissen Grade lernen, sobald der Lehrer aber auch *Erzieher* sein soll, sind «[...] ganz andere Eigenschaften nötig als die blossе Fähigkeit, ein oder mehrere Wissensgebiete beherrschen zu lernen und nach vorgeschriebenem oder selbst erfundenem Methodenschema erfolgreich an die Schüler zu übermitteln» (Kerschensteiner 1969, p. 12). Wer Gemeinschaftsgesinnung oder sozialen Geist im anderen lebendig machen wolle, der müsse diese selber besitzen, d.h. «[...] selbst *ein sozialer Geist sein*» (ebd.). Was wir aber *sind*, das haben wir nicht lernen können, das sind wir *schlechthin geworden*, weil der Kern dazu von vornherein in unserem individuellen Wesen lag [...] und ein gütiges Geschick ihn entwickeln half» (ebd.).

Was Kerschensteiner damit behauptet, dass nämlich pädagogisches Handeln auf einer Basis beruht, die sich nicht *lernen* lässt, wird auch von Spranger behauptet. Zwar meint sein Begriff des «geborenen Erziehers» nicht eine schicksalhafte Begabung, aber doch ein »inneres Getriebensein zur Menschenbildung« (Spranger 1969, p. 284). Gemeint ist der »genuine Erzieher«, d.h. der Pädagoge «[...] von so echter Art, 'als ob' er für das Erziehertum geradezu geboren wäre» (ebd., p. 285). Demnach ist nicht anlagemässig festgelegt, wer zum geborenen Erzieher wird. «Man wird [...] nicht als Erzieher geboren, sondern man muss sich langsam zu der Forderung, die diese Geistesart stellt, emporverwandeln und emporveredeln» (ebd., p. 334). Trotzdem rechnet Spranger mit Eigenschaften des Lehrers, «[...] die nicht auf Einsicht beruhen und daher *weder lehrbar noch lernbar sind*» (ebd., p. 300 – unsere Hervorhebung).

Die hohen Erwartungen an die Persönlichkeit des Lehrers erklären sich aus der Annahme, pädagogische Wirksamkeit ergebe sich in erster Linie aus der *Vorbildlichkeit* des Erziehers. Viel mehr als durch sein Handeln wirke der Lehrer durch seine Person. Bestimmend sei die Atmosphäre, in der Erziehung stattfindet, und diese sei in erster Linie eine Funktion der Lehrerpersönlichkeit. Deshalb gehöre zum Erziehen wesentlich, «[...] dass man an sich selbst ernsthaft gearbeitet habe» (Spranger 1969, p. 323).

Seit den 30er Jahren des 20. Jahrhunderts gibt es eine Reihe von empirischen Studien über die Persönlichkeitsstruktur von Lehrkräften, die zu einer Relativierung der hohen Bedeutung der Lehrerpersönlichkeit geführt haben. Dies vor allem deshalb, weil die Studien kaum nennenswerte Einflüsse der Persönlichkeitsstruktur des Lehrers auf das Lernverhalten und die Leistungen von Schülerinnen und Schülern ausmachen konnten. Die durchaus grosse Zahl an Studien lässt sich auf die mehr oder weniger triviale Aussage reduzieren, «[...] dass gute Lehrer eher freundlich, heiter, sympathisch oder tugendhaft als grausam, depressiv, unsympathisch oder amoralisch seien bzw. sein sollten» (Pause 1973, p. 85). Auch wenn dies als eine etwas überspitzte Formulierung erscheinen mag, ist anzuerkennen, dass die Suche nach der idealen Lehrerpersönlichkeit bisher ergebnislos verlaufen ist und es an der Zeit wäre, sich zu fragen, weshalb dem so ist. Unsere Vermutung ist, dass es – bis auf einige Ausnahmen⁵⁶ – keiner *spezifischen* Persönlichkeitsmerkmale oder Charaktereigenschaften bedarf, um eine gute Lehrerin oder ein guter Lehrer zu sein. Entscheidender als (statische) Persönlichkeitsmerkmale sind (professionelle) Kompetenzen, die sich im Verhalten ausdrücken und nach pädagogischen und didaktischen Qualitätsmerkmalen beurteilt werden können. Der Anspruch auf Professionalität der pädagogischen Berufe, wie er seit langem erhoben wird, rückt die Persönlichkeit in den Hintergrund, denn Professionalität zeigt sich nicht daran, wie einer *ist*, sondern wie er *zu handeln weiss*. Insofern könnte man sagen, dass auch die Persönlichkeit eher eine *distale* Bedingung der unterrichtlichen Wirksamkeit einer Lehrperson ist, während die *proximalen* Bedingungen in ihrem *Verhalten* liegen.

⁵⁶ Dazu gehören gewisse basale Charaktereigenschaften wie Gewissenhaftigkeit, Gerechtigkeit, Empathiefähigkeit, Verständnis, Sinn für Humor, Toleranz, Ordentlichkeit und Geduld (vgl. Gerstenmaier 1975).

5.3.2 Das Prozess-Produkt-Paradigma

Die Persönlichkeit ist ein weitgehend *statisches* Merkmal. Sie lässt wenig Spielraum zur Veränderung des Lehrerhandelns zu, abgesehen davon, dass das Ansinnen, die Persönlichkeit einer angehenden Lehrperson zu *verändern*, ein ethisch kaum vertretbares Ziel der Lehrerinnen- und Lehrerbildung wäre. Die Unterrichtsforschung hat sich daher den *Prozessmerkmalen* des Unterrichts zugewandt, wozu insbesondere das *Verhalten* bzw. *Handeln* der Lehrkräfte gehört, d.h. dasjenige, was Lehrerinnen und Lehrer in der Unterrichtssituation konkret *tun*.

Dementsprechend ist Unterrichtsforschung im wesentlichen Prozess-Produkt-Forschung. Man untersucht Prozessvariablen des Unterrichts und setzt sie in Beziehung zu Produktvariablen. Allgemein gesagt, beanspruchen Prozess-Produkt-Studien «[...] to define the relationships between what teachers do in the classroom (the processes of teaching) and what happens to their students (the products of learning)» (Anderson, Evertson & Brophy 1979, p. 193).

5.3.2.1. Erweiterung des Prozess-Produkt-Paradigmas

Um einen etwas genaueren Einblick in diese Forschungstradition zu gewinnen, wollen wir die Position von Nathaniel Gage etwas näher anschauen. Gage bestimmt die Wissenschaft als Analyse des Zusammenhangs zwischen zwei bis vier oder mehr gleichzeitig wirksamen Variablen, wobei er betont, dass der praktische Nutzen wissenschaftlichen Wissens nachzulassen beginnt, sobald mehr als vier Variablen in die Analyse einbezogen werden, da in der Praxis kaum mehr als vier Variablen gleichzeitig im Auge behalten werden können. Die Variablen, die im Falle einer Unterrichtstheorie von Relevanz sind, sind Variablen des Lehrens auf der einen Seite und solche des Lernens auf der anderen Seite. Unterrichtsforschung ist folglich gemäss Gage die Aufdeckung des Zusammenhangs zwischen Variablen des Lehrens (Lehrerverhalten) und Variablen des Lernens (Schülerverhalten), wobei im letzteren Fall vor allem die Schülerleistungen im Vordergrund stehen. Von daher ergibt sich die Bezeichnung »Prozess-Produkt-Forschung«: Untersucht wird, welche *Prozessmerkmale* des Unterrichts (insbesondere Lehrerverhalten, aber auch andere Merkmale des Unterrichtsgeschehens) welche Produkte (insbesondere Schülerleistungen) erzielen. In den Worten von Gage, suchen wir beim Prozess-Produkt-Paradigma

«[...] nach 'Prozessvariablen' (Lehrerverhaltensweisen und -eigenschaften in Gestalt von Unterrichtsstilen, -methoden, -modellen oder -strategien), mit denen 'Produktvariablen' (d.h. Ergebnisse des Unterrichts in Gestalt von Schülerleistungen und -einstellungen) vorausgesagt oder wünschenswerterweise kausal erklärt werden können» (Gage 1979, p. 71).

Ein wesentlicher Vorzug des Prozess-Produkt-Paradigmas gegenüber der Forschung zur Persönlichkeit von Lehrkräften (vgl. Abschnitt 5.3.1) liegt darin, dass die Forschungsmethoden auf einer weit konkreteren, verhaltensnahen Ebene ansetzen können. Während Persönlichkeitsmerkmale wie «Heiterkeit» oder «Gelassenheit» die Verwendung hoch-inferenter Messinstrumente (wie Persönlichkeitsfragebogen oder Ratingverfahren) voraussetzen, sind Verhaltensweisen wie «einen Schüler loben» oder «einer Schülerin eine Frage stellen» niedrig-inferent messbar, d.h. sie lassen sich mehr oder weniger direkt *beobachten*.

Das Prozess-Produkt-Paradigma ist relativ stabil und auch gegenüber Kritik ziemlich resistent. Zwar kann man berechtigterweise einwenden, es sei zu einfach, da es unter anderem den Beitrag ausklammert, den die Schülerinnen und Schüler zur Verwirklichung ihrer Leistungen erbringen (z.B. Aufmerksamkeit, Lernstile, Gedächtnisleistungen, Interessen etc.), doch lässt sich dieses Manko leicht beheben, indem man Schülerverhaltensweisen als zusätzliche bedingende Faktoren

neben die Prozessvariablen stellt.⁵⁷ Scheinbar ungenügend repräsentiert im Prozess-Produkt-Paradigma sind auch sogenannte ökologische Bedingungen des Unterrichts, wie z.B. die Lehrmittel, die Form der Prüfungen, situative Merkmale des Klassenzimmers, räumliche Bedingungen, der Einfluss anderer Schülerinnen und Schüler oder auch unartikulierte Erwartungen der Lehrperson. Aber auch bei diesen potentiellen Einflussgrößen lassen sich durchaus entsprechende Ergänzungen des Paradigmas vornehmen, indem die Prozessvariablen durch Kontextvariablen ergänzt werden. Die Kontextvariablen erscheinen dann als Hinweise (cues) für die Schülerinnen und Schüler, die ihr Verhalten danach orientieren. Auf diese Weise kommt Gage zu einem *vierstufigen Modell*, das man als *Prozess-Hinweisreiz-Prozess-Produkt-Paradigma* der Unterrichtsforschung bezeichnen könnte (vgl. Abbildung 5.2).

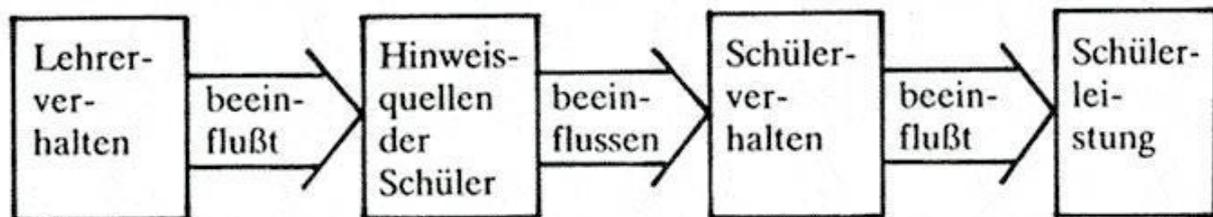


Abbildung 5.2: Erweitertes Prozess-Produkt-Paradigma (aus: Gage 1979, p. 76)

Im Unterschied zu den Kontext- und Inputvariablen, sind die Prozessvariablen im Allgemeinen unmittelbar beeinflussbar. Sie bilden dementsprechend *proximale* Bedingungen des Unterrichts bzw. der schulischen Leistung. Auch Kontext und Input können beeinflussbar sein, aber kaum durch das Lehrerhandeln. Prozessvariablen beziehen sich demnach auf Massnahmen, die Lehrkräfte treffen können, um die Wirksamkeit von Schule und Unterricht zu verbessern.

5.3.3 Ergebnisse der Prozess-Produkt-Forschung

Was sind die wichtigsten Ergebnisse der Prozess-Produkt-Forschung? Wir gliedern unsere Diskussion in die folgenden drei Hauptpunkte:

1. *Bedingungen auf Schülerseite*: kognitive Bedingungen (Intelligenz und fachliches Vorwissen), motivationale und emotionale Bedingungen sowie deren relative Bedeutung
2. *Bedingungen auf Seiten des Unterrichts*: quantitative Bedingungen («time on task»), qualitative Bedingungen sowie Kontextmerkmale und Lehrererwartungen
3. *Bedingungen auf Seiten der Familie*

Am Ende der jeweiligen Ausführungen werden wir die genannten Bedingungen zusammenfassend darstellen und beurteilen.

⁵⁷ Borich und Klinzing (1987) sprechen in diesem Fall von einem Prozess-Prozess- (wenn nur das Lehrer- und das Schülerverhalten erfasst wird) bzw. von einem Prozess-Prozess-Produkt-Paradigma (wenn darüber hinaus auch die Schülerleistungen erfasst werden).

5.3.3.1 Bedingungen auf Schülerseite: kognitive Bedingungen

Zunächst einmal besteht grosse Einmütigkeit dahingehend, dass die *intellektuellen Voraussetzungen* und das *fachliche Vorwissen* starke Determinanten der schulischen Leistung sind. Nimmt man die allgemeine Intelligenz (wie sie von IQ-Tests gemessen wird) als Prädiktor, dann liegen die Korrelationswerte für das Kriterium Schulleistung durchwegs zwischen $r = .5$ und $r = .6$. Dabei ist allerdings zu beachten, dass die Schwankungsbreite (Varianz bzw. Streuung) der Korrelationswerte, abhängig vom verwendeten Intelligenztest und vom erfassten Leistungskriterium, durchaus beträchtlich ist; sie liegt zwischen $r = .1$ und $r = .9$.⁵⁸

Umstritten ist die *relative* Bedeutung dieser beiden Faktoren Intelligenz und Vorwissen. Andreas Helmke ist der Ansicht, dass die Intelligenz – im Sinne einer einheitlichen Fähigkeit, wie sie durch Intelligenztests gemessen und in einem IQ ausgedrückt wird – als Voraussetzung für die Schulleistung im Vergleich zum bereichsspezifischen Vorwissen überschätzt wird (vgl. Helmke 1992, p. 7f.). Dafür spricht die folgende Tabelle 5.1:

Tabelle 5.1: Einfache Korrelationen und Partialkorrelationen zwischen Mathematikleistungen im Verlaufe der Grundschule und Intelligenzleistungen (aus: Helmke & Weinert 1997, p. 107)

Zum Vergleich	Arithmetik	Textaufgaben
Einfache Korrelation zwischen Mathematikleistung (2. Klasse) und Mathematikleistung (4. Klasse)	$r = .57$	$r = .55$
Einfache Korrelation zwischen Intelligenzleistung (1. Klasse) und Mathematikleistung (4. Klasse)	$r = .26$	$r = .47$
Partialkorrelation zwischen Mathematikleistung (2. Klasse) und Mathematikleistung (4. Klasse) bei Auspartialisierung der Intelligenzleistung (1. Klasse)	$r = .53$	$r = .42$
Partialkorrelation zwischen Intelligenzleistung (1. Klasse) und Mathematikleistung (4. Klasse) bei Auspartialisierung der Intelligenzleistung in der 2. Klasse	$r = .05$	$r = .29$

Die in der Tabelle dargestellten Ergebnisse zeigen die Zusammenhänge zwischen Intelligenztestwerten und verschiedenen mathematischen Leistungen (Arithmetik und Textaufgaben) innerhalb eines mehrjährigen Zeitintervalls. Interessant sind vor allem die *Partialkorrelationen*. Bei Partialkorrelationen werden drei oder mehr Variablen in die (statistische) Analyse einbezogen, wobei jeweils eine Variable konstant gehalten bzw. «ausgeschaltet» («auspartialisiert») wird.⁵⁹ Konkret (in Bezug auf Tabelle 5.1): Bei der *einfachen Korrelation* zwischen den beiden Variablen «Mathematikleistung in der 2. Klasse» und «Mathematikleistung in der 4. Klasse» (vgl. die erste Ergebniszeile der Tabelle) bleibt unberücksichtigt, ob allenfalls die Intelligenz als dritte Variable die Höhe der Korrelation beeinflusst. Wird die Intelligenz «auspartialisiert», bleibt nur mehr die

⁵⁸ Korrelation ist ein statistischer Fachbegriff für *Zusammenhang* oder «Assoziation» zweier Variablen. Masse für die Stärke der Korrelation werden meist als Korrelationskoeffizienten r bezeichnet. Korrelationskoeffizienten können Werte zwischen minimal -1 und maximal +1 annehmen, wobei -1 einen perfekten negativen und +1 einen perfekten positiven Zusammenhang bezeichnet (vgl. Internetseite ILMES, Kapitel «Korrelationen»; http://www.lrz-muenchen.de/~wlm/ein_voll.htm). Eine Korrelation von 0 würde bedeuten, dass gar kein Zusammenhang zwischen zwei Variablen besteht.

⁵⁹ Die Auspartialisierung von Variablen ist eine statistische Kontrolltechnik, die als Ersatz für die in der nicht-experimentellen Forschung nicht mögliche Kontrolle der Untersuchungsbedingungen fungiert (zum Experiment als Forschungsdesign vgl. Abschnitt 5.4).

Korrelation zwischen den Mathematikleistungen zurück, unabhängig vom Niveau der Intelligenz eines Schülers oder einer Schülerin (vgl. die dritte Ergebniszeile von Tabelle 5.1). Tatsächlich zeigt Tabelle 5.1, dass die Beziehung zwischen mathematischen Vorkenntnissen in Arithmetik (2. Klasse) und Arithmetikleistungen in der 4. Klasse bei statistischer Kontrolle der Intelligenz fast unverändert bleibt ($r = .57$ bzw. $r = .53$). Partialisiert man demgegenüber die Vorkenntnisse (Intelligenzleistung in der 2. Klasse) aus, so reduziert sich der Zusammenhang zwischen Intelligenz und Mathematikleistungen in Arithmetik praktisch auf Null ($r = .26$ bzw. $r = .05$). Bei den Textaufgaben (dabei geht es um die Fähigkeit, mathematische Probleme zu lösen) sind die Werte weniger ausgeprägt, gehen aber in dieselbe Richtung.

Das Ergebnis hilft, die einfachen Korrelationen zwischen Intelligenz und Schulleistung nicht falsch zu interpretieren. Die Intelligenz stellt zwar eine wichtige individuelle Ressource dar, um schulischen Stoff zu lernen, von kaum zu überschätzender Bedeutung für den Lernerfolg ist aber das bereits vorhandene Wissen in einem Stoffgebiet. Zusätzliche Belege für die Bedeutung des Stoffwissens gibt die sogenannte *Novizen-Experten-Forschung*. Vergleicht man Menschen mit ähnlicher Intelligenz, aber mit einem unterschiedlichen Wissensstand bei Lernaufgaben aus einem bestimmten Stoffgebiet, dann übertreffen diejenigen, die über das bessere Wissen verfügen (Experten) die Novizen in fast allen Belangen. Wobei sich die Überlegenheit auf jene Bereiche beschränkt, für die das Spezialwissen tatsächlich verfügbar ist. «Verallgemeinert man diesen Befund und berücksichtigt zusätzlich viele ähnliche Ergebnisse, so gelangt man zu der theoretisch wichtigen Schlussfolgerung, dass die kognitive Entwicklung nicht nur eine Bedingung, sondern stets auch eine *Folge*, ja sogar ein Ziel des schulischen Lernens darstellt [...]» (Helmke & Weinert 1997, p. 109 – unsere Hervorhebung).

5.3.3.2 Bedingungen auf Schülerseite: motivationale und emotionale Bedingungen

Auch über die Bedeutung der *motivationalen* und *emotionalen* Variablen, welche die Schulleistung beeinflussen, besteht ein relativ grosser Konsens. Dazu gehört das Wissen um die leistungsmindernde Wirkung von Prüfungsangst und die leistungsförderlichen Effekte eines positiven Leistungsselbstbildes sowie die positive Korrelation zwischen dem *Interesse* an einem Fach und den Leistungen in dem betreffenden Fach. Uguroglu und Walberg (1979) kommen aufgrund einer Metaanalyse⁶⁰ von 40 Einzelstudien zum Ergebnis, dass im Mittel lediglich 11.4 % der Varianz der Schulleistungen auf Motivationsfaktoren zurückgehen. In den vierzig Studien werden insgesamt 232 Korrelationskoeffizienten berichtet, die sich über die 12 Klassenstufen hinweg, wie in der folgenden Tabelle 5.2 dargestellt verteilen:

⁶⁰ Was unter einer Metaanalyse zu verstehen ist, erläutern wir im Abschnitt 5.5.

