Deutsches Institut für Internationale Pädagogische Forschung

Z 43

Bibliothek für Bildungsgeschichtliche Forschung BERLIN

# MITTEILUNGEN UND NACHRICHTEN

SONDERHEFT

## DIE LEISTUNGEN IM NATURWISSENSCHAFTLICHEN UNTERRICHT IN DER BUNDESREPUBLIK IM INTERNATIONALEN VERGLEICH

Von Prof. Dr. Walter Schultze

FRANKFURT AM MAIN
1974

74:3988

DEUTSCHES INSTITUT FÜR INTERNATIONALE PÄDAGOGISCHE FORSCHUNG
UND GESELLSCHAFT ZUR FÖRDERUNG PÄDAGOGISCHER FORSCHUNG E.V.

MITTEILUNGEN UND NACHRICHTEN

SONDERHEFT

# Die Leistungen im naturwissenschaftlichen Unterricht in der Bundesrepublik im internationalen Vergleich

Von Prof. Dr. Walter Schultze

Frankfurt am Main
1974

Z 43

74. 3902

Pädagogische
Zentralbibliothek
Berlin

1) Naturwiss. ä.: Allg.
2) Erz. wiss.: Vgl. Päd.

Herausgeber: Forschungskollegium des Deutschen Instituts
für Internationale Pädagogische Forschung

Redaktion: Dr. Rosemarie Klaus-Roeder, DIPF,
6000 Frankfurt a.M. 90, Schloßstr. 29

# I n h a l t


| | | Seite |
|---|---|---|
| 1. | Vorgeschichte der Untersuchung | 2 |
| 1.1. | Entstehung der I.E.A. | 2 |
| 1.2. | Die Mathematikstudie, die 1. Phase der I.E.A.-Untersuchungen | 3 |
| 1.3. | Die 6-Fächer-Studie, die 2. Phase der I.E.A.-Untersuchungen | 4 |
| 2. | Die Anlage und organisatorische Durchführung der Untersuchung | 5 |
| 2.1. | Allgemeines | 5 |
| 2.2. | Die Stichprobe | 5 |
| 2.3. | Die Untersuchungsinstrumente | 8 |
| 2.4. | Die Verarbeitung der Daten | 10 |
| 3. | Die Bundesrepublik im internationalen Vergleich | 12 |
| 3.1. | Unterschiede in den Erziehungssystemen | 12 |
| 3.2. | Schulvariablen | 14 |
| 3.3. | Die Tests und Verhaltensskalen; statistische Werte | 16 |
| 3.4. | Allgemeine Analyse der Schülerleistungen | 19 |
| 3.4.1. | Die Ergebnisse nach Fächern und Leistungskategorien | 19 |
| 3.4.2. | Die Abhängigkeit der Leistungen von den Lehrplänen | 21 |
| 3.4.3. | Unterschiede zwischen den Geschlechtern | 21 |
| 3.4.4. | Unterschiede zwischen den Ländern und die ihnen zugrunde liegenden Faktoren | 23 |
| 3.4.5. | Die Bundesrepublik im Vergleich der relativen Leistungen | 27 |
| 3.5. | Unterschiede zwischen den Schulen und die ihnen zugrunde liegenden Faktoren | 30 |
| 3.5.1. | Das methodische Vorgehen | 31 |
| 3.5.2. | Ergebnisse | 33 |
| 3.6. | Unterschiede zwischen den Schülern und die ihnen zugrunde liegenden Faktoren | 36 |
| 4. | Der naturwissenschaftliche Unterricht in der Bundesrepublik | 38 |
| 4.1. | Die Stichprobe | 38 |
| 4.2. | Die außerschulischen Variablen | 41 |
| 4.3. | Korrelationen zwischen den Populationen | 43 |
| 4.4. | Analyse der Ergebnisse für die drei Populationen | 44 |
| 4.4.1. | Population I | 44 |
| 4.4.2. | Population II | 47 |
| 4.4.3. | Population IV | 51 |
| 5. | Zusammenfassung | 53 |

## 1. Vorgeschichte der Untersuchung

### 1.1 Entstehung der I.E.A.

Ab 1954 traf sich eine Gruppe europäischer Forscher auf Einladung von Dr. W.P.Wall, dem damaligen Direktor der National Foundation of Educational Research, London, einmal jährlich in Eltham bei Greenwich, um in informeller Weise Probleme der pädagogischen Forschung in Europa zu erörtern und sich gegenseitig über neuere Entwicklungen, Forschungsvorhaben usw. zu informieren (die "Eltham-Gruppe"). Daraus entwickelte sich eine Zusammenarbeit mit dem Unesco Institut für Pädagogik in Hamburg, das einzelne dieser Fragen zum Gegenstand internationaler Seminare machte, auf denen eine größere Zahl von Ländern vertreten war, an denen vor allem auch Forscher aus den USA teilnahmen.