Tabelle 5.2: Korrelationen zwischen Schulleistungen und motivationalen Faktoren nach Klassenstufe (aus: Uguroglu & Walberg 1979, p. 380)

Grade	Mean Correlation	Standard Deviation	Number of Correlations
1	.07	.19	9
2	.25	.14	4
3	.25	.08	15
4	.29	.07	6
5	.35	.10	27
6	.36	.12	37
7	.31	.14	40
8	.38	.09	12
9	.41	.18	26
10	.29	.08	13
11	.36	.11	19
12	.44	.18	24

Legende: Grade = Klassenstufe; Mean Correlation = (über die 40 Studien) gemittelte Korrelation; Standard Deviation = Standardabweichung (Mass für die Streuung der Werte); Number of Correlations = Anzahl der in die Auswertung einbezogenen Einzelkorrelationen.

Deutlich ist, dass die Aufklärung der Leistungsvarianz durch motivationale Faktoren auf den unteren Klassenstufen gegen Null geht, während sie auf den oberen Stufen (12. Klasse) fast 20 % ausmacht.⁶¹ Der Einfluss motivationaler Faktoren auf die Schulleistung scheint in den ersten Schuljahren nicht besonders gross zu sein. Tatsächlich kommen die Kinder im Allgemeinen mit hohen Leistungserwartungen und grosser Lernfreude in die Schule. Dass die Bedeutung der Motivation mit den Jahren grösser wird, hat möglicherweise damit zu tun, dass die objektiv erbrachten Leistungen bzw. deren Benotung bei vielen Schülerinnen und Schülern der anfänglich unbefangenen Lernfreude einen Dämpfer aufsetzt. Es braucht mehr «Durchhaltewille», wenn man zur Kenntnis nehmen muss, dass man in der Schule nicht so gut ist, wie man anfänglich geglaubt hat. Die Motivation oder das Interesse können dann zur entscheidenden Grösse werden, die über Erfolg oder Misserfolg in einem Fach befinden.

5.3.3.3 Bedingungen auf Schülerseite: relative Bedeutung der kognitiven und motivationalen Bedingungen

Wenn nach der *relativen* Bedeutung der *kognitiven* und *motivationalen* Faktoren gefragt wird, dann schwindet die zuvor erwähnte Einmütigkeit. Auf der einen Seite wird den kognitiven Faktoren das grössere Gewicht eingeräumt, auf der anderen Seite finden sich Positionen, die beide Anteile zumindest als gleichgewichtig beurteilen. So konnte Helmke (1992) feststellen, dass bei der Vorhersage von Mathematikleistungen im Primarschulbereich der prognostische Wert motivationaler Variablen kaum geringer ist als derjenige der kognitiven Variablen. Allerdings ist in Rechnung zu stellen, dass die neuere Motivationspsychologie einen starken *kognitiven Einschlag* aufweist, so dass nicht immer klar ist, ob unter motivationalen Einflüssen letztlich nicht doch wieder kognitive Variablen thematisiert werden. Helmke selber räumt ein, «[...] dass es erhebliche Beiträge an konfundierter Varianz gibt, also Varianzanteile, die nicht ausschliesslich eine[r] der beiden Prädiktorgruppen zuzuordnen sind» (Helmke 1997, p. 214). Er schreibt weiter, dass 25 - 30 Prozent

⁶¹ Der Anteil der erklärten Varianz in Prozent (R^2) berechnet sich aus dem Quadrat des Korrelationskoeffizienten r . Im Falle der Tabelle 5.2 wäre der Anteil der erklärten Varianz bei der 12. Klasse folglich $R^2 = .44^2 \Rightarrow 19.3\%$. Dies bedeutet, dass 19.3% der Varianz in den Schulleistungen der Zwölfklässlerinnen und Zwölfklässler durch motivationale Faktoren erklärbar sind.

der Gesamtvarianz von schulischen Leistungen durch die Kommunalität (gemeinsame Varianz) von motivationalen und kognitiven Merkmalen aufgeklärt werde.

Nur nebenbei erwähnt, sind es diese beiden Faktoren – Begabung (Intelligenz) und Anstrengung, d.h. kognitive und motivationale Bedingungen –, die auch Lehrkräfte vorzugsweise beiziehen, um die Leistungen von Schülerinnen und Schülern zu erklären und zu bewerten.

5.3.3.4 Zusammenfassende Beurteilung der Bedingungen auf Schülerseite

Zusammengefasst ist ein wiederkehrendes Ergebnis verschiedener Studien zu den Bedingungen schulischer bzw. unterrichtlicher Wirksamkeit der grosse Einfluss von Schülermerkmalen. So schreibt zum Beispiel Gage:

«In jeder Untersuchung zur Schulleistung, die ich kenne, haben solche schulunabhängigen Variablen einen grösseren Anteil der Varianz der Schülerleistungen erklärt, als irgendeine auf die Person des Lehrers oder sein Verhalten bezogene Variable. Der typische Fall ist der, dass der IQ oder die Ergebnisse eines Kindes im prognostischen Test, gemessen vor der Einschulung, viel höher mit der Leistung am Ende einer Schulperiode korrelieren als die Variablen des Lehrerverhaltens. Dieses Ergebnis zeigt sich gleich, ob wir nach einer Viertelstunde Unterricht messen oder nach einem ganzen Schuljahr. Die Korrelationen zwischen schulunabhängigen Schülervariablen und Schülerleistungen können Werte bis zu .7 oder .9 annehmen. Manche Forscher haben in dieser Tatsache einen Anlass gesehen, an der Bedeutung des Lehrerverhaltens zu zweifeln. Zuwenig Varianz in der Schülerleistung bleibt zu erklären übrig [...]» (Gage 1979, p. 64).

Die Interpretation dieses Ergebnisses ist allerdings nicht trivial. Auf methodische Probleme des Ansatzes der «Varianzaufklärung» haben wir bereits im ersten Kapitel hingewiesen. Ein weiteres Problem liegt in der Tatsache, dass Schülervariablen nicht einfach als »letzte« Ursachen verstanden werden können, denn es handelt sich dabei – zumindest teilweise – auch um Ergebnisse (Wirkungen) früherer pädagogischer bzw. schulischer Einflüsse. Wird eine Schülerin in der 7. Klasse untersucht und werden ihre Leistungen in Beziehung zu Unterrichtsvariablen gesetzt, dann dürfte der *aktuelle* Lehrereinfluss tatsächlich vergleichsweise gering sein, was aber nicht ausschliesst, dass der *kumulative* Einfluss aller *früherer* Lehrkräfte zusammen nicht trotzdem bedeutsam ist. Faktisch wird deren Einfluss aber nicht erfasst, weil er aufgrund der Methodik von *Querschnittuntersuchungen* nicht erfasst werden kann. Gage drückt sich wie folgt aus: «Da die Wirkungen sich kumulativ addieren, fügt jeder 'neue' Lehrer dem Einfluss von den 'bisherigen' Lehrern etwas hinzu. Vom letzten Lehrer in der Kette können wir nicht erwarten, dass sein Einfluss der kumulativen Wirkung aller seiner Vorgänger nahe kommt. Es ist aber zwangsläufig immer dieser letzte Lehrer, dessen Wirkung wir in Korrelationsstudien zur Bedeutung des Lehrerverhaltens untersuchen. So gesehen, ist es gar nicht anders zu erwarten, als dass dieser letzte Lehrer einen scheinbar nur unbedeutenden Einfluss ausübt» (ebd., p. 65).

Eine weitere Schwierigkeit der Interpretation der Bedeutung individueller Schülervariablen für die Erklärung von Schulleistungen liegt darin, dass es oft nicht leicht ist, festzustellen, welche individuellen Determinanten der Schulleistung durch familiäre oder schulische Effekte überhaupt beeinflussbar sind. Das lässt sich am Beispiel des Merkmals Geschlecht gut illustrieren. Anders als im Falle der Intelligenz, die auch nach alltagspädagogischer Überzeugung wenigstens in einem gewissen Ausmass beeinflussbar ist, denken wir uns das Geschlecht als eine natürlicherweise gegebene und daher unbeeinflussbare Grösse. Bei der Diskussion der internationalen Schulleistungsstudien (vgl. Kapitel 2 bis 4) haben wir gesehen, wie stark das Geschlecht als bedingende Variable ins Gewicht fällt. Vor allem in der Schweiz sind die schulischen Leistungen der Mädchen in Mathematik und in den Naturwissenschaften (Physik, Chemie) geringer (wenn auch nicht in gravierendem Ausmass) als diejenigen der Knaben. Der Vergleich mit anderen Ländern, in denen die Geschlechterdifferenz entweder kleiner oder gar nicht vorhanden ist, lässt vermuten, dass es

sich nicht um *genetisch* bedingte Unterschiede handelt, sondern um Effekte, die pädagogisch beeinflussbar sind. Das heisst, dass das Geschlecht *als solches* zwar unbeeinflussbar sein mag, gewisse Attribute, die mit dem Geschlecht zusammengehen, jedoch sehr wohl beeinflussbar sind. Dass Frauen mathematisch unbegabt sind, scheint eher ein soziales Stereotyp zu sein als eine biologische Tatsache. Demnach muss es Möglichkeiten pädagogischer Art geben, die Geschlechterdifferenz bezüglich mathematischer und naturwissenschaftlicher Leistungen abzubauen und aufzuheben.

Das Beispiel zeigt im Übrigen nochmals, dass die internationalen Schulvergleichsstudien letztlich keine Kausalaussagen zulassen, sondern lediglich auf einer deskriptiven Ebene feststellen können, was der Fall ist. Dabei stellt die Geschlechterdifferenz ein immer wieder repliziertes Ergebnis der internationalen Schulleistungsvergleiche dar. Bereits in der ersten IEA-Studie über naturwissenschaftliche Schulleistungen («Six-Subject-Survey») wurde festgestellt:

«Whether the causes of this sex difference are innate, which seems unlikely, or whether they are the result of traditional practices in child rearing and formal education, it is impossible to say at this juncture. It is clear that there is a problem that deserves further attention if justice is to be done to girls in the field of science education and if countries are to make the most of their resources in scientific development» (Walker 1976, p. 99).

5.3.3.5 Bedingungen auf Seiten des Unterrichts: quantitative Bedingungen («time on task»)

Bezüglich der Merkmale des *Unterrichts*, welche die Leistungen von Schülerinnen und Schülern bedingen, ist in quantitative und qualitative Faktoren zu unterscheiden. In *quantitativer* Hinsicht ist die Wirksamkeit von Schule von der *Anzahl Lektionen*, während der ein Fach unterrichtet wird, abhängig. Wie trivial auch immer die Erkenntnis sein mag, sie hat sich in einer Fülle von Studien bestätigt: Die Zeit, die auf das Lernen eines Faches verwendet wird, bestimmt die Qualität des erworbenen Wissens. Umgekehrt formuliert: Je mehr Lektionen ausfallen und nicht kompensiert werden, desto schlechter die schulische Leistung. Das gilt auch innerhalb der verfügbaren Lektionenzahl: Je grösser die aktive Lernzeit («time on task»), desto besser die Lernleistung. Wiederum umgekehrt formuliert: Je mehr Zeit auf die Disziplinierung von Schülerinnen und Schülern verwendet wird bzw. werden muss, desto weniger Lernzeit steht zur Verfügung und desto schlechter ist die schulische Leistung. Auch das Thema *Hausaufgaben* gehört in diesen Kontext. Gezielt eingesetzte Hausaufgaben erweitern die Lernzeit und stellen dadurch eine förderliche Bedingung der schulischen Leistung dar.

5.3.3.6 Bedingungen auf Seiten des Unterrichts: qualitative Bedingungen

Was die *qualitativen* Merkmale des Unterrichts anbelangt, so erweisen sich die folgenden Kriterien als günstig für die Leistungsentwicklung in einer Klasse: hohe Leistungserwartungen, intensive individuelle Unterstützung und Hilfe, klar aufgebaute Unterricht, ausgeprägte Lehrstofforientierung, Toleranz von Langsamkeit und effizientes Klassenmanagement.

Eine etwas ausführlichere Zusammenstellung verschiedener Faktoren auf der Unterrichtsebene gibt Lowyck (zit. nach Scheerens & Bosker 1997, p. 148):

1. «*Clarity*: clear presentation adapted to suit the cognitive level of pupils.
2. *Flexibility*: varying teaching behavior and teaching aids, organizing different activities, etc.
3. *Enthusiasm*: expressed in verbal and non-verbal behavior of the teacher.
4. *Task-related and/or businesslike behavior*: directing the pupils to complete tasks, duties, exercises, etc., in a businesslike manner.
5. *Criticism*: much negative criticism has a negative effect on pupil achievement.

6. *Indirect activity*: taking up ideas, accepting pupils' feelings and stimulating self-activity.
7. *Providing the pupils with an opportunity to learn criterion material*, i.e. a clear correspondence between what is taught in class and what is tested in examinations and assessments.
8. Making use of *stimulating* comments: directing the thinking of pupils to the question, summarizing a discussion, indicating the beginning or end of a lesson, emphasizing certain features of the course material.
9. *Varying the level* of both cognitive questions and cognitive interaction.»

Verschiedene Studien zeigen, dass ein Kriterium effektiven Unterrichts in dem liegt, was in der englischsprachigen Literatur «highly structured learning» bzw. «direct teaching» genannt wird. Damit ist ein eher *lehrerzentrierter* Unterricht gemeint, in dem Sinne, dass die Lehrkraft ihrem Unterricht eine klare Struktur gibt. Dazu gehört auch, dass die Lektionen klare Ziele haben, die Lernfortschritte regelmässig überprüft werden und ein ausgiebiges korrekatives Feedback praktiziert wird. In den Worten von Michael Rutter :

«[...] a variety of well-conducted studies of teaching and of classrooms have shown not only that teaching 'matters' but also that the crucial components of effective teaching include a clear focus on academic goals, an appropriate degree of structure, an emphasis on active instruction, a task-focused approach, and high achievement expectations» (Rutter 1983, p. 20f.).

Diese Aufzählung günstiger Bedingungen wird von Andreas Helmke und Friedrich-Wilhelm Schrader in Bezug auf Ergebnisse einer Studie an Münchner Hauptschulen (in der Schweiz: Realschulen) bestätigt: «Die günstigste Leistungsentwicklung war in Klassen zu finden, deren Unterricht durch Klarheit, ausgeprägte Lehrstofforientierung, häufige individuelle fachliche Unterstützung, Effizienz der Klassenführung und ein hohes Anforderungsniveau charakterisiert ist» (Helmke & Schrader 2001, p. 245).

«Direct teaching» ist nicht gleichzusetzen mit Frontalunterricht. »Direct teaching« meint auch nicht zwingend, dass der Unterricht im Frage-Antwort-Stil, d.h. als »fragend-entwickelnder Unterricht« gestaltet wird. Walter Doyle, der sich auf die Sekundarstufe bezieht, nennt folgende Merkmale des «direct teaching»:

1. Teaching goals are clearly formulated.
2. The course material to be followed is carefully split into learning tasks and placed in sequence.
3. The teacher explains clearly what the pupils must learn.
4. The teacher regularly asks questions to gauge what progress pupils are making and whether they have understood.
5. Pupils have ample time to practice what has been taught, with much use being made of prompts and feedback.
6. Skills are taught until mastery of them is automatic.
7. The teacher regularly tests the pupils and calls on the pupils to be accountable for their work» (zit. nach Scheerens & Bosker 1997, p. 150).

Ein «direkter» Unterricht ist allerdings nicht *per se* effizienter. Er scheint besonders bei leistungsschwächeren Schülerinnen und Schülern wirksam zu sein, während leistungstärkere Schülerinnen und Schüler von einem eher «freien» bzw. «offenen» Unterrichtsstil profitieren. Um Franz Weinert zu zitieren:

«Je weniger es um den Erwerb von Wissen und Können geht, je stärker die praktische Anwendung des Gelernten, je mehr das Lernen des Lernens selbst und je mehr die Einübung von methodischen Schlüsselqualifikationen (z.B. kommunikative, muttersprachliche und fremdsprachliche Kompetenzen, Medienkompetenzen, Teamfähigkeiten) als Ziele des Unterrichts fungieren, umso geringer wird die Bedeutung des lehrergesteuerten, aber schülerzentrierten Unterrichts und umso wichtiger werden Formen der Projektarbeit, des Gruppenunterrichts, des selbstorganisierten Lernens, des offenen Klassenzimmers und der lehrlingsanalogen Praktika» (Weinert 2001b, p. 81).

Dieses Ergebnis ist deshalb von Bedeutung, weil es darauf hinweist, dass nicht jeder Unterrichtsstil und nicht jede Art von Unterricht auf alle Schülerinnen und Schüler gleich wirken.

5.3.3.7 Zusammenfassende Beurteilung der bisherigen Bedingungen

Es scheint, dass die vier Faktorenbündel, die wir bisher diskutiert haben (kognitive und motivationale Bedingungen auf Seiten der Schülerinnen und Schüler, Quantität und Qualität des Unterrichts) bereits einen grossen Teil der Bedingungsfaktoren schulischer Leistung abdecken. So kommen Haertel, Walberg und Weinstein zum Ergebnis:

«Lernen im Klassenzimmer ist eine multiplikative [...] Funktion von vier wesentlichen Faktoren – Fähigkeit und Motivation der Schüler sowie Qualität und Quantität des Unterrichts [...] Jeder dieser wesentlichen Faktoren scheint als solcher notwendig, aber nicht hinreichend für das Lernen im Klassenzimmer zu sein, d.h. alle vier Faktoren müssen wenigstens auf einem minimalen Niveau verfügbar sein, damit schulisches Lernen stattfinden kann. Es scheint auch, dass sich die wesentlichen Faktoren wechselseitig substituieren, kompensieren oder ausbalancieren können [...]» (Haertel, Walberg & Weinstein 1983, p. 75f. – eigene Übersetzung).

5.3.3.8 Bedingungen auf Seiten des Unterrichts: Kontextmerkmale und Lehrererwartungen

Trotzdem werden in der Literatur weitere Faktoren diskutiert, so etwa die *Kontextmerkmale* bzw. äusseren Merkmale *der Unterrichtssituation* wie Klassengrösse, Homogenität bzw. Heterogenität der Schülerschaft, Leistungsniveau der Klasse, Klassenklima u.ä. Schliesslich spielen trotz allem auch *Lehrervariablen* eine Rolle. Wichtig scheint vor allem, dass Lehrkräfte eine positive Haltung bewahren und auch unter widrigen Umständen an die Leistungsbereitschaft und -fähigkeit der Schülerinnen und Schüler glauben. Nichts ist schlimmer als eine pessimistische Einstellung. Wie sehr *Lehrererwartungen* das Lernverhalten und die Leistungen von Schülerinnen und Schülern beeinflussen können, zeigt der sogenannte Pygmalioneffekt.⁶²

5.3.3.9 Bedingungen auf Seiten der Familie

Auch in der Prozess-Produkt-Forschung kommt regelmässig zu Tage, dass der familiäre Hintergrund eine bedeutsame Determinante schulischer Leistungen ist. Darauf gehen wir an dieser Stelle nicht weiter ein, möchten aber am Beispiel der Familie nochmals betonen, wie wichtig die Unterscheidung in distale und proximale Einflüsse ist. Denn wenn von der Familie als Bedingung schulischer Leistungen die Rede ist, dann werden im Allgemeinen zwei Gruppen von Faktoren genannt. Auf der einen Seite ist von *strukturellen* Merkmalen der Familie die Rede, wie materielle und ökonomische Situation (Einkommen, Wohnverhältnisse), soziokulturelle Merkmale

⁶² Bekannt geworden ist der Effekt durch die wegweisende und gleichnamige Studie von Rosenthal und Jacobson (1968; dt. Übersetzung 1971) zu «Pygmalion im Unterricht». Allgemein formuliert geht es dabei um den Mechanismus der sich-selbst-erfüllenden Prophezeiungen («self fulfilling prophecies»). Gemeint ist, dass Erwartungen, die wir haben, wahr werden, da wir im Sinne dieser Erwartungen handeln.

(Bildungsinteresse, Kulturkonsum, Wertvorstellungen) und familienorganisatorische Merkmale («Vollständigkeit» der Familie, Zahl der Kinder, Position in der Geschwisterreihe). Auf der anderen Seite erscheinen *prozessuale* Merkmale wie der Kommunikationsstil in der Familie, Erziehungsziele und Erziehungspraktiken der Eltern, Unterstützung der Kinder durch die Eltern, Leistungserwartungen der Eltern, Freizeitgestaltung etc. Es dürfte leicht einzusehen sein, dass die strukturellen Faktoren das Lernverhalten eines Kindes nicht direkt bestimmen. Bedeutsamer sind vielmehr die prozessualen Faktoren, die unmittelbar Einfluss nehmen auf die Entfaltung von kognitiven Kompetenzen, Interessen und Einstellungen von Kindern.