Aus diesen Seminaren, die sich mit Ausleseproblemen, Examen, Schulversagen usw. im Zusammenhang mit Schulsystemen und Formen der Schulorganisation befaßten, entstand das Bedürfnis, die bisherigen Methoden internationaler Vergleiche durch Untersuchungsverfahren zu ergänzen, die es gestatteten, diese Vergleiche dadurch zu objektivieren, daß mit einheitlichen Meßinstrumenten, wie sie sich in nationalen Erhebungen und in bilateralen Studien bewährt hatten, gleichzeitig Untersuchungen in mehreren Ländern unter gleichen Bedingungen durchgeführt werden. So wurde 1958 beschlossen, eine Pilotstudie, an der sich 12 Länder (einschließlich der Bundesrepublik) beteiligten, zu unternehmen. Dabei sollten kurze Leistungstests in verständigem Lesen, Mathematik, Naturwissenschaften, Erdkunde sowie ein sprachfreier Gruppenintelligenztest unter möglichst einheitlichen Bedingungen angewendet werden, wobei zunächst auf Repräsentanz der Population verzichtet wurde. Die Erhebung erfolgte in 8 Sprachen an 9918 Schülern. (Foshay, A.W., ed.: Educational Achievement of Thirteen-Year-Olds in Twelve Countries. Hamburg, Unesco Institut für Päd., 1962.)

Die Ergebnisse zeigten, daß trotz aller Schwierigkeiten eine solche Methode sinnvoll anzuwenden war und ermunterten zu einer groß angelegten systematischen Untersuchung, wobei zunächst der Mathematikunterricht gewählt wurde, ein Bereich, der die geringsten Schwierigkeiten aufwarf.

Zu diesem Zweck strukturierte sich die Arbeitsgruppe 1961 als "Council of the International Project of Evaluation of Educational Achievement" und wählte einen ständigen

Ausschuß sowie ein Büro, das als ständige Einrichtung die geordnete Durchführung der Untersuchungen gewährleisten sollte. 1967 wurde sie als gemeinnützige internationale Gesellschaft in Belgien zugelassen (International Association for the Evaluation of Educational Achievement, I.E.A.). Die Mittel für ihre Untersuchungen werden, was die internationalen Kosten anbelangt, durch Regierungen und Stiftungen zur Verfügung gestellt, während die nationalen Zentren die Mittel für den auf sie entfallenden Teil der Untersuchung aus nationalen Quellen decken müssen. Infolgedessen sind alle Mittel projektgebunden.

Ziel der innerhalb der I.E.A. tätigen Wissenschaftler war es von Anfang an, die Vergleichende Erziehungswissenschaft um eine Methode zu bereichern, deren man sich bisher nur unvollkommen in einzelnen bilateralen Projekten bedient hatte: nämlich quantitative Methoden, die sich bisher in nationalen "Surveys" bewährt hatten, zu internationalen Untersuchungsmethoden auszuweiten und sie zu verwenden, um Hypothesen in vergleichenden Untersuchungen einer wissenschaftlich exakten Prüfung unterziehen zu können. Dabei heißt "quantitativ" in allen Untersuchungen vom Survey-Typ, die Beziehungen zwischen Eingangsvariablen (den Prädiktoren), die sozialer, ökonomischer oder pädagogischer Natur sein können, mit den Kriteriumsvariablen, d.h. in diesem Fall den Schulleistungen, rechnerisch in Beziehung zu setzen, wobei über die Ermittlung von Korrelationen hinaus statistische Methoden das verschiedene Gewicht der Eingangsvariablen am Zustandekommen der Leistung zu bestimmen gestatten.