5.3.3.10 Gesamtbeurteilung der genannten Bedingungen

Unbefriedigend an der Zusammenstellung von Einzelfaktoren, die die Effektivität von Unterricht bestimmen, ist nicht nur die insgesamt eher geringe Varianzaufklärung bei den Effekten, sondern auch die Unklarheit, wie denn die (vielen) Einzelfaktoren zusammengehören bzw. *zusammenwirken*. Wie gehen zum Beispiel die kognitiven Voraussetzungen mit den motivationalen zusammen und diese wiederum mit der Unterrichtsqualität und den Kontextbedingungen des Unterrichts? Etwas despektierlich könnte man von «Schrotflinten-Forschung» (*shotgun research*) sprechen: Die einzelnen Studien nehmen sich aus wie Schüsse mit der Schrotflinte ins wilde Geschehen des Unterrichts. Irgendetwas trifft man fast immer, wie relevant es aber unter einer Gesamtperspektive ist, vermag man nicht zu sagen. Was fehlt, sind *Gesamtmodelle* für die Erklärung schulischer Wirkungen.⁶³ Um hier einen Schritt weiter zu kommen, wären Theorien unabdingbar. Doch gute Unterrichtstheorien sind zurzeit Mangelware. Wir kommen daher nicht darum herum, festzuhalten, dass sowohl der Stand der Forschung als auch der Stand der Theoriebildung im Bereich der Wirksamkeit von Schule und Unterricht unbefriedigend sind. Zwar gibt es eine Vielzahl an empirischen Einzelstudien, aber keine Theorien, welche die Einzelbefunde zufrieden stellend integrieren liessen.

Dafür verantwortlich sind auch methodische Gründe. Wie schon einige Male betont, lassen sich auf der Basis von Survey- bzw. Querschnittstudien lediglich Korrelationen, d.h. Beziehungen zwischen Variablen feststellen, über die Kausalstruktur, die ihnen zu Grunde liegt, können keine verbindlichen Aussagen gemacht werden. Das ist nur im Rahmen von experimentellen oder Längsschnittstudien möglich. Weil dem so ist, wollen wir uns im Folgenden der experimentellen Unterrichtsforschung zuwenden bzw. nach dem Gewinn fragen, den die Nutzung des Experiments in der Pädagogik bietet.

5.4 Experimentelle Unterrichtsforschung

Was wir als Kritik an den Input-Output-Modellen, die den internationalen Leistungsvergleichen zugrunde liegen, formuliert haben, trifft zu einem grossen Teil auch auf die Prozess-Produkt-Forschung zu. Auch wenn hier ein weit konkreteres Niveau gewählt wird, ist auch die Prozess-Produkt-Forschung kaum in der Lage, *Kausalaussagen* zu machen. Die Daten verbleiben auf der Ebene von Korrelationen zwischen Variablen des Lehrerverhaltens auf der einen Seite (Prozess) und Variablen der Schülerleistung auf der anderen Seite (Produkt), ohne über die *Richtung* der Wirksamkeit verbindliche Aussagen machen zu können. Erfasst werden Phänomene, die *zusammen auftreten*, während Kenntnisse erwünscht wären, die den Praktiker und die Praktikerin darüber aufklären, welches Phänomen dem anderen als Ursache *vorausgeht*.

⁶³ Wir werden einige solcher Modelle im Abschnitt 5.6 etwas näher kennen lernen.

5.4.1 Der Begriff des Experiments

Als begriffliche Kriterien eines Experiments werden gelegentlich die «drei W's» genannt: Willkürlichkeit, Wiederholbarkeit und Variierbarkeit. «Willkürlichkeit» meint, dass die Bedingungen, unter denen ein experimenteller Faktor (wie im Beispiel von Lewin: der Führungsstil) wirken soll, aktiv und planmässig hergestellt werden (Manipulation der unabhängigen Variablen). Zur Willkürlichkeit gehört auch, dass allfällige störende Nebenbedingungen ausgeschaltet oder kontrolliert werden (im Beispiel von Lewin wurden das Alter und das Geschlecht der Kinder sowie die Gruppengrösse dadurch kontrolliert, dass sie konstant gehalten wurden; die Persönlichkeit der Gruppenleiter wurde kontrolliert, indem alle Gruppenleiter alle Stile praktizieren mussten). «Wiederholbarkeit» meint, dass ein Experiment unter genau spezifizierten Bedingungen durchgeführt wird, die es ermöglichen, das Experiment ein zweites Mal bzw. weitere Male durchzuführen (z.B. mit anderen 10jährigen Knaben und/oder anderen Gruppenleitern, um die Allgemeingültigkeit der Resultate zu überprüfen). «Variierbarkeit» meint, dass die Bedingungen, unter denen ein Effekt erzeugt wird, variabel sind (im Beispiel von Lewin: der Führungsstil liegt in drei Ausprägungen vor). Zur Variierbarkeit gehört aber auch, dass das Experiment unter anderen Bedingungen wiederholt werden kann, um systematisch die Auswirkungen veränderter Randbedingungen zu überprüfen (z.B. jüngere vs. ältere Kinder; Mädchen vs. Knaben; kleinere vs. grössere Gruppen; männliche vs. weibliche Gruppenleitende).

Wenn wir die drei W's zusammennehmen, dann können wir wie folgt definieren:

«Unter einem Experiment [...] versteht man [...] die planmässige Beobachtung eines Vorganges unter systematisch variierten, hinreichend kontrollierten und möglichst replizierbaren Bedingungen zum Zwecke der Erweiterung der Kenntnisse über die herrschenden Gesetzmässigkeiten oder der Überprüfung von Hypothesen» (Fischer 1977, p. 137).

Eine Hypothese meint einen Konditionalsatz der Art: Wenn (diese oder jene Bedingung gegeben ist), dann (stellt sich dieser oder jener Effekt ein).

- Fakultative Lektüre: Der Begriff des Experiments⁶⁴
- Fakultative Lektüre: Ein Experiment zum fragend-entwickelnden Unterricht

5.4.2 Grenzen der experimentellen Forschung in pädagogischen Situationen

Der experimentellen Forschung in der Pädagogik sind deutliche Grenzen gesetzt. Sie liegen dort, wo die experimentelle Forschungslogik die aktive Manipulation der Bedingungsfaktoren sowie die Kontrolle der Randbedingungen erfordern würde. Bei der Untersuchung von Schulklassen ist es jedoch kaum möglich, die Klassen willkürlich auszuwählen oder zusammenzusetzen. Ebenso wenig ist es möglich, die Lehrkräfte willkürlich auszuwählen oder den Klassen willkürlich zuzuteilen. Unterrichtsforschung muss im Normalfall mit «natürlicherweise» gegebenen Klassenverbänden arbeiten. Kinder, die in eine Schule eintreten, werden nach *administrativen* Kriterien zu Klassen zusammengefügt. Es wäre ethisch kaum vertretbar, wenn Erziehungswissenschaftler fordern würden, dass sie zu Versuchszwecken Klassen auflösen und nach streng experimentellen Kriterien neu zusammensetzen dürfen, v.a. wenn wir in Rechnung stellen, dass Schulklassen relativ lange zusammenbleiben, Experimente – nicht zuletzt wegen des hohen Kontrollaufwands – kaum über längere Zeit durchführbar sind. Des Weiteren wäre es ethisch nicht zu vertreten, die Kontextbedingungen, unter denen ein Unterrichtsexperiment stattfindet, so weit zu kontrollieren, dass deren

⁶⁴ Die Texte zur fakultativen Lektüre sind online in der Lernumgebung des Moduls zu finden.

Einfluss tatsächlich neutralisiert würde, denn bei diesen Kontextbedingungen handelt es sich um Faktoren, deren Standardisierung einen massiven Eingriff in das Privatleben der am Experiment beteiligten Schülerinnen und Schüler bedeuten würde (Gestaltung der Freizeit, Wahl der Freunde, familiäre Einflüsse, Medienkonsum etc.).

Bei pädagogischen Experimenten ist es auch selten möglich, die Einhaltung der experimentellen Bedingungen zu garantieren (vgl. Abschnitt 5.4.3). Im Beispiel von Anderson, Evertson und Brophy (1979) musste extra kontrolliert werden, inwiefern die Lehrkräfte der Experimentalgruppe die 22 Prinzipien tatsächlich umsetzten. Des Weiteren musste kontrolliert werden, ob nicht ein Teil oder sogar alle Lehrkräfte der Kontrollgruppe die 22 Prinzipien *ungewollterweise* ebenfalls umsetzten. Tatsächlich mussten Anderson, Evertson und Brophy (vgl. ebd.) feststellen, dass in Einzelfällen Lehrpersonen der Kontrollgruppe einige der Prinzipien intuitiverweise angewandt hatten, d.h. ohne dass sie dahingehend instruiert worden wären. Solche vorweg kaum kontrollierbaren Einflüsse können ein experimentelles Design beeinträchtigen oder sogar ganz ausser Kraft setzen.

Ein weiteres Problem der experimentellen Forschung, das aber nicht nur für die Pädagogik, sondern auch für andere Sozialwissenschaften relevant ist, liegt darin, dass der Variation der Bedingungsfaktoren Grenzen gesetzt sind. Theoretisch können faktorielle Pläne mit beliebig vielen unabhängigen Variablen, die in beliebig vielen Ausprägungen vorliegen, entworfen werden. Praktisch wird man jedoch kaum je mehr als drei oder vier Variablen mit zwei bis fünf Ausprägungen kombinieren. Dies aus dem einfachen Grund, weil die Kombination von mehr als vier Variablen bzw. von Variablen mit mehr als fünf Kategorien zu unübersichtlichen Situationen führt. Man kann dies anhand eines Experiments von Gage (1976) gut illustrieren. Gage experimentierte mit 3 Variablen à 2 Ausprägungen, d.h. mit $2^3 = 8$ Zellen. Wollen wir *mehr* Variablen überprüfen, wie das aufgrund der Ergebnisse der Prozess-Produkt-Forschung naheliegender wäre, da dort eine grössere Zahl von Bedingungsfaktoren für wirksamen Unterricht genannt wird⁶⁵, werden die Verhältnisse sehr rasch unübersichtlich. Bei 7 unabhängigen Variablen à 2 Ausprägungen ergeben sich $2^7 = 128$ Zellen. Wollten wir nicht nur mit 2, sondern mit 4 Ausprägungen pro Variable arbeiten, ergäben sich $4^7 = 16'384$ Zellen! Es ist schlicht unmöglich, ein solches experimentelles Design zu realisieren. Stellt man weiter in Rechnung, dass die Zellen nicht nur mit je einer Person, sondern zum Beispiel – wie im Experiment von Gage – mit ca. 13 Personen zu besetzen wären, würde man für den Versuchsplan $16'384 \times 13 = 212'992$ Schülerinnen und Schüler benötigen, eine völlig unrealistische Vorgabe – abgesehen davon, dass ein Forscher oder eine Forscherin, ja eine ganze Forschungsgruppe praktisch ihr Leben lang damit beschäftigt wären, dieses Design umzusetzen.

Nach Nathaniel Gage (1976; vgl. auch Abschnitt 5.3.2.1) beginnt der praktische Nutzen des wissenschaftlichen Wissens nachzulassen, sobald mehr als vier Variablen in die Analyse einbezogen werden, da man kaum mehr als vier Variablen gleichzeitig im Auge behalten könne. Offensichtlich gilt diese Feststellung nicht erst für die Praktikerinnen und Praktiker, sondern bereits für die Forscherinnen und Forscher, denn auch diese sind nicht in der Lage, multifaktorielle Versuchspläne beliebiger Grössenordnung zu beherrschen. Wobei zur Unübersichtlichkeit von Versuchsplänen mit mehr als vier Variablen beiträgt, dass faktorielle Pläne auch Interaktionen (Wechselwirkungen) zwischen den unabhängigen Variablen aufdecken lassen. Solche Interaktionen stellen aber sehr bald Interpretationsprobleme.

⁶⁵ Diskutiert haben wir individuelle Faktoren wie kognitive und motivationale Voraussetzungen auf Seiten der Schülerinnen und Schüler, didaktische Faktoren wie Quantität und Qualität des Unterrichts, Kontextmerkmale wie Grösse und Heterogenität der Schulklasse sowie familiäre Bedingungen, also insgesamt sieben Variablenbündel (vgl. Abschnitt 5.3).

5.4.3 Interventionsstudien

Trotz der offensichtlichen Grenzen des Experiments in der pädagogischen Forschung, spielt das Denken in Kategorien der experimentellen Methodik eine wichtige Rolle bei jeder Art von empirischer Untersuchung. Wo nicht im strengen Sinn experimentell geforscht werden kann, da ist man bemüht, wenigstens *einige* Kriterien der experimentellen Forschungslogik zu realisieren. Daraus ergeben sich Untersuchungspläne wie die «quasi-experimentelle» Forschung oder die Evaluationsforschung. Charakteristisch für diese Forschungsansätze ist, dass dem Moment des aktiven Eingriffs Genüge getan wird. Man kann daher auch von Interventionsstudien sprechen. Im Unterschied zur Survey- und zur Prozess-Produkt-Forschung, die die Wirklichkeit so nehmen müssen, wie sie ist, und denen damit ein *konservativer* Charakter zukommt⁶⁶, wird bei Interventionsstudien – wie es schon der Name sagt – in die Wirklichkeit eingegriffen («interveniert»), was ihnen einen *progressiven* Anstrich gibt. Ein Beispiel dafür sind die Programme zur kompensatorischen Erziehung.

Kompensatorische Erziehungsprogramme haben zum Ziel, benachteiligte Kinder zu fördern, um deren schulische Chancen zu optimieren. Verschiedene solche Programme wurden in den 1960er und 1970er Jahren durchgeführt und auf ihre Wirksamkeit hin evaluiert. Da es sich dabei um Interventionen handelt, kommt diesen Studien ein «quasi-experimenteller» Charakter zu. Damit ist gemeint, dass zwar die strengen Bedingungen eines Experiments nicht erfüllt sind, trotzdem aber eine gewisse Nähe zu einem experimentellen Design besteht.

Faktisch waren die Programme kompensatorischer Erziehung nicht besonders erfolgreich, wobei ein ausgiebiger Streit darüber ausgetragen wurde, wie weit die Wirksamkeit trotzdem gegeben war (vgl. Bronfenbrenner 1974). Die Frage ist auch, welche Wirkungen von solchen Interventionen erwartet werden können, wenn wir in Rechnung stellen, wie stark die familialen Bedingungen und die individuellen Voraussetzungen der Kinder das Lernverhalten und die schulischen Leistungen beeinflussen. Zu berücksichtigen ist auch, dass sich diese kompensatorischen Programme an eine besondere Gruppe von Kindern richteten (Kinder im Vorschulalter oder in den ersten Schuljahren aus sozio-ökonomisch oder kulturell benachteiligten Familien), so dass aus den Ergebnissen der Studien keine *generellen* Schlüsse gezogen werden können. Schliesslich ist das Zielkriterium zu berücksichtigen, das – wie eben erwähnt – in der Angleichung der Bildungschancen von Kindern aus verschiedenen sozialen Schichten lag. Die Studien stehen damit klar im Kontext eines *bestimmten* Verständnisses von pädagogischer Wirksamkeit und schliessen andere Zielkriterien aus (vgl. Abschnitt 1.2.5).

⁶⁶ Das gilt etwa für die Studien zum «direct teaching», die wir im Abschnitt 5.3.3.6 diskutiert haben. Korrelationsstudien erfassen, was der Fall ist, sie vermögen nicht aufzuzeigen, was möglicherweise ebenfalls gut wäre oder sogar besser sein könnte.

5.5 Metaanalysen

Wir haben schon erwähnt, dass es angesichts der Fülle an empirischen Einzelstudien, deren Ergebnisse oftmals nicht leicht auf einen gemeinsamen Nenner zu bringen sind, schwierig ist, den Aussagewert der Unterrichtsforschung zu beurteilen (vgl. die Schlussbemerkungen im Abschnitt 5.2.4 und Abschnitt 5.3.3.10). Seit den 1970er Jahren gibt es ein statistisches Verfahren bzw. eine Familie solcher Verfahren, die eine solche vergleichende Beurteilung verschiedener empirischer Studien zu einem identischen Forschungsthema auf quantitativer Basis erlaubt. Das Verfahren nennt sich «Metaanalyse» («meta-analysis»). Die Metaanalyse ist keine Erhebungstechnik, sondern ausschliesslich ein Verfahren zur *sekundären* Bearbeitung bereits vorliegender Forschungsergebnisse. Sie kann als eine besondere Form der Forschungsübersicht angesehen werden, die sich von qualitativen («narrativen») Reviewmethoden unterscheidet.

Ein besonderes Anliegen von Metaanalysen ist die Beurteilung der *Effektstärke*⁶⁷ eines gewonnenen Befundes. Gerade in einem so viel verzweigten und komplexen Feld wie der Bildungsforschung ist es wichtig, im Einzelnen beurteilen zu können, wie *stark* eine festgestellte, statistisch signifikante Beziehung zwischen zwei oder mehr Variablen ist. Die am häufigsten verwendeten Effektstärkenmasse sind einerseits die Differenz zwischen zwei Gruppenmittelwerten, dividiert durch ihre gemeinsame Standardabweichung (Cohen's *d*), und andererseits die Produkt-Moment-Korrelation (Pearson's Korrelationskoeffizient *r*).

Im Folgenden gehen wir exemplarisch auf die Ergebnisse zweier solcher Metaanalysen ein. Im ersten Fall handelt es sich um eine Arbeit von Fraser, Walberg, Welch und Hattie (1987), im zweiten um eine Arbeit von Wang, Haertel und Walberg (1993).

(1) Barry J. Fraser, Herbert J. Walberg, W. Welch und J. A. Hattie haben insgesamt 7'827 Untersuchungen, die ihrerseits bereits in 137 Metaanalysen zusammengefasst wurden, analysiert. Tabellarisch (vgl. Tabelle 5.3) dargestellt, ergeben sich folgende Determinanten der Schulleistung, wobei als Kriterium der Bedeutung eines Faktors der Korrelationswert aufgeführt wird.

Wenn man die einzelnen Determinanten und ihre Effekte überblickt, dann fallen die im Allgemeinen eher geringen Effektstärken auf, so dass man pauschal sagen könnte, dass alle untersuchten Bedingungsfaktoren zwar eine gewisse, aber durchwegs geringe Bedeutung für die Schulleistungen haben. Im Übrigen bestätigen sich Aussagen, die wir bereits im Zusammenhang mit unserer Diskussion der Prozess-Produkt-Forschung gemacht haben: Von vergleichsweise grossem Einfluss auf die schulische Leistung sind erstens die kognitiven Voraussetzungen der Schülerinnen und Schüler ($r = .49$), zweitens die Qualität des Unterrichts ($r = .47$) und drittens die Quantität des Unterrichts ($r = .38$).

⁶⁷ Die *Effektstärke* (engl. «effect size») ist ein objektives und standardisiertes Mass für die Grösse eines beobachteten Effekts (vgl. Field 2005, p. 34f.). Da das Effektstärkemass standardisiert ist, bedeutet dies, dass es zwischen verschiedenen Studien (mit unterschiedlichen Variablen, Skalen oder Messmethoden) ohne weiteres verglichen werden kann. Der Korrelationskoeffizient *r* nach Pearson ist nicht nur ein Mass für den Zusammenhang zweier Variablen (so kennen wir das *r* normalerweise), sondern kann auch als Mass für die Effektstärke herangezogen werden. Durchgesetzt hat sich der Vorschlag von Cohen wonach ein Effekt von $r = .10$ als *kleiner* Effekt, $r = .30$ als *mittlerer* Effekt und $r = .50$ als *grosser* Effekt interpretiert werden kann (vgl. Cohen 1988, 1992 zit. nach Field 2005, p. 32).

Tabelle 5.3: Determinanten der Schulleistung (aus: Helmke & Weinert 1997, p. 78, basierend auf Fraser, Walberg, Welch & Hattie 1987, p. 207)

Determinanten	Anzahl der Studien	Durchschnittl. r
Soziale Kontextbedingungen	153	.18
Beziehung zu Gleichaltrigen	12	.19
Häusliche Umwelt	118	.31
Konsum von Massenmedien	23	-.06
Schule	781	.12
Ziele und Politik	307	.12
Organisation (z. B. Klassengröße, traditionelle oder offene Klassenzimmer)	372	-.02
Lernumwelt (z. B. Zusammenhalt in der Klasse, curriculares Schwierigkeitsniveau)	201	.26
Lehrer	329	.21
Lehre (Instruktion)	1854	.22
Quantität	110	.38
Qualität	41	.47
Lehrmethoden	1763	.17
Spezielle Instruktionmethoden	2541	.14
Individualisierung	467	.07
Computerunterstützung	557	.15
Tutorensysteme	218	.25
Zielerreichendes Lernen	106	.25
Hausaufgaben	44	.21
Instruktionsmedien	657	.14
Schülermerkmale	1455	.24
kognitive	484	.44
affektive	355	.12
Lernstrategien	714	.28
Bekräftigungslernen	76	.49
remediales Lernen	97	.30

Legende: Determinanten = Faktoren, welche Schulleistungen beeinflussen oder bedingen; Anzahl Studien, in welcher die Determinante untersucht wurde und deren Einzel-Effektstärke in die Metaanalyse einbezogen wurde; Durchschnittl. r = durchschnittliche Effektgröße (über alle einbezogenen Studien gemittelt).⁶⁸

(2) Die zurzeit wohl umfassendste Metaanalyse der Forschungsliteratur zu den Bedingungen schulischer Leistungen stammt von Margareth C. Wang, Geneva D. Haertel und Herbert J. Walberg (1993). Die beiden Autorinnen und der Autor stützen sich nicht nur auf Metaanalysen, sondern zusätzlich auf konventionelle Überblicksdarstellungen der Forschungsliteratur (insbesondere auf eine Inhaltsanalyse) sowie eine Expertenbefragung. Die Inhaltsanalyse basiert auf 179 Review-Artikeln in wissenschaftlichen Zeitschriften und Handbüchern zu den Bedingungen schulischen Lernens. Bei der Expertenbefragung sind 134 Personen angeschrieben worden⁶⁹; davon haben 61 geantwortet (46%). Im Falle der Metaanalysen sind insgesamt 91 solche Analysen in die Untersuchung einbezogen worden, wobei es sich im Wesentlichen um die bereits erwähnten Analysen von Fraser, Walberg, Welch und Hattie (1987) handelt, die um sechs weitere Forschungssynthesen ergänzt worden sind.