## 1.2. Die Mathematikstudie, die 1. Phase der I.E.A.-Untersuchungen

Die Erfahrungen mit der Pilotstudie führten zur ersten Phase der I.E.A.-Arbeiten, einer vergleichenden Leistungsuntersuchung im Mathematikunterricht von 12 Ländern, wobei insgesamt 132 773 Schüler, 12 364 Lehrer und 5 348 Schulen einbezogen und insgesamt 50 Millionen Einzelinformationen statistisch aufgearbeitet wurden. (Walter Schultze und Lothar Riemenschneider: Eine vergleichende Studie über die Ergebnisse des Mathematikunterrichts in zwölf Ländern. In: Mitteilungen und Nachrichten, (1967) 46/47, S.1-34).
Bei den Überlegungen, welche Altersgruppen herangezogen werden sollten, einigte man sich auf 4 Populationen:

1. das Ende der Primarstufe, d.h. das 4.Schuljahr bzw. alle Zehnjährigen,
2. den letzten Jahrgang, der sich in allen Ländern noch zu 100% in der Schule befindet, d.h. das 8. Schuljahr, bzw. die Vierzehnjährigen,

3. den Jahrgang am Ende der Sekundarstufe I, bzw. die 15/16-jährigen,

4. den letzten Jahrgang der Sekundarstufe II, unabhängig von der Dauer der Schulzeit.

Den Ländern blieb es überlassen, welche Populationen sie einbeziehen wollten. Doch sollten aus diesen Jahrgängen Zufallsstichproben gezogen werden. Da es seinerzeit in der Bundesrepublik nicht möglich war, eine repräsentative Stichprobe für alle Länder aufzustellen, beschränkten wir uns auf zwei Länder, die sich aufgrund bisheriger Erfahrungen bei anderen Untersuchungen als charakteristisch für die Bundesrepublik ausgewiesen hatten: auf Hessen und Schleswig-Holstein.[1)]

1.3. Die 6-Fächer-Studie, die 2. Phase der I.E.A.-Untersuchungen

1966 wurde beschlossen, eine weitere Untersuchung vom Survey-Typ auf multi-nationaler Basis für mehrere Fächer durchzuführen. Man beschloß, Naturwissenschaften, verständiges Lesen, Literaturverständnis (durchgeführt 1970), Französisch und Englisch als erste Fremdsprachen sowie Sozialkunde (durchgeführt 1971) gleichzeitig in Angriff zu nehmen, wobei neben den Leistungstests wiederum Schüler-, Lehrer-, Schul- und nationale Fragebogen verwendet werden sollten. Außerdem wurden Attitüdenskalen entwickelt, um bestimmte Einstellungen und Haltungen der Schüler zu erfassen. Die Zahl der Länder im "Rat" war inzwischen auf 22 gestiegen; aber nicht alle Länder zogen sämtliche Bereiche sowie alle Populationen in die Untersuchung ein. Die Bundesrepublik beteiligte sich in den Naturwissenschaften, in Englisch und in Sozialkunde mit den Populationen I, II und IV.

Im folgenden soll über die Ergebnisse der Untersuchung in den Naturwissenschaften berichtet werden, wobei

---

1) Wegen weiterer Einzelheiten muß auf die genannte Veröffentlichung Schultze/Riemenschneider sowie auf folgende Veröffentlichungen hingewiesen werden:
   1. Torsten Husén (ed.): International Study of Achievement in Mathematics, A Comparison of Twelve Countries. 2 volumes. Stockholm, New York, London, Sydney/ 1967.
   2. Neville Postlethwaite: Leistungsmessung in der Schule. Eine internationale Untersuchung am Beispiel des Mathematikunterrichts. Aus dem Englischen übersetzt von Lothar Riemenschneider. Frankfurt, Berlin, Bonn, München, 1968.

sich die Darstellung hinsichtlich des internationalen Vergleichs auf die Veröffentlichung von Comber/ Keeves stützt. (L.C.Comber and John P.Keeves: Science Education in Nineteen Countries. Stockholm 1973.)

## 2. Die Anlage und organisatorische Durchführung der Untersuchung

### 2.1. Allgemeines

An der Untersuchung sollten nur diejenigen Länder teilnehmen, in denen im Sinne der Übereinkunft repräsentative Stichproben aufgestellt werden konnten. Dafür mußten in der Bundesrepublik zunächst die Voraussetzungen geschaffen werden, indem auf Sitzungen mit Vertretern der Kultusministerien der Zweck und die Durchführung der Untersuchung eingehend erörtert wurden. Die Kultusministerien erklärten sich bereit, die Schulen durch Rundschreiben zu bitten, an der Erhebung teilzunehmen.

Die finanziellen Voraussetzungen für die Teilnahme der BRD wurden dadurch gewährleistet, daß die Stiftung Volkswagenwerk sich bereit erklärte, die für die Durchführung der Untersuchung erforderlichen Mittel zur Verfügung zu stellen.[1] So war es möglich, für die Gesamtdauer der Untersuchung einen technischen Leiter sowie darüber hinaus zeitweise einen Diplompsychologen zu beschäftigen.[2]

### 2.2. Die Stichprobe

Für die Untersuchung waren drei der genannten Populationen vorgesehen: Pop.I, Pop.II und Pop.IV. Die Stichproben I und II sollten nach dem Lebensalter und nicht nach Schulklassen gezogen werden, obgleich die damit verbundenen erhöhten Schwierigkeiten durchaus gesehen wurden. Die Population IV umfaßte alle die Schüler des 13. Schuljahres, gleichgültig ob sie Naturwissenschaften als Wahlfach hatten oder nicht.[3]

---

1) Der Stiftung Volkswagenwerk gebührt dafür unser Dank, vor allem aber auch dafür, daß sie sich außerdem an der Finanzierung eines Teils der internationalen Ausgaben beteiligte.

2) Lothar Riemenschneider, Realschullehrer, hat dankenswerterweise die Gesamtleitung der technischen Durchführung auf sich genommen, Dipl.-Psych. Bernd Kötter ist ihm dabei während des Hauptanfalls der Arbeit eine große Hilfe gewesen.

3) Für die Angehörigen des naturwissenschaftlichen Zweiges waren zusätzliche Aufgaben vorgesehen (s.Tab.3). Es stand den Ländern frei, sie in die Untersuchung einzubeziehen. Das ist in der BRD geschehen, doch beziehen sich die Vergleichsangaben auf den Haupttest (60 Aufgaben).

Nicht einbezogen werden sollten Schüler aus Sonderschulen und aus Schulen, die ausgesprochen untypisch im Zusammenhang einer solchen Untersuchung sind, z.B. keine Versuchsschulen. Im übrigen war es den einzelnen Ländern freigestellt, die Untersuchung auszuweiten. So wurden in Schweden z.B. nicht nur die letzten Jahrgänge allgemeinbildender Gymnasien, sondern auch die letzten Klassen der Fachhochschulen (als Teil des schwedischen Gymnasiums) in diese Untersuchung einbezogen. Italien erweiterte darüber hinaus die Stichprobe zu einem nationalen Survey, indem es sämtliche Schulen in die Untersuchung einbezog.

Allgemeiner Grundsatz für die Aufstellung der Stichprobe war, von jeder Schule etwa dreißig Schüler in sie aufzunehmen und so viele Schulen wie möglich in die Untersuchung einzubeziehen, sofern das die zur Verfügung stehenden Mittel zuließen. Dabei war in der BRD die Ermittlung der in den Bundesländern auf die einzelnen Schülerjahrgänge entfallenden Schülerzahlen und ihre Verteilung auf die Schularten keine ganz einfache Aufgabe. Die Ministerien waren nicht immer imstande, die Listen aller Schulen mit den erforderlichen Angaben bereitzustellen; infolge der in den einzelnen Ländern laufenden Schulreformen und Versuche standen nicht einmal überall ausgedruckte Schullisten zur Verfügung. Eine Auswahl der Schulen für die drei Populationen erfolgte dann aufgrund der eingereichten Listen im Deutschen Institut.

Dabei wurden die Schulen weiterhin stratifiziert nach der Größe (groß, mittel und klein), sowie nach Knaben-, Mädchen- und Koedukationsschulen. Die einzelnen Schulen wurden angeschrieben, um Teilnahme und gegebenenfalls um Schülerlisten für die vorgesehene Population gebeten. Das war für die Schulen deshalb eine nicht ganz leichte Aufgabe, weil sich die Populationen I und II auf verschiedene Klassen verteilten. Wie aus Tabelle 1 ersichtlich, verteilten sich die für die Erhebung herangezogenen Schüler der Populationen I und II über jeweils vier Jahrgangsklassen, was die Durchführung der Untersuchung insofern erschwerte, als die Schüler für diesen Zweck aus den Klassen herausgezogen werden mußten.