⁶⁸ Anmerkung zur Tabelle: r kann auch im negativen Bereich liegen (z.B. -.06 für Konsum von Massenmedien). Das Vorzeichen sagt zusätzlich etwas über die *Richtung des Zusammenhanges* aus; in diesem Falle wäre der Zusammenhang also sehr schwach negativ, d.h. je mehr ein Kind Massenmedien konsumiert, desto schlechter sind seine Schulleistungen.

⁶⁹ Im Wesentlichen handelte es sich um die Autorinnen und Autoren der in die Inhaltsanalyse einbezogenen Artikel und Handbuchbeiträge.

Bei allen drei Verfahren – Inhaltsanalyse, Expertenbefragung und Metaanalyse – sind jeweils 228 Variablen in die Auswertung einbezogen worden. Untersucht wurde der Einfluss dieser Variablen auf die schulische Leistung, wobei als Mass wiederum Korrelationswerte angegeben werden.

Die 228 Variablen wurden in 30 Kategorien gruppiert, die ihrerseits zu sechs Konstrukten zusammengefasst wurden:

- State and District Governance and Organization
- Home and Community Educational Contexts
- School Demographics, Culture, Climate, Policies, and Practices
- Design and Delivery of Curriculum and Instruction
- Classroom Practices
- Student Characteristics

Die 179 Review-Artikel und die 91 Metaanalysen ergeben zusammen eine Basis von 270 Studien. Diese Studien decken ein breites Spektrum an potentiellen Determinanten der Schulleistung ab, von distalen Merkmalen wie demographischen oder bildungspolitischen Faktoren bis zu proximalen Faktoren wie Schülermerkmalen oder Unterrichtspraktiken. Um die Studien vergleichen zu können, sind sogenannte T-Werte berechnet worden. Die durchschnittlichen T-Werte der sechs Konstrukte sind in der folgenden Tabelle 5.4 dargestellt:

Tabelle 5.4: Durchschnittliche T-Werte der sechs theoretischen Konstrukte
(aus: Wang, Haertel & Walberg 1993, p. 270)

theoretical constructs	Average T score
Student Characteristics	54.7
Classroom Practices	53.3
Home and Community	51.4
Design and Delivery of Curriculum and Instruction	47.3
School Demographics, Culture, Climate, Policies, and Practices	45.1
State and District Governance and Organization	35.0

Das Konstrukt mit dem stärksten Einfluss auf die Schulleistung sind die Merkmale der Lernenden, gefolgt von Merkmalen der Unterrichtspraxis und den Hintergrundmerkmalen der Familie und der Schulgemeinde. Die Autorinnen und der Autor stellen fest:

«Student Characteristics, Classroom Practices, and Home and Community Educational Contexts are direct determinants of student learning while Design and Delivery of Curriculum and Instruction, School Demographics, Culture, Climate, Policies, and Practices, and State and District Governance and Organization are indirect determinants of student learning. Constructs that are close to students' defining characteristics and educative experiences – for example, psychological aptitudes, classroom practices, and home and community environments – exert the most influence on student learning. On the other hand, constructs that are removed from students and their everyday learning experiences, like State and District Policy, exert the least influence on student learning» (Wang, Haertel & Walberg 1993, p. 270).

Das Ergebnis bleibt praktisch unverändert, wenn nicht auf der relativ abstrakten Ebene der sechs Konstrukte, sondern auf der konkreteren Ebene der 28 Kategorien analysiert wird (vgl. Tabelle 5.5).

Tabelle 5.5: T-Werte, durchschnittliche T-Werte und Anzahl Quellen sowie Anzahl statistischer Zusammenhänge für jede Kategorie in Inhaltsanalysen, Expertenratings und Metaanalysen (aus: Wang, Haertel & Walberg 1993, p. 272f.)

Category	Content ratings	Expert ratings ^a	Meta-Analysis	Average	Number of sources in content ratings	Number of statistical relationships in meta-analyses
Classroom Management	59.5	64.9	70.0	64.8	42	15
Metacognitive	60.0	68.0	61.1	63.0	76	186
Cognitive	55.5	58.1	40.2	61.3	101	825
Home Environment and Parental Support	51.9	62.1	61.3	58.4	47	92
Student and Teacher Social Interactions	57.3	56.1	-	56.7	44	-
Social and Behavioral	55.5	55.0	-	55.2	35	-
Motivation and Affective	53.3	64.9	46.2	54.8	81	899
Peer Group	56.4	56.1	49.3	53.9	18	122
Quantity of Instruction	57.3	50.2	53.7	53.7	69	168
School Culture	49.2	57.7	52.8	53.3	49	185
Classroom Climate	56.8	54.2	45.9	52.3	75	734
Classroom Instructional	49.7	59.3	47.2	52.1	156	4095
Curriculum Design	51.0	51.0	52.0	51.3	97	752
Student and Teacher Academic Interactions	51.5	41.9	59.3	50.9	29	14
Classroom Assessment	51.5	52.6	47.3	50.4	61	45
Community	47.4	50.6	-	49.0	15	-
Psychomotor	71.2	36.3	39.3	48.9	6	637
Teacher/Administrator Decision Making	40.7	56.1	-	48.4	21	-
Curriculum and Instruction	52.8	44.3	46.0	47.7	108	1001
Parental Involvement Policy	41.6	43.1	52.6	45.8	23	1
Classroom Implementation Support	49.2	48.6	39.3	45.7	66	27
Student Demographics	43.0	41.1	50.4	44.8	90	904
Student Use of Out-of-School Time	53.7	46.6	32.6	44.3	17	274
Program Demographics	55.1	39.5	33.9	42.8	23	725
School Demographics	44.8	36.3	43.0	41.4	25	491
State and District Policies	22.4	32.8	56.0	37.0	19	22
School Policy and Organization	29.5	39.1	40.8	36.5	74	120
District Demographics	32.2	33.6	-	32.9	14	-
Accessibility	*	*	*	*	*	*
History of Educational Placement	*	*	*	*	*	*

^aThere were 61 respondents who rated each of the 30 categories.

- Data were unavailable for these categories; in analyses, they were coded as missing data.

*The Accessibility and History of Educational Placement categories were removed prior to computation of the *T* scores.

Anmerkung: Die Werte sind geordnet vom grössten (oberste Zeile; Average 64.8) bis zum kleinsten (Average 32.9; unterste Zeile) durchschnittlichen T-Wert der Kategorie.

Auch zwischen den drei Methoden ergeben sich kaum wesentliche Differenzen. In der Diskussion ihrer Ergebnisse schreiben Wang, Haertel und Walberg:

«Distal variables, like State, District, and School level policy and demographics, have little influence on school learning. [...] Based on the results of content analysis, expert ratings, and meta-analyses, summarized in this article, proximal variables strongly influence school learning. Proximal variables like psychological, instructional, and home environment variables have more impact on learning than most of the variables studied and should be part of an effective strategy to promote student learning. [...] If practitioners and teacher educators wish to enhance school learning, they must attend to proximal variables such as: (a) psychological variables, especially metacognition and cognition; (b) classroom instruction and management, and student and teacher social and academic interactions; and (c) the home environment» (ebd., p. 276, 278).

Und, zusammenfassend: «[...] the actions of students, teachers, and parents matter most to student learning [...] Ironically, State, District and School Policies that have received the most attention in the last decade of educational reform appear least influential on learning» (ebd., p. 279f.).

Diese relativ umfassende Gesamtschau verschiedenster Bedingungsfaktoren schulischer Leistungen kommt also zum Ergebnis, dass *proximale* Faktoren (wie Merkmale der Schülerinnen und Schüler, Unterrichtspraktiken und familiäre Bedingungen) weit einflussreicher sind für die Erklärung von Schulleistungen als *distale* Faktoren (wie demographische Merkmale, bildungspolitische Massnahmen oder organisatorische Merkmale der Schule). Wie auch Helmke und Weinert schreiben, gibt es «[...] keinen vernünftigen Zweifel daran, dass der Lernende mit seinen dispositionalen Merkmalen und seinem aktuellen Verhalten die wichtigste dynamische Determinan-

te der Schulleistungen und der Entstehung von Schulleistungsunterschieden darstellt» (Helmke & Weinert 1997, p. 99). Von deutlich nachweisbarer Bedeutung ist aber auch die Qualität des Unterrichts, so dass man mit Weinert feststellen kann: «Niemand wird daran zweifeln, dass Schulleistungen sowohl von der Qualität des Unterrichts als auch von den Kompetenzen der Schüler abhängen. Mit anderen Worten: Schulleistungen sind stets Leistungen der Schule und der Schüler!» (Weinert 2001b, p. 73). Aber auch damit ist die Aufzählung noch nicht abgeschlossen, denn zweifellos zeigen die Daten der vorgestellten Metaanalysen auch, wie wichtig die ausserschulischen Bedingungen schulischer Leistungserbringung sind: der familiäre Hintergrund, Einflüsse der Medien und der Peers, die Nutzung von Nachhilfeunterricht etc.

Es sind diese drei grossen Faktorenbündel – Merkmale der Schüler, Qualität des Unterrichts und ausserschulischen Einflüsse –, die letztlich die Hauptverantwortung für die schulischen Leistungen tragen.

Wenn wir in Bezug auf die Bedeutung der Lehrerinnen und Lehrer als Bedingung schulischer Wirksamkeit eine Bilanz ziehen wollen, dann ist festzustellen, dass deren «Persönlichkeit» wohl nicht die grosse Bedeutung zukommt wie oft behauptet wird, dass aber der Lehrer bzw. die Lehrerin mit der Art und Weise, *wie sie unterrichten*, sehr wohl Wesentliches zur Qualität von Schule und Unterricht beitragen. In den Worten von Weinert:

«Der Qualifikation von Lehrern kommt für die Qualität des Unterrichts und für die Entwicklung der Schüler grössere Bedeutung zu als dies einige modernistische Strömungen wahrhaben wollen. Lehrer spielen nicht nur als Lehrende eine wichtige, oft unterschätzte Rolle, sondern auch als Moderatoren von Lerngruppen, als Tutoren einzelner Schüler, als Gestalter anregender Lernumwelten, als Verhaltensmodelle und als verantwortliche Garanten für die simultane Erreichung unterschiedlicher Bildungsziele durch möglichst viele Schüler» (Weinert 2001a, p. 23).

Bei aller Wertschätzung einer aktiven und «teilautonomen» Schule, eines engagierten Kollegiums, einer professionellen schulinternen Fortbildung, eines guten Schulklimas und eines überzeugenden schulischen Leitbildes, auf schulischer Seite sind es die proximalen Faktoren des Unterrichts und das Lehrerhandeln, die letztlich über die Wirksamkeit einer Schule, gemessen anhand der Schulleistungen der Schülerinnen und Schüler, entscheiden.

5.6 Modelle

Wir haben bereits im 1. Kapitel das Problem angesprochen, dass die Rede von Effektivität bzw. Wirksamkeit von Schule und Unterricht unter Umständen mehr analytische Gewissheit verspricht, als faktisch bzw. methodisch eingelöst werden kann. Was Schülerinnen und Schüler leisten, kann selbst dann, wenn es als Ziel des Lehrplans formuliert ist, nicht einfach als Effekt von Schule verstanden werden. Immer ist damit zu rechnen, dass auch andere Einflüsse schulische Leistungen mitbedingen. Wenn man mit Andreas Krapp unter Schulleistung das «Ergebnis von Lernprozessen» versteht, «die durch Unterrichtsmassnahmen initiiert und/oder gesteuert wurden» (Krapp 1976, p. 92)⁷⁰, dann wird aller Wahrscheinlichkeit nach etwas behauptet, das sich wissenschaftlich nicht einlösen lässt, nämlich eine Analyse der Bedingungen schulischer Leistungen, die *durch den Unterricht* verursachte Leistungen von *andern* verursachten Leistungen zu differenzieren vermag.

Um diesem Fehler nicht anheimzufallen, sind wir auf ein *Modell* angewiesen, das uns zeigen kann, welche Faktoren in welcher Verbindung schulische Leistungen verursachen. Ein solches Modell

⁷⁰ Analog definieren Krapp und Mandl (1976, p. 196) sowie Rüdiger: «Schulleistung soll verstanden sein als Verlauf und Ergebnis schulisch-unterrichtlicher Lernprozesse, getragen von Unterrichtsmassnahmen (Lehrer), von Interaktionen (Lehrer-Schüler, Schüler-Schüler) und von persönlichen Aktivitäten (Einzelschüler)» (Rüdiger 1980, p. 129).

gibt es zurzeit jedoch nicht. Immerhin gibt es *Ansätze* dazu, die zeigen können, wo die Probleme liegen und wie man zu einem solchen Modell kommen könnte. Wir diskutieren im Folgenden fünf solche Ansätze: zwei einfache Modelle von Krapp (1976), das Modell von Carroll (1973), Modelle im Zusammenhang von Qualitätssystemen, Komplexe- und Mehrebenenmodelle sowie Modelle, welche Kontextbedingungen mit einbeziehen.

5.6.1 Zwei einfache Modelle von Krapp (1976)

Mit den Bedingungsfaktoren von Schulleistungen, wie sie in der resümierten Forschungsliteratur veranschlagt werden (der Schüler bzw. die Schülerin, die Familie des Schülers bzw. der Schülerin und die Schule bzw. der Unterricht, inklusive Lehrer bzw. Lehrerin), lässt sich ein *einfaches* Modell entwerfen (vgl. Abbildung 5.3).

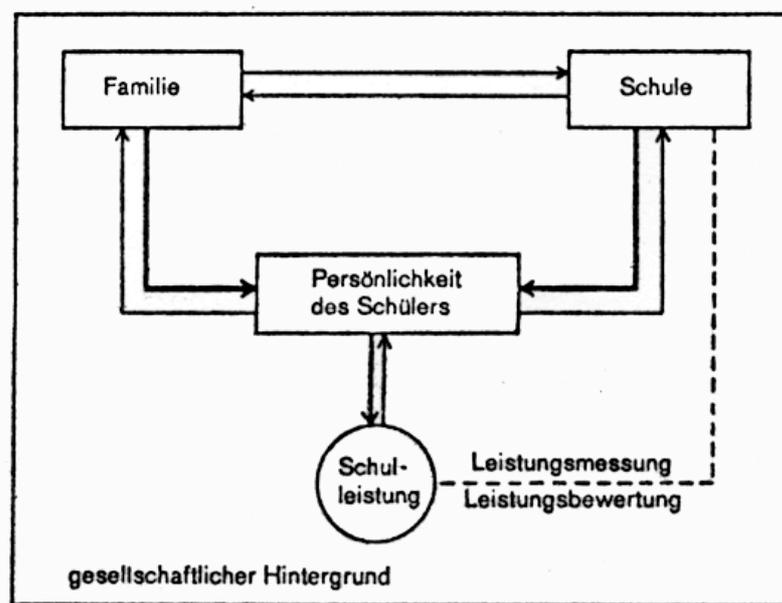


Abbildung 5.3: Grobstruktur eines Modells möglicher Determinantenbereiche der Schulleistung (aus: Krapp 1976, p. 94)

Krapp (1976) erläutert sein Modell wie folgt: Die Determinanten der Schulleistung sind als Kästchen dargestellt, das Kriterium in Form eines Kreises. Die Beziehung zwischen den Variablenbereichen wird durch Pfeile symbolisiert.

»Mit dem Kriterium, der Schulleistung, stehen unmittelbar nur die Variablen der Schülerpersönlichkeit in direkter Verbindung, denn jedes Verhalten ist zunächst Ausdruck und Ergebnis der psychophysischen Konstellation der Schülerpersönlichkeit. Die für uns wichtige Relation ist primär von der Schülerpersönlichkeit auf das Leistungsverhalten gerichtet, d.h. die Merkmale des Schülers sind Erklärungsvariablen für die Schulleistung. Die Beziehung lässt sich auch umkehren: das Leistungsverhalten hat über das Erlebnis von Erfolg und Misserfolg Rückwirkungen auf die Persönlichkeit des Schülers [...] Hier und an anderer Stelle wird man deshalb *Interaktionsbeziehungen* zu Grunde legen müssen» (Krapp 1976, p. 94).

Die Bedingungsmerkmale aus den beiden Bereichen Familie und Schule wirken nicht unmittelbar, sondern nur indirekt auf das Kriterium, indem sie die Persönlichkeit des Schülers, der Schülerin kurz- oder langfristig beeinflussen. Sie sind wesentlich beteiligt an der Entstehung von relativ

stabilen Lernvoraussetzungen und aktuellen schülerspezifischen Leistungsbedingungen. Auch die Variablen dieser beiden Merkmalsbereiche stehen untereinander in einem wechselseitigen Abhängigkeitsverhältnis. Die Beziehungen sind also nicht eingleisig zu interpretieren, wenngleich die Stärke der jeweiligen Abhängigkeiten durchaus ungleich verteilt sein kann.

Die gestrichelte Linie in Abbildung 5.3 zwischen dem Kriterium Schulleistung und dem variablen Bereich Schule soll deutlich machen, dass die Art, wie Schulleistung gemessen und bewertet wird, und die Definition dessen, was als Schulerfolg oder als Schulversagen zu gelten hat, vom jeweiligen Schulsystem festgelegt wird und von ihm auch verändert werden kann.

«Das Variablenmodell, d.h. alle darin vorgesehenen Variablen und die Beziehungen zwischen ihnen, gilt jeweils nur für bestehende *gesellschaftliche Rahmenbedingungen*. Deshalb denken wir uns das Modell auf einer Ebene angeordnet, die den 'gesellschaftlichen Hintergrund' repräsentiert. Ändert sich dieser Hintergrund, dann ändert sich u. U. das gesamte Modell, sowohl die Zusammensetzung der in die variablen Bereiche einzuordnenden Merkmale wie die Konstellation der Beziehungen zwischen diesen Merkmalen» (Krapp 1976, p. 95).

Ein wesentliches Manko dieses Modells liegt darin, dass es keine Zeitdimension aufweist. Erziehung und Unterricht sind jedoch *zeitbezogene Phänomene*, die prozessualen Charakter haben. Krapp (1976) hat daher eine Erweiterung seines Modells vorgeschlagen, die in der folgenden Abbildung 5.3 dargestellt wird.

Die Zeit kommt dadurch zu Gesicht, dass bei den Bedingungsfaktoren der Schulleistung jeweils in *frühere Entwicklungsbedingungen* und *aktuelle Bedingungsfaktoren* unterschieden wird.

«Frühere Entwicklungsbedingungen sind Einflussgrößen, die in Familie und Schule in der Vergangenheit dazu beigetragen haben, dass bestimmte Lernvoraussetzungen bei den Schülern entwickelt oder nicht entwickelt wurden. Der Erziehungsstil der Eltern ist in diesem Sinn eine typische Entwicklungsbedingung. Aktuelle Bedingungsfaktoren wirken in der Gegenwart, zum Beispiel das Unterrichtsverhalten des Lehrers. Das gleiche Merkmal kann in der Vergangenheit als Entwicklungsbedingung oder in der Gegenwart zusätzlich als aktuelle Einflussgrösse interpretiert werden (z.B. Sprachverhalten der Eltern)» (ebd., p. 95).