Tabelle 1: Verteilung der Populationen I und II auf die Jahrgangsklassen (in %)

| Klassen | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Durchschnittsalter |
|---|---|---|---|---|---|---|---|---|
| Population I | 8 | 40 | 51 | 1 | | | | 10,5 |
| Population II | | | | 2 | 9 | 44 | 45 | 14,5 |

Quelle: Comber/Keeves, Tabelle 3.2

Diese Tatsache und der relativ große Umfang der Untersuchung schreckte viele der in die Stichprobe aufgenommenen Schulen ab. Das war vorhergesehen, und für diesen Fall war von vornherein eine zweite Stichprobe aufgestellt worden. Aber nicht immer war auch die vorgesehene Ersatzschule bereit, die Untersuchung durchzuführen, so daß dann infolge der fortschreitenden Zeit Ausfälle nicht immer ersetzt werden konnten. Die Zahl der in den Ländern vorgesehenen Schulen bzw. Schüler wurde deshalb nicht immer erreicht. Tabelle 2 gibt eine Übersicht über die Zahl der beteiligten Schulen, Lehrer und Schüler.

Tabelle 2: Zahl der in der BRD beteiligten Schulen, Lehrer und Schüler

| | Population I | Population II | Pop.IV |
|---|---|---|---|
| Zahl der Schulen | 68 | 83 | 80 |
| Zahl der Lehrer | 398 | 432 | 709 |
| Zahl der Schüler | 1742 | 2131 | 1989 |

Die Auswahl der Schüler erfolgte ebenfalls im Institut als Zufallsauswahl. Daß einzelne Schüler dann am Tage der Durchführung nicht anwesend waren oder aus anderen Gründen ausfielen, mußte in Kauf genommen werden. So wurden nicht immer die vorgesehenen dreißig Schüler pro Schule erreicht.

Die aufgestellten Stichproben wurden von einem internationalen Stichprobenexperten daraufhin überprüft, ob sie die erforderlichen Voraussetzungen erfüllten, bevor sie akzeptiert wurden.

Es sollte eine Verfahrensweise angestrebt werden, die unter Berücksichtigung der zur Verfügung stehenden finanziellen Mittel und Arbeitskräfte ohne allzu große Beeinträchtigung der laufenden Schularbeiten zu brauchbaren Ergebnissen führte. Das dürfte im allgemeinen erreicht worden sein. Doch sind im Hinblick auf die Auswertung einige Einschränkungen zu machen. Die geringe Zahl der Schulen bzw. der Schüler, die, bedingt durch die Größe der Gesamtstichprobe, auf kleine Bundesländer entfiel, macht es unmöglich, aufgrund der Ergebnisse gültige Aussagen für diese Bundesländer zu machen. Es kommt hinzu, daß gerade in diesen Ländern Schulen ausfielen, weil sie im Zusammenhang einer Schulreform aufgelöst bzw. zu Mittelpunkt- oder Gesamtschulen zusammengefaßt wurden oder aber aus undurchsichtigen Gründen das Material zum

Teil ohne Kommentar zurückschickten. Ersatzschulen konnten in manchen Fällen nicht mehr herangezogen werden; bzw. auch sie lehnten eine Teilnahme ab. Daraus erklärt sich die z.T. niedrige Beteiligung einzelner Länder, die deshalb für Vergleiche innerhalb der BRD ausfallen. (s.Tabelle 13)

## 2.3. Die Untersuchungsinstrumente

Die Verantwortung für die Tests, Fragebögen usw. lag bei den einzelnen internationalen Ausschüssen.

Nach einem international einheitlichen Kategorienschema wurden zunächst in den verschiedenen Ländern Lehrpläne für den naturwissenschaftlichen Unterricht für die betreffenden Stufen analysiert.

Aufgrund dieser Analysen wurden die ersten Aufgaben entwickelt und mit den gleichzeitig entworfenen Fragebogen an die nationalen Zentren zur Stellungnahme und Vorlage von Änderungsvorschlägen gesandt. Das führte z.T. zu erneuten Verhandlungen in den internationalen Ausschüssen mit Vertretern einzelner Nationen. Da die ersten Entwürfe etwa die vierfache Zahl von Aufgaben enthielten als für den endgültigen Test vorgesehen war, ergab sich für die nationalen Zentren hieraus eine erhebliche, nicht immer leicht durchzuführende Arbeit bei der Übersetzung und Adaption. Die Übersetzungen wurden vor der Probetestung dem internationalen Büro eingeschickt. Während die Tests für Population I vollständig von allen beteiligten Nationen erprobt wurden, wurde für Pop. II und IV von jedem Land etwa die Hälfte aller Aufgaben an 100 bis 200 Studenten je Population vorgetestet. Insgesamt wurden auf diese Weise etwa 700 items in mindestens sechs Ländern vorgetestet. Die Auswertung dieser Vortests erfolgte in den nationalen Zentren, die Ergebnisse dienten den internationalen Ausschüssen zur Auswahl der endgültigen Items.

Die Tests für jede Population waren folgendermaßen zusammengesetzt: Einmal mußten hinreichend viele Aufgaben (mindestens 6) für die verschiedenen Sachbereiche vorhanden sein, zweitens wurden die Aufgaben hinsichtlich der Lernziele ausgewählt, wobei man sich an die Bloom'sche Taxonomie anlehnte:

A. Kenntnisse, B. Verständnis, C. Anwendung,
D. Höhere Denkprozesse.

In fachlicher Hinsicht wurden neben Erdkunde, Biologie, Chemie, Physik "praktische Fähigkeiten" berücksichtigt, d.h. Tests, die einen Rückschluß z.B. auf die

Fähigkeit zuließen, Versuchsanordnungen zu beurteilen. Die Tabellen 3 und 4 geben einen Überblick über die zur Anwendung gekommenen Tests.

Tabelle 3: Die angewandten kognitiven Tests

| Population | Test | Zahl der Aufgaben | Zeit (in Min.) |
|---|---|---|---|
| I | 1 A | 20 | 30 |
| | 1 B | 20 | 30 |
| II | 4 A | 40 | 60 |
| | 4 B | 40 | 60 |
| IV | 10 A | 36 | 60 |
| | 11 B (P-U) | 30 | 60 |
| IV S | 15 (Biologie) | 40 | 60 |
| | 16 (Chemie) | 40 | 60 |
| | 17 (Physik) | 40 | 60 |

Quelle: Comber/Keeves, Tab. 2.3

Anmerkungen: Die Pop. IV S besteht aus Schulen mit dem Wahlfach Naturwissenschaften. Um Vergleiche mit denjenigen zu ermöglichen, die andere Zweige bzw. Kurse gewählt haben, wurden in Test 11B 6 Aufgaben (je 2 der drei Fächer) aus IV S aufgenommen. Um hinreichende Information zu erhalten, wurden 6 verschiedene Gruppen von Aufgaben gewählt, die so gegeben wurden, daß je 6 Schüler insgesamt 12 Aufgaben aus den drei Bereichen für die Pop. IV S bearbeiteten (sog. "rotierende Aufgaben").

Tabelle 4: Struktur der Tests

| Pop. | Zahl d. Tests | Zahl der Aufgaben aus: Erdk. | Biol. | Chem. | Phys. | Prakt. | nach Kategorien: A | B | C | D | Prakt. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 2 | 9 | 13 | 4 | 14 | - | 11 | 14 | 9 | 6 | - |
| II | 2 | - | 19 | 19 | 22 | 20 | 17 | 22 | 12 | 9 | 20 |
| IV | 2 | - | 18 | 18 | 18 | 12 | 7 | 22 | 15 | 4 | 12 |
| IV S | Biol. 1 | - | 40 | - | - | - | 6 | 13 | 11 | 10 | - |
| | Chem. 1 | - | - | 40 | - | - | 8 | 19 | 9 | 4 | - |
| | Phys. 1 | - | - | - | 40 | - | 4 | 18 | 14 | 4 | - |

Für jede Population wurde eine Reihe von Anker-Items aufgenommen, d.h. von Aufgaben, die in jeweils zwei benachbarten Populationen gleich waren, um einen Zuwachs der Leistung feststellen zu können. (Pop. I/II 11, Pop. II/IV 20)

Mit den Leistungstests wurde auf allen Stufen gleichzeitig ein Wortschatztest gegeben. Wortschatztests korrelieren erfahrungsgemäß sehr hoch mit der allgemeinen schulischen Leistungsfähigkeit. Sie wurden deshalb an Stelle von umständlicher zu administrierenden und auszuwertenden Gruppenintelligenztests gewählt.

Die Prädiktoren wurden durch einen Schüler/Lehrer- und Schulfragebogen sowie in Form umfangreicher Daten über das Schulwesen, die wirtschaftliche und gesellschaftliche sowie kulturelle Situation eines jeden Landes ermittelt. Auf diese Weise erhielt man eine große Zahl von Eingangsvariablen, von denen sich etwa die Hälfte als brauchbar erwies und später in die Auswertung eingangen ist.