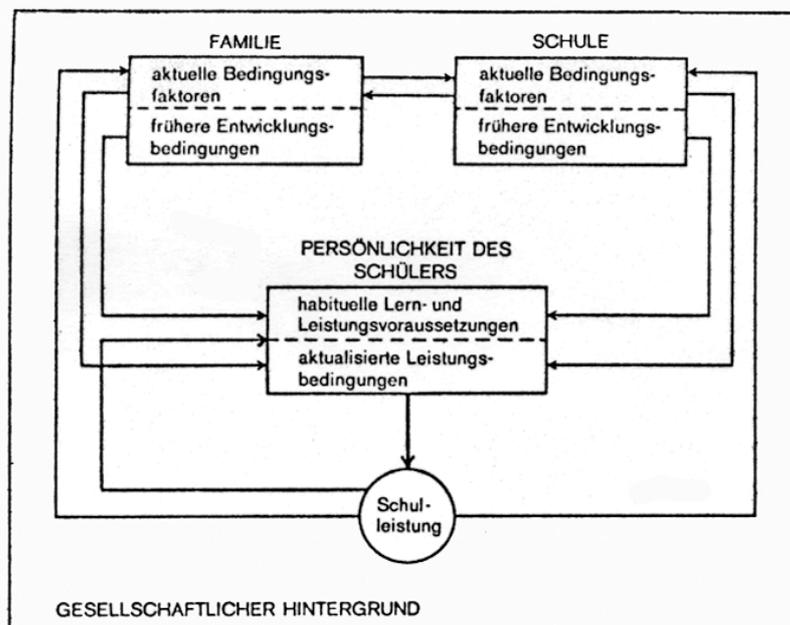


Abbildung 5.4: Erweitertes Modell zur Darstellung der Bedingungsfaktoren der Schulleistung (aus: Krapp 1976, p. 96)

Dieses zweite Modell zeigt nochmals, weshalb der Einfluss der schulischen Bedingungen unterschätzt wird, weil nämlich in der dominierenden Art von Unterrichtsforschung die früheren Einflussfaktoren nicht erfasst werden. Was wir an früherer Stelle mit einem Zitat von Nathaniel Gage (1979) ausgedrückt haben (vgl. Abschnitt 5.3.3.4), bringt Krapp (1976) dahingehend zum Ausdruck, dass Schule und Unterricht im Normalfall *über längere Zeit hinweg* wirksam sind und wir deshalb diesen Zeitraum *als ganzen* erforschen müssten. Während das Zitat von Gage (1979) im Grundsätzlichen zum Ausdruck bringt, weshalb die üblichen Methoden der Unterrichtsforschung vermutlich zu einer Unterschätzung der Einflüsse von Schule und Unterricht führen, zeigt das zweite Modell von Krapp (1976), wie die Aussage von Gage zu verstehen ist: Insofern Schule und Unterricht prozessualen Charakter haben, müssten wir die Gesamtheit der Prozesse erfassen, um den Anteil von Schule und Unterricht an den schulischen Leistungen angemessen zu erfassen.

Was die Schülerbedingungen anbelangt, so unterscheidet das Modell von Krapp:

«[...] zwischen habituellen Lern- oder Leistungsvoraussetzungen und aktualisierten Leistungsbedingungen. Habituelle Lernvoraussetzungen sind relativ zeitstabile, generalisierte Persönlichkeitsmerkmale, die im Entwicklungsprozess entstanden sind und durch zukünftige Entwicklungen verändert werden können. Intelligenz als Lernfähigkeit oder Art und Richtung des Leistungsmotivs als 'Motivdisposition' (Heckhausen) sind Beispiele dafür. Habituelle Lernvoraussetzungen erklären noch nicht das konkrete Leistungsverhalten. Erst aus der Wechselwirkung zwischen habituellen Lern- bzw. Leistungsvoraussetzungen und den in den Leistungssituationen vorhandenen Randbedingungen ergeben sich aktualisierte Leistungsbereitschaften, zum Beispiel die Intensität der Motivierung» (Krapp 1976, p. 96).

Gerade die *Intelligenz* ist ein Faktor, der nicht nur als Prädiktor für schulische Leistung in Frage kommt, da sich nachweisen lässt, dass der Schulbesuch die Intelligenz nicht unbeeinflusst lässt. Wir haben bereits im ersten Kapitel darauf hingewiesen, dass die Steigerung der Intelligenz eine (nichtintendierte) Wirkung von Schule und Unterricht darstellt.

Das erweiterte Modell von Krapp (1976) kann auch nochmals die Begriffe *unabhängige* und *abhängige Variable* illustrieren. Als abhängige Variable (Kriteriumsvariable) gilt die Schulleistung. Diese wird bedingt durch die unabhängigen Variablen in den Kästchen Familie, Schule und Persönlichkeit des Schülers. In allen drei Fällen können aber die aktuellen Bedingungsfaktoren auch als abhängige Variablen begriffen werden, die ihrerseits bedingt werden durch frühere Entwicklungsbedingungen, die dann als unabhängige Variablen erscheinen. Was als abhängige und was als unabhängige Variable bezeichnet wird, ist folglich eine Frage der Problemstellung.

Etwas Weiteres kann das Modell von Krapp (1976) nochmals illustrieren, nämlich das Ungenügen reiner Querschnittuntersuchungen. Querschnittuntersuchungen, die lediglich zu *einem* Zeitpunkt durchgeführt werden, klammern die Prozesscharakteristik von Erziehung und Unterricht aus. Sie vermögen lediglich die *aktuellen* Bedingungsfaktoren zu erheben. Dadurch erscheinen die Individuenvariablen («Persönlichkeit des Schülers») nur als unabhängige Größen, während sie in einer prozessualen Sichtweise (Längsschnittstudien) auch als abhängige Variable begriffen werden könnten. Ausgeklammert wird bei Querschnittstudien die Möglichkeit, dass die individuellen Bedingungen der Schulleistung ihrerseits auf pädagogische Einflüsse zurückgehen, womit – wie gesagt – Schule und Unterricht, aber auch andere Einflussfaktoren wie Familie, Peers und Medien, in ihrer Wirksamkeit falsch beurteilt werden.

5.6.2 Das Modell von Carroll (1973)

Die beiden Modelle von Krapp (1976) lassen sich auf verschiedene Weise erweitern. Eine Möglichkeit der Erweiterung bietet das Modell von *John Carroll*. Carroll (1973) geht von der einfachen, intuitiv allerdings leicht nachvollziehbaren Annahme aus, dass der Lernerfolg von der

Zeit abhängt, die, gemessen an den Lernvoraussetzungen, aufgewendet werden muss, um eine Aufgabe zu bewältigen. Das Niveau der Schulleistungen ist demnach m.a.W. eine direkte Funktion der aufgewendeten im Verhältnis zur benötigten Lernzeit. Noch einfacher formuliert, postuliert Carroll, «[...] dass der Lernende eine gegebene Aufgabe in dem Masse bewältigt, in dem er die für ihn nötige Lernzeit darauf verwendet» (Carroll 1973, p. 237).

Carroll (1973) unterscheidet *fünf* Faktoren, die er als Bedingungen des Lernerfolgs thematisiert: Die aufgabenspezifische Begabung (1), die Fähigkeit, dem Unterricht zu folgen (2), und die Qualität des Unterrichts (3) bilden die Determinanten der *benötigten* Lernzeit; die zugestandene Lernzeit (4) sowie die Zeit, die der Lernende bereit ist, mit Lernen zu verbringen (5) (seine Ausdauer), bilden die Determinanten der *aufgewendeten* Lernzeit. Die *ersten beiden* Faktoren sind trivial. Der *dritte* Faktor (Qualität des Unterrichts) umfasst die didaktische und methodische Qualität der Lehrkraft, aber auch die Qualität der Lehrbücher, Arbeitsunterlagen, beigezogenen Medien etc. Sofern die Qualität des Unterrichts aus irgendeinem Grunde nicht optimal ist, wird der Lernende mehr Zeit zur Bewältigung einer Lernaufgabe benötigen als er sonst gebraucht hätte. Damit wird offensichtlich, dass die Menge an Zeit, die eine Person benötigt, um eine gegebene Aufgabe zufrieden stellend zu lernen, nicht nur von ihrer Begabung abhängig ist, sondern auch von der Qualität des Unterrichts, der ihr geboten wird. Wobei das Kriterium, das Carroll ansetzt, um die Qualität des Unterrichts zu bemessen, nämlich die Entsprechung zwischen aufgewendeter und benötigter Lernzeit, unmittelbar zur Forderung führt, der Unterricht sei zu individualisieren. Das bestätigt auch der *vierte* Faktor (zugestandene Lernzeit), der sich auf die Zeit bezieht, die eine Schule oder eine Lehrkraft zur Verfügung stellen, damit eine Aufgabe gelernt werden kann.

Es bildet seit jeher ein Ideal des schulischen Unterrichts, dass die unterrichtlichen Bedingungen so weit individualisiert werden, dass das Lerntempo individuell gestaltet werden kann. Das bedeutet nichts anderes, als dass die Schule den einzelnen Schülerinnen und Schülern so viel Lernzeit zugesteht, wie sie individuell benötigen. Normalerweise ist dieser Grad an Individualisierung jedoch nicht erreicht, so dass die von der Schule zugestandene Lernzeit fast immer nicht der individuell benötigten Lernzeit entspricht. Der *fünfte* Faktor schliesslich, die Bereitschaft, eine Aufgabe zu bewältigen, ist wiederum eher trivialer Art. Je grösser die Ausdauer des Lernenden, desto eher ist zu erwarten, dass er die Lernaufgabe bewältigen wird. Eine formale Darstellung des Modells von Carroll (1973) gibt Abbildung 5.5:

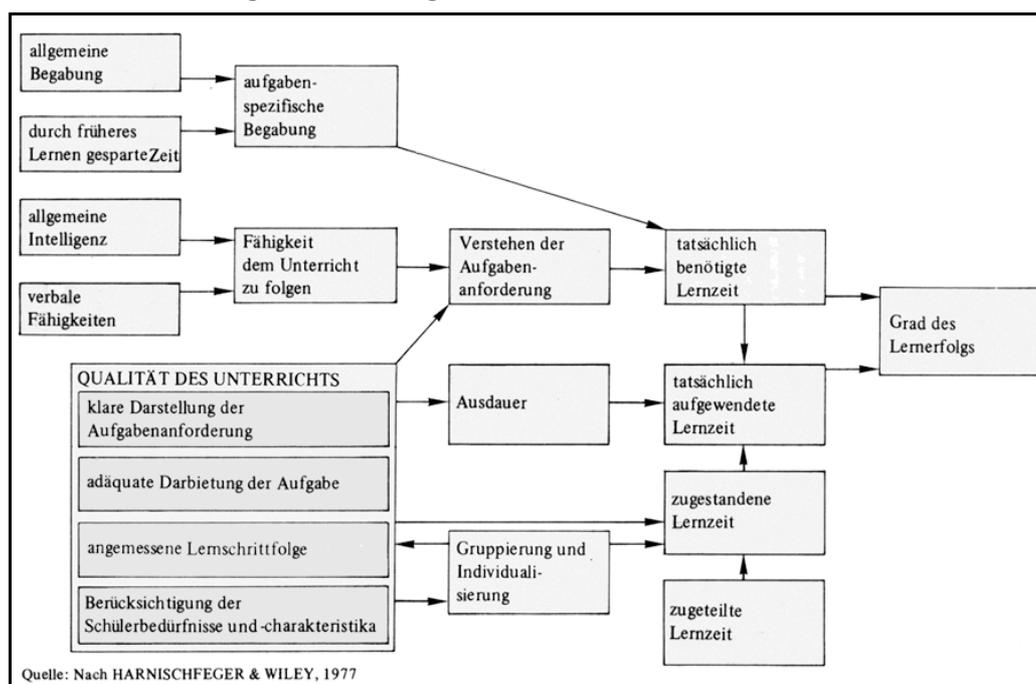


Abbildung 5.5: Modell schulischen Lernens von John Carroll (aus: Fend 1981, p. 323)

Wir wollen nicht weiter auf das Modell von Carroll eingehen, möchten aber darauf hinweisen, dass die im Rahmen der Diskussion des Prozess-Produkt-Paradigmas herausgearbeitete Bedeutung der *Unterrichtszeit* («time on task») als Bedingung schulischer Wirksamkeit aus diesem Modell unmittelbar ersichtlich wird (vgl. Abschnitt 5.3.3.5).

5.6.3 Modelle im Zusammenhang mit Qualitätssystemen

Eine andere Möglichkeit, die Modelle von Krapp zu erweitern, bieten die sogenannten *Qualitätssysteme*. Tatsächlich finden sich in der Qualitätsdiskussion Modelle, die komplexer angelegt sind als die beiden Krapp-Modelle, wobei insbesondere die Organisationsebene der Schule mitberücksichtigt wird.

a) Das sei zunächst am Beispiel von *Q2E* illustriert, einem von der Nordwestschweizerischen Erziehungsdirektorenkonferenz (NW EDK) entwickelten System zur Qualitätsevaluation und Qualitätsentwicklung auf der Sekundarstufe II. Vorgegeben werden zwölf Dimensionen der Schulqualität, aufgeteilt in vier Bereiche und drei Qualitätssegmente (vgl. Abbildung 5.6):

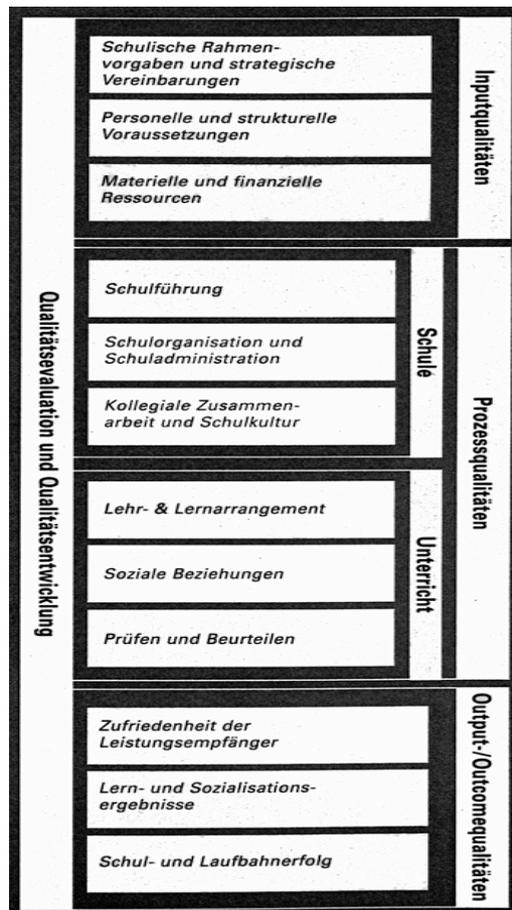


Abbildung 5.6: Modell von Q2E (aus: Landwehr 1999, o. S. - Q2E Basisinstrument zur Schulevaluation⁷¹)

⁷¹ Die Broschüre zum Q2E Basisinstrument für Schulevaluation und Schulentwicklung, welches von Landwehr (1999) gemeinsam mit der Nordwestschweizerischen Erziehungsdirektorenkonferenz (NW EDK) entwickelt worden ist, ist im Semesterapparat einsehbar.

Die drei Qualitätssegmente sind die Input-, die Prozess- und die Output- bzw. Outcomequalitäten einer Schule. Als Bedingungen für den Output bzw. das Outcome werden ebenfalls drei Bündel von Faktoren genannt: äussere Rahmenbedingungen, schulische Faktoren und Unterricht. Diese Dreiteilung entspricht recht gut dem Ansatz der Schuleffektivitätsforschung (vgl. Abschnitt 5.2).

(b) Ein dem Q2E ähnliches Modell hat Hartmut Ditton (2002) entwickelt, das insgesamt sechs Felder unterscheidet (vgl. Abbildung 5.7).

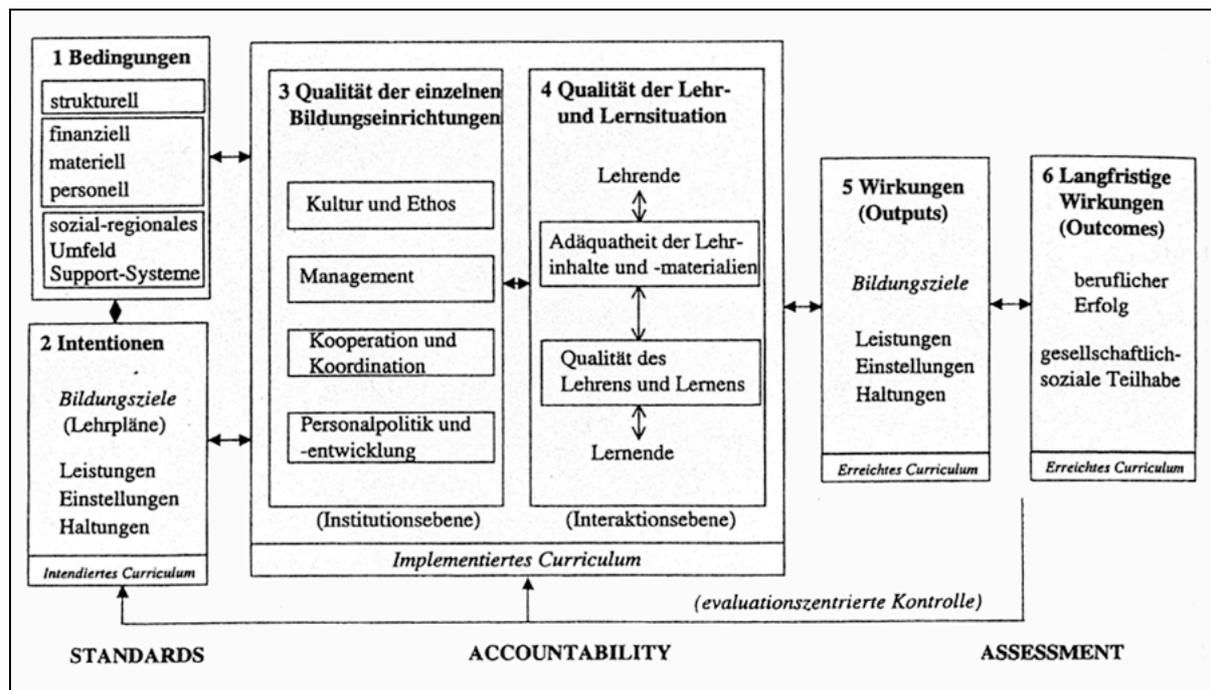


Abbildung 5.7: Modell zu Evaluation und Qualitätssicherung im Bildungswesen (aus: Ditton 2002, p. 776)

Genauso wie das Vier-Faktoren-Modell des Q2E ist das Modell von Ditton (2002) theoretisch schwach artikuliert. Letztlich erlauben auch die Modelle zur Qualitätsentwicklung keine klaren Aussagen zur relativen Bedeutung der verschiedenen Bedingungsfaktoren schulischer Wirksamkeit. So äussert sich das Basisinstrument Q2E nicht zur Bedeutung der drei Ebenen Bildungssystem, Schule und Unterricht und suggeriert damit eine *Gleichbedeutung*, die faktisch nicht besteht. Wir haben bei der Auseinandersetzung mit den Metaanalysen festgestellt, dass der Ebene der Einzelschule kaum jene Bedeutung zukommt, die ihr in den Qualitätssystemen zugeschrieben wird. Zwar gibt es durchaus einige Faktoren, die auf der organisatorischen Ebene liegen und für eine hohe Effektivität der Schule sprechen. Ditton nennt folgende Merkmale (vgl. Tabelle 5.6):

Tabelle 5.6: Bedeutsame Faktoren auf der Schulebene (aus: Ditton 2000, p. 85)

Tab. 2: Bedeutsame Faktoren auf der Schulebene	
Schulkultur	Schulmanagement
Gemeinsam akzeptierte, handlungsrelevante und eindeutige Ziele; Einigkeit hinsichtlich der primär zu erfüllenden Aufgabe; Organisatorische und pädagogische Leitung; Geregelte Zuständigkeiten und Verantwortlichkeiten;	Gemeinsam geteiltes Aufgabenverständnis; Eine gemeinsame <i>Vision</i> Geklärte Entscheidungsbefugnisse und -verfahren; Geregelte Aufgabenverteilungen
Kooperation und Koordination	Personalpolitik und Personalentwicklung
Koordinierter Schul- und Unterrichtsbetrieb; Kooperation mit Partnern außerhalb der Schule (Eltern, Administration, Berater ...) Einführung neuer Lehrer; Erfahrungsaustausch / Wissens-Sharing	Kooperation innerhalb der Schule (Schulleitung-Lehrer-Schüler); Rekrutierung, Sozialisation und Weiterbildung der Lehrer; Regelungen der Fort- und Weiterbildung;

Auf der anderen Seite muss man einräumen, dass die Forschung relativ deutlich zeigt, dass die Ebene der Schule nur bedingt die Qualität des Unterrichts vorhersagen lässt. In keiner Weise kann davon ausgegangen werden, dass eine hohe Schulqualität im Sinne institutioneller Struktur- und Prozessfaktoren Lernerfolge der Schülerinnen und Schüler *garantiert*. Entscheidender als die Rahmenbedingungen auf der Schulebene ist die Qualität des Unterrichts.

Tatsächlich kommt Ditton in einem Überblick zum Stand der empirischen Forschung zur Qualitätssicherung in Schule und Unterricht zum Ergebnis, dass sich die Bemühungen um Verbesserung der Schulqualität primär auf die Ebene des *Unterrichts* beziehen müssen – «[...] in der Regel auf die Ebene der einzelnen Schulklassen» (Ditton 2000, p. 73). Auch Fend, der ansonsten dazu neigt, die verschiedenen Ebenen eher als gleichwertig zu behandeln, betont, dass «[...] letztlich [...] die Unterrichtsebene für die Qualität des Lernens entscheidend ist» (Fend 2000, p. 62). «Die Qualität eines Bildungswesens materialisiert sich [...] in den konkreten Lehr-Lernprozessen einer Schulklasse» (ebd., p. 64). Erstaunlicherweise kommt Fend sogar auf einen Begriff zurück, der in der neueren erziehungswissenschaftlichen Literatur kaum von Bedeutung ist: die Persönlichkeit der Lehrperson. Es gelte, in beide Bereiche zu investieren, «[...] in qualitätsförderliche Rahmenbedingungen und in die Unterrichtskompetenzen bzw. in die Persönlichkeitsressourcen von Lehrern» (ebd.).

Aus pädagogischer Sicht heisst dies, dass die entscheidenden Faktoren der Effektivität von Bildungsbemühungen im situativen Bereich der pädagogischen Interaktionen liegen. Welche Faktoren gemeint sind, kann nochmals Ditton (2000) zeigen. Er gibt folgendes Schema, das auf der Basis verschiedener empirischer Studien entwickelt worden ist und sich an das sogenannte *QAIT-Modell* von Robert Slavin anlehnt (vgl. Tabelle 5.7)⁷²:

⁷² QAIT ist das Akronym für «Quality – Appropriateness – Incentive – Time of instruction». Gemeint sind damit Merkmale von Unterricht, die uns inzwischen hinlänglich bekannt sind, wie ein gut strukturierter und unterstützender Unterricht (Quality), der dem Lernniveau der Schülerinnen und Schüler angepasst ist (Appropriateness), der motivierend ist (Incentive) und in dem ausreichend Lernzeit zur Verfügung steht (Time of instruction).

Tabelle 5.7: Bedeutsame Faktoren des Unterrichts (aus: Ditton 2000, p. 82)

Tab. 1: Bedeutsame Faktoren des Unterrichts (QAIT-Modell)	
Qualität (Quality)	Motivierung (Incentives)
Struktur und Strukturiertheit des Unterrichts Klarheit, Verständlichkeit, Prägnanz Variabilität der Unterrichtsformen Angemessenheit des Tempos (Pacing) Angemessenheit des Medieneinsatzes Übungsintensität Behandelte Stoffumfang Leistungserwartungen und Anspruchsniveau	Bedeutungsvolle Lehrinhalte und Lernziele Bekannte Erwartungen und Ziele Vermeidung von Leistungsangst Interesse und Neugier wecken Bekräftigung und Verstärkung Positives Sozialklima in der Klasse
Angemessenheit (Appropriateness)	Unterrichtszeit (Time)
Angemessenheit des Schwierigkeitsgrades Adaptivität Diagnostische Sensibilität/ Problemsensitivität Individuelle Unterstützung und Beratung Differenzierung und Individualisierung Förderungsorientierung	Verfügbare Zeit Lerngelegenheiten Genutze Lernzeit Inhaltsorientierung, Lehrstoffbezogenheit Klassenmanagement, Klassenführung

Auch wenn als Zielkriterium die Chancengleichheit angesetzt wird, scheint die Unterrichtsebene von wesentlicher Bedeutung zu sein. Denn gleiche Chancen werden letzten Endes nicht durch Budgetzuteilung gewährt, sondern durch einen Unterricht, der auf unterschiedliche Lernvoraussetzungen der Schülerinnen und Schüler eingeht und deren Lernverhalten den Voraussetzungen gemäss unterschiedlich zu fördern vermag. So gesehen, muss das primäre Ziel der Schule sein, «[...] zunächst eine breite und verlässliche Grundlage für den weiteren Lernerfolg aller Schüler zu schaffen, um bei möglichst vielen Schülern die Voraussetzungen zum Erreichen eines hohen Leistungsniveaus zu gewährleisten» (Ditton 2000, p. 88). Ähnlich heisst es bei Fend, die Qualität eines Bildungswesens ergebe sich

«[...] letztlich daraus, ob es gelingt, die Schule für möglichst alle Schüler zu produktiven Räumen des Lernens und ihrer längerfristigen Entwicklung werden zu lassen. Aus Schülersicht sollten Schulen [...] vor allem Lernökologien sein, in denen [die Schüler] ihr Wissen, ihr Welt- und Selbstverständnis, ihre kulturellen Fertigkeiten und ihre motorischen Ausdrucksmöglichkeiten entfalten können» (Fend 2000, p. 56).

5.6.4 Komplexe- und Mehrebenenmodelle

Wenn man die Modelle von Q2E und Ditton (2002) mit den Modellen von Krapp (1976) vergleicht, dann fällt auf, dass die beiden Bedingungsfaktoren Familie und «Persönlichkeit des Schülers» verloren gegangen sind. Wir haben es also nach wie vor nicht mit *vollständigen* Modellen zu tun. Ein Ansatz zu einem solchen vollständigen Modell der Schulleistungsdeterminanten findet sich bei Helmke und Weinert (1997) (vgl. Abbildung 5.8).

Auch dieses Modell ist allerdings mit dem Mangel behaftet, dass es keine Auskunft erteilt über die Bedeutung der einzelnen Bedingungsfaktoren der schulischen Leistung, d.h. es ist theoretisch noch wenig artikuliert. Weder ist auf Anhieb ersichtlich, wie die Vielzahl an Faktoren zusammenspielen, noch welches Gewicht den einzelnen Faktoren im Einzelnen zukommt. Abgesehen davon, ist dem Modell schwer zu entnehmen, wie die zeitlichen Bedingungen schulischer Wirksamkeit zu fassen sind.

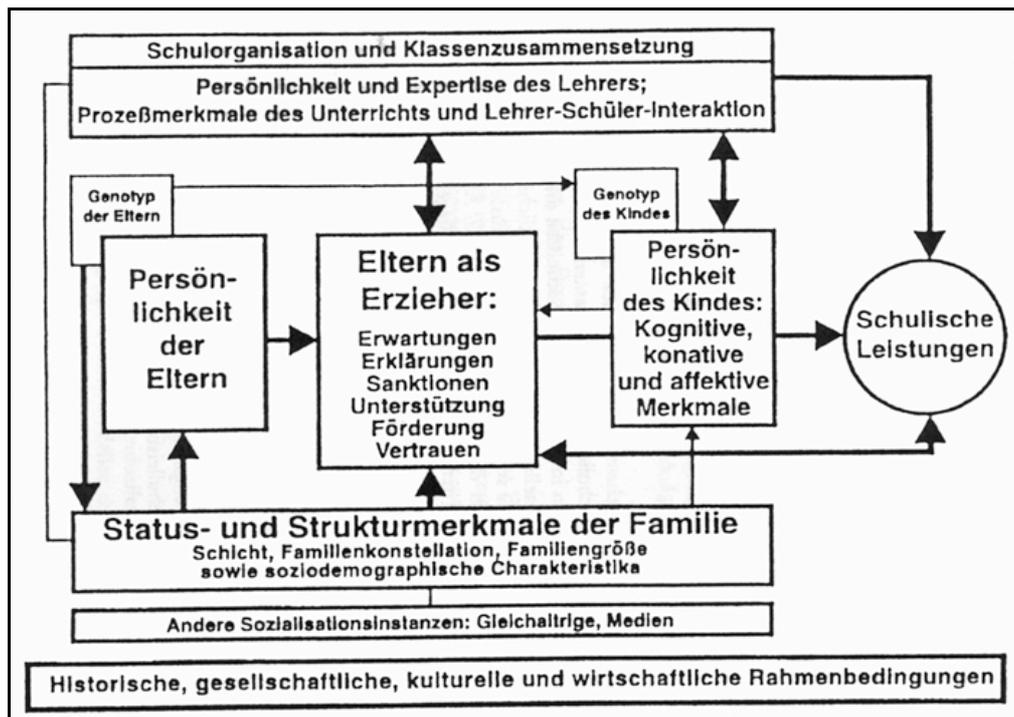


Abbildung 5.8: Komplexes Schema der Schulleistungsdeterminanten (aus: Helmke & Weinert 1997, p. 86)

Eine gewisse Abhilfe schaffen in jüngster Zeit sogenannte *Mehrebenenmodelle*. Dabei werden verschiedener *Ebenen* der Wirksamkeit unterschieden. Das ist insofern nichts Neues, als wir bei unserer Diskussion verschiedener Forschungsparadigmen implizit schon bisher eine solche Ebenenunterscheidung getroffen haben. Allerdings haben wir bisher keine entsprechenden Modelle zur Diskussion gestellt.

Es besteht eine Art Konsens, wonach für die Beurteilung der Wirksamkeit von Bildungssystemen vier Ebenen zu unterscheiden sind:

- Ebene des Schul- bzw. Bildungssystems
- Schulebene (Schultypus; Einzelschule)
- Klassenebene (inkl. Lehrkräfte)
- Individualebene (Schüler und Schülerinnen)

5.6.5 Einbezug von Kontextbedingungen

In den Gesamtmodellen schulischer und unterrichtlicher Wirksamkeit kommt oft ein Element vor, das wir separat anschauen wollen: die *Kontextbedingungen*. Kontextbedingungen betreffen insbesondere das soziale Umfeld einer Schule bzw. die sozio-ökonomische Lage der Schülerschaft. Im Modell von Krapp (1976) ist vom «gesellschaftlichen Hintergrund» die Rede, bei Helmke und Weinert (1997) von «soziokulturellen Rahmenbedingungen» und »Status- und Strukturmerkmalen der Familie«. Scheerens (1991) spricht von »context variables« auf den Ebenen der Schule und der Region. Er sieht im Einbezug kontextueller Merkmale eine «contingency perspective» und meint dazu:

«The general view of contingency theory is that it depends upon contextual characteristics whether or not specific organizational structures or managerial processes will be effective [...] At first glance contingency theory seems at odds with school effectiveness research, which is concerned with a set of school characteristics that is very robust in predicting effectiveness. On closer analyses of the research and literature on school effectiveness it becomes clear that claims regarding the generalization of

effective predictors across contexts have only partly survived empirical tests [...]» (Scheerens 1991, p. 385).

Auch intuitiverweise macht es Sinn, davon auszugehen, dass die Kontextbedingungen, unter denen Lehrkräfte unterrichten müssen, ihre Effektivität beeinflussen.

Wenn wir auf die IEA Lesestudie (Kapitel 2) und die PISA-Studie (Kapitel 4) zurückblicken, dann ist leicht ersichtlich, dass die Wirksamkeit einer Schule in Bezug auf die Vermittlung von Lesekompetenzen in der schulischen Erstsprache davon abhängig ist, wie gross der Anteil fremdsprachiger Kinder in einer Schule oder Klasse ist. Eine faire Beurteilung der Effektivität von Schule und Unterricht hat daher die Kontextbedingungen, unter denen in einem Schulsystem, einem Schultyp oder einer Einzelschule unterrichtet wird, in Rechnung zu stellen. Das ist durchaus wörtlich gemeint: Bei der Berechnung der Effektgrössen sind die Kontexteffekte gewissermassen abzuziehen. Man spricht dann von «bereinigten Werten» oder «korrigiertem Zuwachs» (adjusted gain). Die Rede ist auch von «residualisierten Werten» bzw. «Residuen». Der Gesamtzuwachs an Leistung (raw gain) wird korrigiert um das Eingangsniveau, von dem man ausgegangen ist und das von Schule zu Schule bzw. Klasse zu Klasse verschieden sein kann.

Kontextbedingungen können neben dem Anteil an fremdsprachigen Kindern auch sein: der sozio-ökonomische Hintergrund der Kinder, die Intelligenz der Kinder⁷³, deren Geschlecht, die materielle Ausstattung der Schule, Verhaltensauffälligkeiten der Kinder etc. Je vorteilhafter die Kontextbedingungen sind, desto eher ist mit guten Schulleistungen zu rechnen. Diese sind dann aber kein Effekt der Schule, sondern ein Effekt der positiv vorselegierten Schülerschaft bzw. der schulischen Infrastruktur. Die Frage allerdings, welche Kontextbedingungen für eine solche «Residualisierung» der Werte relevant sind, ist bis heute nicht definitiv entschieden.

5.7 Rückblick und Ausblick

Blicken wir nach Abschluss unseres Durchgangs durch verschiedene Ansätze der Wirkungsforschung auf die Anfänge dieser Forschungstradition bei James Coleman (1966) und Christopher Jencks (1972) zurück. Wie steht es um das eher pessimistische Urteil des Coleman-Reports, wonach Schule und Unterricht lediglich für etwa 10% der Varianz in den Schulleistungen verantwortlich sind, 90% also anderweitig bedingt sind? Eine neuere Schätzung gibt der niederländische Forscher Jaap Scheerens: «In interpreting school effects we should bare in mind that up to 75% of the variance in pupil achievement can be accounted for by factors like intelligence and social economic factors which are to a large extend beyond the reach of educative manipulation. This means that the total direct influence of schooling is relatively small as far as the pupils' overall achievement level is concerned» (Scheerens 1991, p. 392).

Nach Scheerens (1991) wären es demnach nicht lediglich 10%, sondern immerhin 25% der Varianz schulischer Leistungen, die von Einflüssen der Schule und des Unterrichts bedingt sind. Das mag noch immer als relativ bescheiden anmuten. Wenn Sie auf die Zitate von Comenius bis Brezinka, die wir im ersten Kapitel aufgeführt haben, zurückblenden, dann könnten Sie enttäuscht sein. Die pädagogische Tradition hat sich Erziehung und Unterricht als ein höchst wirksames Instrument der Menschenverbesserung vorgestellt. Offensichtlich ist die pädagogische Wirksamkeit aber bedeutend geringer einzuschätzen. Allerdings darf man nicht vergessen, dass angesichts der vielfältigen Einflüsse auf das Lernverhalten von Schülerinnen und Schülern ein Anteil von 25%,

⁷³ Diese dürfte im Normalfall nach dem Zufall verteilt sein, da die Intelligenz aber keine rein biologische Grösse darstellt, wird sie beeinflusst durch familiäre Bedingungen wie ein entsprechender Anregungsgehalt. Je nach familiärer Herkunft der Kinder, kann deren Intelligenz daher ungleich verteilt sein.

der auf Schule und Unterricht zurückgeht, relativ gross ist. Dazu kommt, dass die Forschungsansätze, die wir diskutiert haben, den Einfluss von Schule und Unterricht möglicherweise unterschätzen. Diesen Punkt wollen wir abschliessend nochmals erörtern, indem wir auf einige Grenzen der bisherigen Forschung zur schulischen und unterrichtlichen Wirksamkeit eingehen. Nur so kann es gelingen, die zuvor erwähnten Ergebnisse einer Vielzahl von Studien angemessen zu interpretieren.

Fragen wir also, welche *Grenzen* den Studien zur schulischen Wirksamkeit gesetzt sind. Wir wollen neun Punkte diskutieren:

1. Was wir unter schulischer Wirksamkeit verstehen, ist *kriteriumsabhängig*.
2. Es ist nicht möglich, in einer einzigen Untersuchung *alle* Bedingungen schulischer Wirksamkeit zu erfassen.
3. Je nach *Operationalisierung* der unabhängigen und abhängigen Variablen, kann das Ergebnis anders ausfallen.
4. Die Ergebnisse schulischer Wirksamkeitsstudien sind *stichprobenabhängig*.
5. Über den *Einzelfall* lassen die Studien keine verbindlichen Aussagen zu.
6. Die Studien unterschlagen den *zeitlichen* Charakter pädagogischer Prozesse.
7. Die Studien unterschätzen die Bedeutung *nicht-variierender* und *konstanter* Bedingungen.
8. Die *Vergleichsbasis* der Studien ist oft (zu) schmal.
9. Die Studien erfassen keine *nicht-linearen* Wirkungen.

(1) Wir haben schon im ersten Kapitel darauf hingewiesen, dass die Analyse schulischer Wirksamkeit kriteriumsabhängig ist. Zwar steht bei den meisten Untersuchungen die *Leistung* im Vordergrund (z.B. die Leistung in einem einzelnen Fach oder die übergreifende Kompetenz in einem bestimmten Bereich). Es gibt aber sehr wohl andere Kriterien, die man anlegen kann, um den «Schulerfolg» bzw. die Wirksamkeit von Schule und Unterricht zu beurteilen. Einige solcher alternativen Kriterien haben wir im ersten Kapitel erwähnt (vgl. Abschnitt 1.2.5).

(2) Bereits im ersten Kapitel ist auch auf Probleme der «anderen Seite» hingewiesen worden: Während beim Kriterium noch eher Konsens auszumachen ist, besteht bei den *Prädiktoren* der Schulleistung eine recht grosse Vielfalt, die sich bisher nur beschränkt in ihrer Bedeutung hat vereinheitlichen und bestimmen lassen. Von individuellen über familiäre, unterrichtliche und schulische bis zu gesellschaftlichen Bedingungen steht eine Fülle an Einflussfaktoren zur Auswahl, ohne dass sich eine klare Struktur dieses Faktorengefüges hätte herauskristallisieren lassen.

Es ist bisher nicht gelungen, ein Grundproblem der pädagogischen Wirkungsforschung zu lösen, nämlich die Abhängigkeit der gewonnenen Ergebnisse von den einbezogenen Variablen. Da schulische Leistungen offensichtlich von einer Vielzahl von Faktoren beeinflusst werden, ist es letztlich unmöglich, in einer *einzig* Untersuchung *sämtliche* denkbaren Einflussgrössen zu erfassen. Wenn beispielsweise in einer Studie 15 Bedingungen untersucht werden, in einer anderen ebenfalls 15, aber andere Bedingungen, dann lassen sich die Ergebnisse der beiden Studien nur bedingt miteinander vergleichen. So kann es sein, dass in der ersten Studie die motivationalen Voraussetzungen schulischer Leistungen berücksichtigt worden sind, in der zweiten aber nicht, dafür ist den familiären Bedingungen mehr Gewicht eingeräumt worden. Wenn nun in den Ergebnissen der ersten Studie die Familie und in denjenigen der zweiten die Motivation *nicht* im Bedingungsgefüge der Schulleistung erscheint, dann heisst dies nicht, dass diese Faktoren bedeutungslos wären – sie sind lediglich in den betreffenden Studien nicht berücksichtigt worden.

Denken Sie auch nochmals an die Probleme, die sich der experimentellen Analyse der Bedingungen schulischer Leistungen stellen: Es ist praktisch ausgeschlossen, im Rahmen eines faktoriellen

Versuchsplanes sämtliche denkbaren Ursachen schulischer Wirksamkeit systematisch zu überprüfen.

Diese Grenze der herkömmlichen Wirkungsforschung kann auch in Metaanalysen nicht überwunden werden. Sie könnte nur auf der Basis gründlicher Theoriebildung korrigiert werden.

(3) Empirische Studien sind aber nicht nur von der *Wahl* der unabhängigen und abhängigen Variablen abhängig, sondern auch von deren *Operationalisierung*. Wenn zum Beispiel die Leistung im Fach Mathematik als Kriterium gewählt wird, dann braucht es ein Messverfahren, um die Mathematikleistung zu erfassen. Das kann dadurch geschehen, dass die Zeugnisnote in Mathematik als Operationalisierung gewählt wird, oder es kann ein standardisierter Mathematiktest verwendet werden, oder es kann das Urteil der Lehrkraft als Indikator beigezogen werden etc. Vergleichbares gilt für die Prädiktoren: Für die Messung der Intelligenz als Prädiktor gibt es verschiedene Verfahren, die zwar durchaus zu ähnlichen Ergebnissen kommen, aber trotzdem nicht eins zu eins übereinstimmen. Wird die Qualität des Unterrichts als Prädiktor verwendet, dann gibt es eine Reihe von Möglichkeiten, diese empirisch zu erfassen (über das Urteil der Schülerinnen und Schüler, über Unterrichtsbeobachtungen, über Videoaufzeichnungen, über den Bericht des Schulinspektors oder des Schulleiters etc.). Zwar ist es ein Ziel der Bildungsforschung, möglichst einheitliche und gut bewährte Messinstrumente einzusetzen, doch dieses Ziel ist noch lange nicht erreicht, so dass sich viele der vorliegenden Untersuchungen auf der Ebene der Operationalisierung der unabhängigen und abhängigen Variablen nur bedingt miteinander vergleichen lassen, und dies wohlverstanden auch dann, wenn sie *dieselben* Konzepte zu messen vorgeben.