Bei den Attitüdenskalen handelte es sich um Einstellungen zur Schule, Interesse für Naturwissenschaften und um Motivationsskalen (Pop. I) bzw. um Beurteilung der Rolle der Naturwissenschaften in der gegenwärtigen Welt (Pop. II und IV).

Ein 1969 durchgeführter Probelauf an zehn Schulen jeder Population gab endgültigen Aufschluß über die Verwendbarkeit dieses Instrumentariums. Das endgültige Instrumentarium wurde dann bei den einzelnen nationalen Zentren vervielfältigt bzw. gedruckt und nach entsprechender Kodierung für jeden beteiligten Schüler einzeln verpackt und an die Schulen versandt. Die Tests wurden in allen Ländern etwa 3 Monate vor Ende des Schuljahres administriert.

## 2.4. Die Verarbeitung der Daten

Die univariate Datenverarbeitung erfolgte in einem Rechenzentrum in New York, das Punktwerte, Mittelwerte, Standardabweichungen und Häufigkeitsverteilungen für alle Kriterien und für jedes Land und jedes Stratum ermittelte. Hier wurden gleichfalls gewichtete Mittelwerte errechnet, um unzureichende Schülerzahlen in einzelnen Populationen auszugleichen. Insgesamt waren bei 400 unabhängigen Variablen für 258 000 Schüler und 9 700 Schulen auf drei Stufen in 20 Ländern die Univariaten zu ermitteln. Die multivariate Analyse wurde im Rechenzentrum Stockholm durchgeführt.

Zum Verfahren der Verarbeitung:

Um das Gewicht und den Anteil der verschiedenen Faktoren (Variablen) an der Variabilität der Leistungen (dem Unterschied der Leistungen sowohl zwischen den Schülern wie zwischen den Schulen) zu ermitteln, bediente man sich der Regressionsanalyse. Sie ist ein statistisches Verfahren, mit dessen Hilfe auf rechnerischem Wege nach Ermittlung der Korrelation zwischen den Eingangsvariablen (Prädiktoren)[1] und den Kriteriumsvariablen[2] bestimmte Prozentanteile der Leistungsvariabilität erklärt werden.

Da die Eingangsvariablen aufgrund hypothetischer Annahmen über ihren Zusammenhang mit der Leistung festgelegt werden, kann nur so viel an Varianz durch eine Regressionsanalyse geklärt werden, wie durch die Eingangsvariablen in den Griff bekommen wurde. Daraus erklärt es sich, daß niemals die gesamten Faktoren, die für Leistungsunterschiede in Frage kommen, durch eine Regressionsanalyse erfaßt werden. Die Differenz zwischen der Summe der Prozente der erklärten Varianz und 100 stellt dann das Gewicht der durch die Untersuchung nicht erfaßten Faktoren dar. Diese nicht berücksichtigten Variablen können komplexer Natur sein, so daß sie als Eingangsvariable nicht klar kategorisiert werden konnten; sie können aber auch komplexe Variable (z.B. Lehrerverhalten) darstellen, die nicht meßbar sind. So bleibt bei der vorliegenden Untersuchung trotz der großen Fülle der

1) Eingangsvariablen (unabhängige Variablen) erfassen Faktoren, die in einem in ihrem Ausmaß und ihrem mathematischen Wert unbekannten Maße in einer Beziehung zur Leistung stehen und deshalb voraussichtlich für die Unterschiede der Leistungen verschiedener Systeme usw. verantwortlich sind. Dazu gehören z.B. Klassen- und Schulgröße, Schultyp, Lehrerausbildung, Stundenzahl für den betreffenden Unterricht, die Möglichkeit, den betreffenden Inhalt der Testaufgaben im Unterricht gelernt zu haben, aber auch sozio-ökonomische Faktoren wie das Elternhaus, die schulische Ausbildung der Eltern, die Zahl der im Haus vorhandenen Bücher usw.

2) Kriteriumsvariable sind die gemessenen Leistungen, die in diesem Zusammenhang möglichst normal verteilt sein sollen (zumeist in Form der Gauss'schen Verteilungskurve). Sie werden auch abhängige Variable genannt, im Gegensatz zu den unabhängigen Variablen.

berücksichtigten Eingangsvariablen bei allen Populationen ein hoher Prozentsatz der Varianz ungeklärt.