(4) Ins Kapitel Methodik gehört auch die *Stichprobenabhängigkeit* der erzielten Ergebnisse. Zunächst ist es trivial, dass Ergebnisse einer Studie im Kindergarten- oder Primarschulbereich nicht ohne weiteres auf das Gymnasium übertragen werden können. Auch ist in Rechnung zu stellen, dass Studien in verschiedenen Ländern nicht ohne weiteres miteinander vergleichbar sind. So gibt es Besonderheiten des amerikanischen Schulsystems, die die Übertragbarkeit von Ergebnissen auf das schweizerische Schulsystem begrenzen. Es ist daher bei jeder Studie zu berücksichtigen, auf welcher Stufe und in welchem Schulsystem sie durchgeführt worden ist. Dann aber spielt es zusätzlich eine Rolle, wie die Klassen, die in eine Stichprobe aufgenommen worden sind, ausgewählt wurden: Nach einem Zufallsverfahren? Weil sie gerade verfügbar waren? Weil man die Lehrkräfte gekannt hat? Weil es leicht war, eine Bewilligung für die Untersuchung zu erhalten? Auch die Zusammensetzung der Schulklassen kann von Bedeutung sein: Sind viele fremdsprachige Kinder in der Klasse? Sind Kinder der sozialen Mittelschicht überrepräsentiert? Sind es vorwiegend städtische Klassen, die man untersucht hat? Sind beide Geschlechter vertreten, d.h. handelt es sich um koedukativ unterrichtete Klassen?

(5) Zur Stichprobenabhängigkeit der Ergebnisse gehört auch, dass die diskutierten Determinanten der Schulleistung immer nur von *relativer* Bedeutung sind. Auf diesen Punkt haben wir schon mehrfach hingewiesen. Die Ergebnisse von Wirkungsstudien haben fast ausnahmslos *vergleichenden* Charakter. Es gibt kaum Studien, die in einem *absoluten* Sinn Aussagen über pädagogische Kausalprozesse zulassen. Fast immer sind die Daten so zu interpretieren, dass ein Effekt *relativ* zu den Schulen bzw. Variablen, die untersucht worden sind, besteht.

Eine nicht unwesentliche Konsequenz dieses Relativismus ist, dass wir im *Einzelfall* keine verbindlichen Aussagen machen können. Selbst wenn wir aufgrund der bisherigen Forschung Anlass zur Annahme haben, dass kognitive Variablen wichtiger sind als motivationale, oder dass die Qualität des Unterrichts wichtiger ist als der familiäre Hintergrund, kann daraus nicht abgeleitet werden, dass die Verhältnisse auch im Einzelfall so liegen. Ganz im Gegenteil müssen wir davon ausgehen, dass beim Schüler X oder bei der Schülerin Y die Anteile *anders* verteilt sind. Die in den

vorausgehenden Abschnitten diskutierten Forschungsergebnisse gelten nur *im Durchschnitt*, d.h. auf der Ebene sämtlicher einbezogener Einzeldaten. Da die Einzelfälle streuen, kann es ohne weiteres sein, dass beim Schüler X die kognitiven Bedingungen durch die motivationalen überspielt oder kompensiert werden, so dass es für eine Lehrkraft Sinn machen könnte, den Schüler vermehrt zu motivieren. Oder es kann sein, dass bei der Schülerin Y die Zusammenarbeit mit den Eltern mehr bringen würde als eine Anpassung des Unterrichtsstils. Die gleiche Schulleistung kann im einen Fall auf höherer kognitiver Kompetenz oder auf ungenügender Anregung durch das Elternhaus und mangelhaftem Unterricht beruhen, im anderen Fall auf geringer kognitiver Kompetenz, aber hoher Leistungsmotiviertheit, grossem Interesse am Fach und didaktisch hochstehendem Unterricht. Die auf Massendaten ausgerichtete Bildungsforschung erlaubt also keine Individualdiagnosen. Allerdings ermöglicht sie, bei der Individualdiagnose die *potentiell* relevanten Faktoren in Rechnung zu stellen. Doch wie deren Gewicht im Einzelnen liegt, muss in Bezug auf jeden einzelnen Schüler und jede einzelne Schülerin von der Lehrkraft erst überprüft werden.

(6) Eine Einschränkung, die wir ebenfalls schon verschiedentlich angesprochen haben, betrifft den *Zeitcharakter* von Erziehung und Unterricht. Die meisten Studien zu den Wirkungen von Schule und Unterricht haben punktuellen Charakter, d.h. sie sind zu einem bestimmten Zeitpunkt durchgeführt worden und vermögen nichts über den Zeitverlauf auszusagen. Es sind Momentaufnahmen von Schule und Unterricht, die deren Prozesscharakter systematisch ausblenden. Wir haben bereits darauf hingewiesen, dass auf diese Weise die Bedeutung von Schule und Unterricht systematisch unterschätzt wird, da die individuellen Bedingungen der Schulleistung keine statischen Grössen sind, sondern ihrerseits durch die Qualität von Schule und Unterricht beeinflusst werden. Dieser Einfluss, der von Schule und Unterricht auf die individuellen Leistungsbedingungen ausgeht, wird jedoch bei Studien ausgeblendet, die lediglich punktuell im Sinne von Querschnittanalysen Daten erheben.

Es ist auch nicht auszuschliessen, dass sich die Wichtigkeit einzelner Prädiktoren im Verlaufe der schulischen Karriere verschiebt. So gibt es plausible Hinweise dahingehend, dass die Intelligenz eine wichtige *Startbedingung* für den Schulbesuch ist, dass aber schon bald nicht mehr die Intelligenz per se, sondern das *erworbene Wissen* für den weiteren Schulerfolg verantwortlich ist. Ebenso haben wir gesehen, dass offensichtlich die motivationalen Faktoren im Verlaufe der Schulkarriere im Verhältnis zu den kognitiven an Bedeutung gewinnen, d.h. dass die motivationalen Ressourcen im Sekundarschulbereich wichtiger sind als im Primarschulbereich (vgl. Abschnitt 5.3.3.2).

Es ist bisher kaum gelungen, den Prozesscharakter von Schule und Unterricht angemessen in die Wirkungsforschung einzubeziehen. Das heisst auch, dass wir bis zur Stunde wenig wissen über pädagogische *Kausalprozesse*. Es ist zwar gelungen – und davon zeugt die zuvor geführte Diskussion in den Abschnitten 5.2 bis 5.6 –, die Variablen ausfindig zu machen, die *aller Wahrscheinlichkeit nach* für die Wirksamkeit von Schule und Unterricht die Hauptverantwortung tragen, aber wir haben noch kaum konkrete Vorstellungen darüber, wie diese Variablen kausal wirken. Insgesamt gesehen haben wir einige gute Kenntnisse über Variablen, die für schulische Leistungen verantwortlich sind, doch über die kausalen Mechanismen, welche die Korrelationen erklären, wissen wir ziemlich wenig.

(7) Auf ein weiteres Problem haben wir ebenfalls schon im ersten Kapitel aufmerksam gemacht. Die vorherrschende Methodik der Varianzaufklärung mittels korrelations- und regressionsanalytischer sowie ähnlicher Verfahren setzt streuende Werte sowohl auf Seiten der unabhängigen wie auf Seiten der abhängigen Variablen voraus. Damit werden *nicht-streuende* Einflussgrössen übersehen. So ist es relativ schwierig, schulische Bedingungen *konstanter* Art in die Modelle schulischer Wirksamkeit aufzunehmen, wie zum Beispiel Schulgesetze, Lehrpläne, Einschulungs-, Übertritts- und Selektionsreglemente, das System der Jahrgangsklassen, das Fachlehrersystem, die

Klassengrösse, die Ausbildung der Lehrkräfte etc. Zweifellos lassen sich auch diese Faktoren im Rahmen von interkantonalen und internationalen Studien in einem gewissen Ausmass vergleichend überprüfen, doch ist die Varianz selten gross genug, um deren Wichtigkeit sichtbar werden zu lassen. Das Reguläre ist zu unauffällig, als dass es vom Verfahren der Varianzaufklärung erkannt werden könnte. Hier liegt vermutlich ein weiterer Grund, weshalb der Einfluss von Schule und Unterricht auf die schulischen Leistungen in der bisherigen Forschung systematisch unterschätzt wird.

Das Modell der Varianzaufklärung ist auch nicht geeignet, um absolute Lernzuwächse zu beurteilen. Da es keine Kausalanalyse zulässt, vermag es die fundamentale Frage, ob Kinder und Jugendliche etwas auch *ausserhalb der Schule* hätten lernen können, nicht zu beantworten. Der Vergleich findet immer *innerhalb* derselben Institution (Schule) statt und erfasst lediglich Varianten *derselben* Institution. Es gibt keine Möglichkeit, die Leistungen der Schule als Ganzer gegenüber den *denkbaren Alternativen* auszutesten. Erst wenn wir Eltern hätten, die versuchen würden, denselben Stoff, der in der Schule vermittelt wird, ihren Kindern selber beizubringen, wäre ein solcher Vergleich möglich. Da uns aber allein schon die gesetzlichen Grundlagen für solche private Formen der Unterrichtung fehlen, können wir nicht wissen, was der Vorzug der Schule im Vergleich zur Familie ist.

(8) Mit dem zuvor genannten Punkt hängt zusammen, dass der Aussagewert vieler Studien zur schulischen Wirksamkeit möglicherweise dadurch beschränkt ist, dass ihre *Vergleichsbasis* zu schmal ist. Das gilt insbesondere für die grosse Zahl an Studien, die auf das Schulsystem der USA ausgerichtet sind. Es ist nicht auszuschliessen, dass Merkmale, die in amerikanischen Studien als hoch bedeutsam erscheinen, bei uns kaum auffallen, weil sie im schweizerischen Schulsystem nicht oder nur gering variieren.

Forschungsmethodisch spricht man von «ceiling-effects», d.h. «Deckeneffekten» bzw. «Grenzeffekten». Damit ist gemeint, dass bei einer Variable, die per definitionem verschiedene Ausprägungsstufen aufweist (z.B. im Sinne einer Likert-Skala mit den Vorgaben: 1 = sehr tief, 2 = tief, 3 = mittel, 4 = hoch, 5 = sehr hoch), in einer Stichprobe der Durchschnitt der erfassten Werte »an der Decke« liegt (z.B. bei 4.5 oder 4.9). Die Variable weist damit praktisch keine Streuung auf. Damit sprechen wir nochmals ein Problem an, das wir mit anderen Worten schon früher artikuliert haben: Wo keine Streuung (Varianz) ist, da kann auch keine Varianz *aufgeklärt* werden (vgl. Abschnitt 1.2.7).

Sofern wir davon ausgehen, dass die Qualität der Lehrerinnen- und Lehrerbildung in der Schweiz hoch ist (höher z.B. als in den USA), ist anzunehmen, dass die Qualität des Unterrichts ebenfalls hoch und auf einem hohen Niveau *einheitlich* ist, d.h. keine grossen Unterschiede zwischen Lehrkräften bestehen. Nach der Logik der statistischen Datenanalyse fallen demnach gerade die Qualitätsmerkmale des *Unterrichts* «unter den Tisch» – eben weil sie zu wenig streuen, d.h. der Unterricht tendentiell überall von hoher Qualität ist. Vergleichbares gilt möglicherweise für das *Fachwissen der Lehrkräfte*, obwohl dazu erstaunlicherweise kaum Untersuchungen vorliegen.

Damit lässt sich abschliessend nochmals auf die Frage nach der Bedeutung des Lehrers bzw. der Lehrerin für die Wirksamkeit von Schule und Unterricht zurückkommen. Wenn wir festgestellt haben, dass der Lehrkraft nicht die grosse Bedeutung zukommt, die ihr in der geisteswissenschaftlichen Pädagogik gerne eingeräumt wird, dann heisst dies nicht, dass es überhaupt nicht auf den Lehrer bzw. die Lehrerin ankommt. Es scheint aber, dass in einem gut entwickelten Schulsystem wie dem schweizerischen viele Stützfunktionen gegeben sind, die die Lehrerpersönlichkeit gewissermassen entlasten bzw. in den Hintergrund drängen. Zu diesen Stützfunktionen gehören: strukturierte Lehrpläne, gute Lehrmittel, standardisierte Prüfungsverfahren, eine gute Grund- und Weiterbildung der Lehrkräfte, ein stark differenziertes Schulsystem, eine strenge Schulaufsicht u.a.

Tatsächlich konnte Fend (1984) in einer Studie in Deutschland feststellen, dass in Schulen mit geringem Organisationsniveau der Einfluss der einzelnen Lehrkräfte bedeutend grösser ist als in

stark differenzierten Schulen. Im ersten Fall handelte es sich um Gesamtschulen (die bekanntlich ohne äussere Differenzierung auskommen), im letzteren Fall um herkömmliche Schulen (differenziert in Hauptschulen, Realschulen und Gymnasien). Je offener die Strukturen, d.h. je weniger das Lehrerverhalten von aussen durch organisatorische Massnahmen gestützt bzw. gelenkt wird, desto stärker treten Merkmale der Lehrerpersönlichkeit als Determinanten des Schulerfolgs in den Vordergrund. In den Worten von Fend: «Im administrativ und schulorganisatorisch weniger stabilen Kontext der Gesamtschulen werden Lehrerhaltungen bedeutsamer als dies bei der starken institutionellen Vorstrukturierung von Schulleistungen durch Schulvorgaben, Lehrbücher, Lehrpläne und Schulaufsicht im herkömmlichen Bildungswesen der Fall ist» (Fend 1984, p. 82). Je mehr Freiheit der Lehrkraft eingeräumt wird, d.h. je mehr sie auf eigene Leistungen bzw. Entscheidungen angewiesen ist, desto stärker wird ihr Einfluss im Bedingungsgefüge schulischer Wirksamkeit. Das ist eine durchaus wichtige Erkenntnis, denn sie zeigt nochmals, dass die Determinationsstruktur von Schulleistungen je nach Kontextbedingungen verschieden sein kann.

(9) Ein letzter Punkt betrifft die Modellvorstellungen über Kausalität. Alle Modelle, die wir angetroffen haben, gehen von einer *linearen* Kausalität aus: das Input-Output-Modell, das Prozess-Produkt-Modell und die experimentelle Unterrichtsforschung. Damit werden Feedbackprozesse ausgeblendet. Und es werden Kausalitätsunterbrechungen übergangen. Der letzte Punkt ist besonders wichtig, da wir es im Falle von schulischen Wirkungen nicht mit materiellen oder energetischen Prozessen zu tun haben. Übertragen wird im Unterricht nicht Materie oder Energie, sondern *Information* (Wissen), und diese Art von Übertragung funktioniert nach anderen Prinzipien als die physikalische Kausalität. Lebewesen sind sich in Bezug auf ihre informationellen Prozesse selbstgenügend. Ihr Wissen wird nicht kausalursächlich durch das Wissen anderer Lebewesen bedingt. Der Lernende muss sich sein Wissen selber aneignen; kein Lehrender vermag in das Bewusstsein eines Lernenden einzudringen. Insofern sind Menschen Kausalitätsunterbrecher, die die Wirkungen, die von anderen ausgehen, durch ihre eigenen Zustände determinieren. Darin liegt ihre Autonomie.⁷⁴

Die Idee der Autonomie widerspricht dem Input-Output-Denken, das von durchschlagenden Effekten ausgeht. Im Alltag kennen wir die Spruchweisheit «Kleine Ursache, grosse Wirkung». Davon finden wir in den referierten Studien zur schulischen Wirksamkeit wenig bis gar nichts. Noch am ehesten kommt der Gedanke im Angebot-Nutzungs-Modell zum Ausdruck (vgl. Abschnitt 1.3.2). Die Annahme linearer Verursachung impliziert eine *gleichförmige* Beziehung zwischen Prädiktor- und Kriteriumsvariablen: Mehr A führt zu mehr B, weniger A zu weniger B. Dass viel A nicht immer wenig B oder wenig A gelegentlich viel B auslösen könnte, lässt sich mit den diskutierten Modellen kaum erfassen. Die Beachtung der erwähnten Volksweisheit würde voraussetzen, dass wir die Autonomie der Schülerinnen und Schüler ernst nehmen. Wir hätten die Lernenden als *nicht-triviale Maschinen* im Sinne Heinz von Foersters (1985) zu verstehen (vgl. Abschnitt 1.3.1). Was daraus für die pädagogische Forschung und Wissenschaft folgen würde, wäre jedoch Thema für eine weitere Lehrveranstaltung.

⁷⁴ Darin liegt ihre «transzendente Freiheit» im Sinne von Kant.

6 Literatur

- Anderson, Linda M., Evertson, Carolyn M. & Brophy, Jere E. (1979). An Experimental Study of Effective Teaching in First-grade Reading Groups. *The Elementary School Journal*, 79, p. 193-223.
- Arnold, Karl-Heinz (1999). *Fairness bei Schulsystemvergleichen. Diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation*. Münster: Waxmann.
- Artelt, Cordula et al. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Ed.). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich, p. 69-140.
- Aurin, Kurt (1991). *Gute Schulen - Worauf beruht ihre Wirksamkeit?* Bad Heilbrunn: Klinkhardt.
- Baumert, Jürgen (1998). Internationale Schulleistungsvergleiche. In Rost, Detlef H. (Ed.), *Handwörterbuch Pädagogische Psychologie*. Weinheim: Beltz, p. 219-225.
- Baumert, Jürgen; Bos, Wilfried & Lehmann, Rainer (Eds.) (2000). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Band I: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.
- Baumert, Jürgen, & Köller, Olaf (1998). Nationale und internationale Schulleistungsstudien. Was können sie leisten, wo sind ihre Grenzen? *Pädagogik*, 50 (6), p. 12-18.
- Baumert, Jürgen; Stanat, Petra & Demmrich, Anke (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Ed.). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich, p. 15-68.
- Beaton, Albert E.; Mullis, Ina V.S.; Martin, Michael O.; Gonzales, Eugenio J. Kelly, Dana L. et al. (1996). *Mathematics Achievement in the Middle School Years. IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill: Boston College.
- BFS & EDK (2002). *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen – Nationaler Bericht der Erhebung PISA 2000*. Bildungsmonitoring Schweiz. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- BFS & EDK (2004). *Kompetenzmessung bei 15-Jährigen. 2005: Piloterhebung für den dritten Zyklus*. Bildungsmonitoring Schweiz. Neuenburg: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- BFS & EDK (2004). *PISA 2003: Kompetenzen für die Zukunft. Erster nationaler Bericht*. Bildungsmonitoring Schweiz. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- BFS & EDK (2005). *PISA 2003: Kompetenzen für die Zukunft. Zweiter nationaler Bericht*. Bildungsmonitoring Schweiz. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- BFS & EDK (2007). *PISA 2006: Kompetenzen für das Leben – Schwerpunkt Naturwissenschaften. Nationaler Bericht*. Statistik der Schweiz. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- Binkley, Marilyn; Rust, Keith & Winglee, Marianne (1995). *Methodological Issues in Comparative Educational Studies. The Case of the IEA Reading Literacy Study*. Washington D.C.: National Center for Education Statistics.

- Borich, Gary D. & Klinzing, Hans Gerhard (1987). Paradigmen der Lehreffektivitätsforschung und ihr Einfluss auf die Auffassung von effektivem Unterricht. *Unterrichtswissenschaft*, 15 (1), p. 90-111.
- Bos, Wilfried & Postlethwaite, T. Neville (2002). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. In Tippelt, Rudolf (Ed.), *Handbuch Bildungsforschung*. Opladen: Leske + Budrich, p. 241-261.
- Brezinka, Wolfgang (1981). *Grundbegriffe der Erziehungswissenschaft. Analyse, Kritik, Vorschläge*. München: Reinhardt (4. Aufl.).
- Bronfenbrenner, Urie (1974). *Wie wirksam ist kompensatorische Erziehung?* Stuttgart: Klett.
- Brookover, W. B.; Beady, C.; Flood, P.; Schweitzer, J. & Wisenbaker, J. (1979). *School Social Systems and Student Achievement. Schools Make a Difference*. New York: Praeger.
- Carroll, John B. (1973). Ein Modell schulischen Lernens. In Wolfgang Edelstein & Diether Hopf (Eds.), *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule*. Stuttgart: Klett, p. 234-250.
- Ceci, Stephen J. (1991). How Much Does Schooling Influence General Intelligence and Its Cognitive Components? A Reassessment of the Evidence. *Developmental Psychology*, 27 (5), p. 703-722.
- Chen, Chuansheng & Stevenson, Harold W. (1995). Motivation and Mathematics Achievement: A Comparative Study of Asian-American, Caucasian-American, and East Asian High School Students, *Child Development*, 66 (4), p. 1215-1234.
- Coleman, James S.; Campbell, E. Q.; Hobson, C.; McPartland, J.; Mood, A. M.; Winefield, F. D. & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, DC: Office of Education, U.S. Department of Health, Education, and Welfare.
- Comenius [Komensky], Johann Amos (1959). *Analytische Didaktik und andere pädagogische Schriften*. Ausgewählt und eingeleitet von Franz Hofmann. Berlin: Volk und Wissen.
- Comenius, Johann Amos (1965). *Pampaedia*. Nach der Handschrift hrsgg. von Dmitrij Tschizewskji in Gemeinschaft mit Heinrich Geissler & Klaus Schaller. Heidelberg: Quelle & Meyer (2. Aufl.).
- Comenius, Johann Amos (1993). *Grosse Didaktik*. Übersetzt und hrsgg. von Andreas Flitner. Stuttgart: Klett-Cotta (8. Aufl.).
- Demmer, Marianne (2000). Risiken und Nebenwirkungen von Schulleistungsvergleichen. Argumente gegen die Testeuphorie. *Pädagogik*, 52 (12), p. 33-35.
- Deutsches PISA-Konsortium (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Diekman, Andreas (1995). *Empirische Sozialforschung*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Ditton, Hartmut (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Eds.), *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik und Hochschule*. Weinheim: Beltz, p. 73-92.
- Ditton, Hartmut (2002). Evaluation und Qualitätssicherung. In Tippelt, Rudolf (Ed.), *Handbuch Bildungsforschung*. Opladen: Leske + Budrich, p. 775-790.
- Döbert, Hans (2002). *Trends in Bildung und Schulentwicklung: Deutschland und Europa*, [Online]. Verfügbar unter: <http://www.dipf.de/publikationen/tibi/trends_doebert_1.pdf> [November, 2002].

- Dolch, Josef (1961). Worte der Erziehung in den Sprachen der Welt. Ein Programm. In Brezinka, Wolfgang (Ed.), *Weltweite Erziehung*. Freiburg: Herder, p. 163-176.
- Fend, Helmut (1981). *Theorie der Schule*. München: Urban & Schwarzenberg (2. Aufl.).
- Fend, Helmut (1984). Determinanten von Schulleistungen: Wie wichtig sind die Lehrer? *Unterrichtswissenschaft*, 12 (1), p. 68-86.
- Fend, Helmut (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fend, Helmut (2000). Qualität und Qualitätssicherung im Bildungswesen. Wohlfahrtsstaatliche Modelle und Marktmodelle. In Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Eds.), *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule*. Weinheim: Beltz, p. 55-72.
- Field, A. (2005). *Discovering Statistics Using SPSS*. London: SAGE Publications (Second Edition).
- Fischer, Gerhard (1977). Experiment. In Herrmann, Theo; Hofstätter, Peter R.; Huber, Helmut P. & Weinert, Franz E. (Eds.), *Handbuch psychologischer Grundbegriffe*. München: Kösel, p. 136-143.
- Foerster, Heinz von (1985). Entdecken oder Erfinden. Wie lässt sich Verstehen verstehen? In Gumin, von Heinz & Mohler, Armin (Eds.), *Einführung in den Konstruktivismus. Schriften der Carl Friedrich von Siemens Stiftung*, Bd. 10. München: Oldenbourg, p. 27-68.
- Fraser, Barry J.; Walberg, Herbert J.; Welch, W. W. & Hattie, J. A. (1987). Syntheses of Educational Productivity Research. *International Journal of Educational Research*, 11 (2), p. 145-252.
- Fuller, Bruce & Heyneman, Stephen P. (1989). Third World School Quality. Current Collapse, Future Potential. *Educational Researcher*, 18 (2), p. 12-19.
- Gage, Nathaniel L. (1976). A Factorially Designed Experiment on Teacher Structuring, Soliciting, and Reacting. *Journal of Teacher Education*, 27 (1), p. 35-38.
- Gage, Nathaniel L. (1979). *Unterrichten – Kunst oder Wissenschaft?* München: Urban & Schwarzenberg.
- Gerstenmaier, Jochen (1975). *Urteile von Schülern über Lehrer*. Weinheim: Beltz.
- Gonon, Philipp; Hügli, Ernst; Landwehr, Norbert; Ricka, Regula & Steiner, Peter (1999). *Qualitätssysteme auf dem Prüfstand*. Die neue Qualitätsdiskussion in Schule und Bildung. Aarau: Sauerländer (2. Aufl.).
- Haertel, Geneva D.; Walberg, Herbert J. & Weinstein, Thomas (1983). Psychological Models of Educational Performance: A Theoretical Synthesis of Constructs. *Review of Educational Research*, 53 (1), p. 75-91.
- Helmke, Andreas (1992). Determinanten der Schulleistung: Forschungsstand und Forschungsdefizite. In Verband Bildung und Erziehung (Ed.), *Leistung in der Schule zwischen pädagogischem Anspruch und gesellschaftlichen Erwartungen*. Bonn: VBE-Dokumentationen, p. 7-14.
- Helmke, Andreas (1997). Individuelle Bedingungsfaktoren der Schulleistung: Ergebnisse aus dem SCHOLASTIK-Projekt. In Weinert, Franz E. & Helmke, Andreas (Eds.), *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlags Union, p. 203-216.
- Helmke, Andreas (2000). Von der externen Leistungsevaluation zur Verbesserung des Lehrens und Lernens. In Trier, Uri Peter (Ed.), *Bildungswirksamkeit zwischen Forschung und Politik*. Chur: Rüegger, p. 135-164.
- Helmke, Andreas (2001). Internationale Schulleistungsvergleichsforschung. Schlüsselprobleme und Perspektiven. *Zeitschrift für Pädagogik*, 47 (2), p. 155-160.

- Helmke, Andreas & Schrader, Friedrich-Wilhelm (2001). Jenseits von TIMSS: Messungen sprachlicher Kompetenzen, komplexe Längsschnittstudien und kulturvergleichende Analysen. In Weinert, Franz E. (Ed.), *Leistungsmessungen in Schulen*. Weinheim: Beltz, p. 237-250.
- Helmke, Andreas & Weinert, Franz E. (1997). Determinanten der Schulleistung. In Weinert, Franz E. (Ed.), *Psychologie der Schule und des Unterrichts. Enzyklopädie der Psychologie, Pädagogische Psychologie*, Bd. 3. Göttingen: Hogrefe, p. 71-176.
- Herbart, Johann Friedrich (1964). Pädagogisches Gutachten über Schulklassen. In ders.: *Sämtliche Werke, Bd. 4*. Hrsgg. von Karl Kehrbach & Otto Flügel. Aalen: Scientia, p. 519-556.
- Herbart, Johann Friedrich (1991). [Anonyme Rezension von Immanuel Kant: Über Pädagogik]. In Langewand, Alfred: *Moralische Verbindlichkeit oder Erziehung*. Herbarts frühe Subjektivitätskritik und die Entstehung des ethisch-edukativen Dilemmas. Freiburg: Alber, p. 305-308.
- Jencks, C.; Smith, M.; Acland, H.; Ban, M. J.; Cohen, D.; Gintis, H.; Heyns, B. & Mitchelson, S. (1972). *Inequality. A Reassessment of the Effects of Family and Schooling in America*. New York: Basic Books.
- Jensen, Arthur R. (1973). Wie sehr können wir Intelligenzquotient und schulische Leistung steigern? In Skowronek, Helmut (Ed.), *Umwelt und Begabung*. Stuttgart: Klett, p. 63-155.
- Keller, Carmen (1997). Geschlechterdifferenzen: Trägt die Schule dazu bei? In Moser, Urs; Ramseier, Erich; Keller, Carmen & Huber, Maja (Eds.), *Schule auf dem Prüfstand. Eine Evaluation der Sekundarstufe I auf der Grundlage der «Third International Mathematics and Science Study»*. Chur: Rüegger, p. 138-79.
- Kerschensteiner, Georg (1949). *Die Seele des Erziehers und das Problem der Lehrerbildung*. München: Oldenbourg (4. Aufl.).
- Kerschensteiner, Georg (1969). Der Lehrer als Erzieher. In Gerner, Berthold (Ed.), *Der Lehrer und Erzieher*. Bad Heilbrunn: Klinkhardt, p. 10-17.
- Krapp, Andreas (1976). Bedingungsfaktoren der Schulleistung. *Psychologie in Erziehung und Unterricht*, 29, p. 91-109.
- Krapp, Andreas & Mandl, Heinz (1976). Vorhersage und Erklärung der Schulleistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 8, p. 192-219.
- Landwehr, Norbert (1999): *Basisinstrument zur Schulevaluation: Bereiche, Dimensionen, Kategorien und Merkmale zu einer ganzheitlichen Analyse der Schulqualität*. Aarau: Nordwestschweizerische Erziehungsdirektorenkonferenz NW EDK.
- Lange, Hermann (2001). Die bildungspolitische Bedeutung von Schulleistungstudien. *Recht der Jugend und des Bildungswesens*, 49 (3), p. 262-282.
- Lustenberger, Werner (1996). *Pädagogische Rekrutenprüfungen*. Ein Beitrag zur Schweizer Schulgeschichte. Chur / Zürich: Rüegger.
- Malti, Tina (2002). *Bildung für Mädchen und Knaben*. In BFS/EDK (Ed.). Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Nationaler Bericht der Erhebung PISA 2000. Neuenburg: Bundesamt für Statistik. p. 136 - 155.
- Meyer, Andrea Hans & Zahner, Claudia (2002). *Kompetenzen in Lesen, Mathematik und Naturwissenschaften*. In BFS & EDK (Ed.), *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Nationaler Bericht der Erhebung PISA 2000*. Neuenburg: Bundesamt für Statistik, p. 23-52.
- Mislevy, Robert J. (1995). What Can We Learn from International Assessments? *Educational Evaluation and Policy Analysis*, 17 (4), p. 419-437.
- Mortimore, Peter; Sammons, Pamela; Stoll, Louise; Lewis, David & Ecob, Russell (1988). *School Matters. The Junior Years*. London: Open Books.

- Moser, Urs (2002). Kulturelle Vielfalt in der Schule: Herausforderung und Chance. In BFS & EDK (Ed.), *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Nationaler Bericht der Erhebung PISA 2000*. Neuenburg: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK), p. 113-135.
- Moser, Urs (2001). *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen – Kurzfassung des nationalen Berichtes PISA 2000*. Bildungsmonitoring Schweiz. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- Moser, Urs & Notter, Philipp (2000). *Schulische Leistungen im internationalen Vergleich. Ausgewählte schweizerische Ergebnisse aus internationalen Leistungsvergleichen bei Kindern, Jugendlichen und Erwachsenen*. Aarau: Schweizerische Konferenz der Kantonalen Erziehungsdirektoren (EDK).
- Moser, Urs; Ramseier, Erich & Berweger, Simone (2002). *Die Grundbildung in den drei Kantonen*. In BFS & EDK (Ed.), *Bern, St. Gallen, Zürich: Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Kantonaler Bericht der Erhebung PISA 2000*. Neuenburg: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK), p. 17-34.
- Moser, Urs; Ramseier, Erich; Keller, Carmen & Huber, Maja (1997). *Schule auf dem Prüfstand. Eine Evaluation der Sekundarstufe I auf der Grundlage der «Third International Mathematics and Science Study»*. Chur: Rüegger.
- Moser, Urs & Angelone, Domenico (2008). *PISA 2006: Porträt des Kantons Zürich*. Zürich: Institut für Bildungsevaluation, Assoziiertes Institut der Universität Zürich.
- Mullis, Ina V.S.; Martin, Michael O.; Beaton, Albert E.; Gonzales, Eugenio J. ; Kelly, Dana L. et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School. IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill: Boston College.
- Nieuwenboom, Jan W. (1996). Die Lesekompetenz der Drittklässler/innen. In Notter, Philipp et al. (Eds.), *Lernziel Lesen – Lesekompetenz von Jugendlichen in der Schweiz*. Schweizerischer Bericht über die IEA Reading Literacy Study, Aarau: Sauerländer, p. 51-77.
- Notter, Philipp (1996a). Die Lesefertigkeiten im internationalen Vergleich. In Notter, Philipp et al. (Eds.), *Lernziel Lesen – Lesekompetenz von Jugendlichen in der Schweiz*. Schweizerischer Bericht über die IEA Reading Literacy Study, Aarau: Sauerländer, p. 97-107.
- Notter, Philipp (1996b). Funktionaler Analphabetismus in der Schule? In Notter, Philipp et al. (Eds.), *Lernziel Lesen – Lesekompetenz von Jugendlichen in der Schweiz*. Schweizerischer Bericht über die IEA Reading Literacy Study, Aarau: Sauerländer, p. 108-118.
- Notter, Philipp; Meier-Civelli, Ursula; Nieuwenboom, Jan W.; Rüesch, Peter & Stoll, François (1996). *Lernziel Lesen. Lesekompetenz von Kindern und Jugendlichen in der Schweiz*. Aarau: Sauerländer.
- OECD (2007). *PISA 2006 – Schulleistungen im internationalen Vergleich. Naturwissenschaftliche Kompetenzen für die Welt von morgen*. Bericht. Paris Cedex: OECD.
- OECD (2001). *Lernen für das Leben. Erste Ergebnisse von PISA 2000. Ausbildung und Kompetenzen*. Paris Cedex: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung OECD.
- Pause, Gerhard (1973). Merkmale der Lehrerpersönlichkeit. In Ingenkamp, Karlheinz (Ed.), *Lehrer und soziale Interaktion in der Unterrichtsforschung*. Weinheim: Beltz, p. 11-97.
- Pekrun, Reinhard (2002). Vergleichende Evaluationsstudien zu Schülerleistungen: Konsequenzen für die Bildungsforschung. In: *Zeitschrift für Pädagogik* 2002, 48 (1), p. 111-128.

- Purkey, Stewart & Smith, Marshall S. (1991). Wirksame Schulen – Ein Überblick über die Ergebnisse der Schulwirkungsforschung in den Vereinigten Staaten. In Aurin, Kurt (Ed.), *Gute Schulen – Worauf beruht ihre Wirksamkeit?* Bad Heilbrunn: Klinkhardt, p. 13-45.
- Ramseier, Erich; Keller, Carmen & Moser, Urs (1999). *Bilanz Bildung: eine Evaluation am Ende der Sekundarstufe II auf der Grundlage der «Third International Mathematics and Science Study»*. Chur: Rüegger.
- Rheinberg, Falko (2002). Bezugsnormen und schulische Leistungsbeurteilung. In Weinert, Franz E. (Ed.). *Leistungsmessungen in Schulen*. Weinheim: Beltz, p. 59-71.
- Rosenthal, Robert & Jacobson, Lenore (1971). *Pygmalion im Unterricht*. Weinheim: Beltz.
- Rüdiger, Dietrich (1980). Determinanten der Schulleistung – die Relativierung empirisch ermittelter Einflussgrößen. In Schermaier, Josef; Rothbucher, Monika & Zecha, Gerhard (Eds.), *Aspekte praxisbezogener Pädagogik. Ambivalenz als methodisches Prinzip pädagogischen Denkens*. Salzburg: Müller, p. 129-143.
- Rüegg, Hans Rudolf (1866). *Die Pädagogik in übersichtlicher Darstellung. Ein Handbuch für Lehramtskandidaten, Volksschullehrer und Erzieher*. Bern: Verlag der J. Dalp'schen Buch- und Kunsthandlung.
- Rüesch, Peter (1996). Die Lesekompetenz der Achtklässler/innen. In Notter, Philipp et al. (Eds.), *Lernziel Lesen – Lesekompetenz von Jugendlichen in der Schweiz*. Schweizerischer Bericht über die IEA Reading Literacy Study. Aarau: Sauerländer, p. 78-97.
- Rüesch, Peter & Nieuwenboom, Jan W. (1996). Das Messen der Lesekompetenz. In Notter, Philipp et al. (Eds.), *Lernziel Lesen – Lesekompetenz von Jugendlichen in der Schweiz*. Schweizerischer Bericht über die IEA Reading Literacy Study. Aarau: Sauerländer, p. 40-50.
- Rutter, Michael (1983). School Effects on Pupil Progress: Research Findings and Policy Implications. *Child Development*, 54 (1), p. 1-29.
- Rutter, Michael; Maughan, Barbara; Mortimore, Peter & Ouston, Janet (1979). *Fifteen Thousand Hours. Secondary Schools and Their Effects on Children*. Cambridge: Harvard University Press.
- Rutter, Michael; Maughan, Barbara; Mortimore, Peter & Ousten, Janet (1980). *Fünftehtausend Stunden. Schulen und ihre Wirkung auf die Kinder*. Weinheim: Beltz.
- Scheerens, Jaap (1991). Process Indicators of School Functioning: A Selection based on the Research Literature on School Effectiveness. *Studies in Educational Evaluation*, 17 (2-3), p. 371-403.
- Scheerens, Jaap & Bosker, Roel J. (1997). *The Foundations of Educational Effectiveness*. Oxford: Pergamon.
- Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (1996). *Pressemitteilung*. Bern: SPP-Pressestelle.
- Spranger, Eduard (1969). Der geborene Erzieher. In Bräuer, Gottfried & Flitner, Andreas (Eds.), *Gesammelte Schriften*, Bd. I. Heidelberg: Quelle & Meyer, p. 280-338.
- Stanat, Petra & Kunter, Mareike (2001). *Geschlechterunterschiede in Basiskompetenzen*. In Deutsches PISA-Konsortium (Ed.). PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich, p. 251 - 270.
- Stevenson, Harold W. & Stigler, James W. (1994). *The Learning Gap. Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education*. New York: Touchstone.
- Thorndike, Robert L. (1973). *Reading comprehension education in fifteen countries: an empirical study*. New York: John Wiley & Sons.

- Tillmann, Klaus-Jürgen (2001). Leistungsvergleichstudien und Qualitätsentwicklung, oder: Auf dem Weg zu holländischen Verhältnissen? *Neue Sammlung*, 41 (3), p. 359-368.
- Uguroglu, Margareth E. & Walberg, Herbert, J. (1979). Motivation and Achievement: A Quantitative Synthesis. *American Educational Research Journal*, 16 (4), p. 375-389.
- Walker, D. (1976). *The IEA Six Subject Survey. An Empirical Study of Education in Twenty-one Countries*. Stockholm: Almqvist & Wiksell.
- Wang, Margareth C.; Haertel, Geneva D. & Walberg, Herbert J. (1993). Toward a Knowledge Base for School Learning. *Review of Educational Research*, 63 (3), p. 249-294.
- Weinert, Franz E. (2001a). Vergleichende Leistungsmessungen in Schulen – eine umstrittene Selbstverständlichkeit. In Weinert, Franz E. (Ed.), *Leistungsmessungen in Schulen*. Weinheim: Beltz, p. 17-31.
- Weinert, Franz E. (2001b). Schulleistungen – Leistungen der Schule oder der Schüler? In Weinert, Franz E. (Ed.), *Leistungsmessungen in Schulen*. Weinheim: Beltz, p. 73-86.
- Weinert, Franz E. (2002). Vergleichende Leistungsmessungen in Schulen – eine umstrittene Selbstverständlichkeit. In Weinert, Franz E. (Ed.), *Leistungsmessungen in Schulen*. Weinheim: Beltz, p. 18-31.
- Winkler, Michael (2000). Qualität und Jugendhilfe. Über Sozialpädagogik und reflexive Modernisierung. In Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Eds.). *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik und Hochschule*. Weinheim: Beltz, p. 137-159.
- Zahner, Claudia & Meyer, Andrea Hans (2002). *Die PISA-Studie in ihrem Kontext*. In BFS & EDK (Ed.), *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Nationaler Bericht der Erhebung PISA 2000*. Neuenburg: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK), p. 9-22.
- Ziller, Tuiskon (1884). *Allgemeine Pädagogik*. Hrsgg. von Karl Just. Leipzig: Verlag von Heinrich Matthes (2. Aufl.).

6.1 Zitierte Internetseiten

- Homepage des Max-Planck-Instituts für Bildungsforschung (Testaufgaben Mathematik 7./8. Klasse TIMSS): [Online] Verfügbar unter: <<http://www.mpib-berlin.mpg.de/TIMSSII-Germany/index.htm>> [Februar 2010].
- Homepage des Max-Planck-Instituts für Bildungsforschung (Ergebnisse von TIMSS): [Online] Verfügbar unter: <<http://www.timss.mpg.de/>> [Februar 2010].
- Homepage der OECD/PISA: [Online] Verfügbar unter: <http://www.pisa.oecd.org/document/35/0,2340,en_32252351_32236225_33664291_1_1_1_1,00.html> [Februar 2010].
- Homepage PISA Schweiz: [Online] Verfügbar unter: <<http://www.pisa.admin.ch>> [Februar, 2010].
- Internetseite ILMES (Internet-Lexikon der Methoden der empirischen Sozialforschung): [Online] Verfügbar unter: <http://www.lrz-muenchen.de/~wlm/ein_voll.htm> [Februar 2010].
- Homepage der Schweizerischen Eidgenossenschaft (Statistik der Schweiz, Interaktive Karten: Gymnasiale Maturitätsquote): [Online] Verfügbar unter: <<http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/06/dos/blank/05/01.html>> [Februar, 2010].