## 3. Die BRD im internationalen Vergleich

### 3.1. Unterschiede in den Erziehungssystemen

3.1.1. Die Entwicklungsländer, Chile, Iran, Thailand fallen in vieler Hinsicht, vor allem auch in den erzielten Leistungen, so stark aus der Gruppe der übrigen Länder heraus, daß sie zumeist einer Sonderdarstellung bedürfen. Ihre Leistungen liegen im Durchschnitt um eine Standardabweichung niedriger als die der entwickelten Länder. Sie werden hier für die weitere Darstellung ausgeschlossen.

3.1.2. Die Dauer der Schulpflicht schwankt in den übrigen Ländern zwischen 8 und 10 Jahren, was naturgemäß nicht ohne Einfluß auf die Lehrpläne in Naturwissenschaften sein kann. Das Lebensalter, das die Schüler der Pop.IV zur Zeit der Testung hatten, reicht von 17 Jahren (Australien, Japan, Neuseeland, Schottland) bis zu 20 Jahren (Niederlande). Das Durchschnittsalter in der Bundesrepublik betrug 18 Jahre.

3.1.3. Der Anteil der Schüler eines Jahrgangs im letzten Schuljahr der Sekundarstufe II (Population IV) schwankt zwischen 9% (Bundesrepublik) und 75% (USA). (Niederlande 13, Neuseeland 13, England 20, Frankreich 29, Schweden 45, Japan 70%). Diese Unterschiede erklären sich z.T. dadurch, daß unterschiedliche Zielgruppen in die Untersuchung einbezogen wurden; wir wiesen schon darauf hin, daß in der Bundesrepublik lediglich die Gymnasien herangezogen wurden, während in Schweden z.B. die gesamte gymnasiale Oberstufe, d.h. einschließlich der Fachschulen, einbezogen wurde. Andererseits schlägt sich in diesen Zahlen die unterschiedliche Schulorganisation nieder. Der verschiedene Grad an "Haltekraft" (holding power) wirkt sich auf die erreichten Durchschnittswerte aus.

3.1.4. Die verschiedenen Schultypen, Züge und Abschlüsse in einem gemeinsamen Kategorienschema zu erfassen, ist fast unmöglich, da die Unterschiede zu sehr in den Schultypen selbst, in ihrer historischen Entwicklung und ihren Reformen verwurzelt sind. Man hat versucht, die 'Programme' nach folgenden Kategorien zu unterscheiden:

Kategorie O : ohne bestimmten Schwerpunkt

Kategorie I-III : allgemeine Kurse (General Courses im amerikanischen Sinne)

Kategorie IV-VI : berufsbildende Kurse

Kategorie VII-IX: "akademische"Kurse (auf die Hochschule vorbereitend).

Länder wie Japan, Schweden und Finnland waren nicht in der Lage, Kurse oder Typen nach diesem Kategoriensystem zu beschreiben, während in der BRD mindestens für die Population IV mit den Kategorien VII-IX die altsprachlichen, neusprachlichen und naturwissenschaftlichen Typen bzw. Züge erfaßt werden konnten. Die uneinheitliche Kategorisierung wirkte sich dahingehend aus, daß für die Regressionsanalysen für jedes Land besondere Typen bzw. Ordnungssysteme verwendet werden mußten.

3.1.5. Stellung des naturwissenschaftlichen Unterrichts im Lehrplan

Drei Fragen müssen in diesem Zusammenhang beantwortet werden:

a) Wann setzt der systematische Unterricht in den Naturwissenschaften ein?

b) Wie sind die Kurse aufgebaut?

c) Wann und in welcher Form erfolgt eine besondere Förderung mathematisch-naturwissenschaftlich interessierter bzw. begabter Schüler?

In zehn der entwickelten Länder wird ein naturwissenschaftlicher Unterricht drei Jahre lang auf der Primarstufe erteilt, wenn auch der Unterricht im wesentlichen in einer Hand liegt, in fünf Ländern von Schulbeginn an. Die Bundesrepublik und Neuseeland indessen kennen keinen systematischen naturwissenschaftlichen Unterricht auf der Primarstufe. In diesen Ländern ist es natürlich schwer, den Umfang des naturwissenschaftlichen Unterrichts im Zusammenhang des heimatlichen Gesamtunterrichts zu bestimmen.

In den Sekundarschulen fast aller Länder wird zunächst ein Jahr lang (z.T. 2 Jahre) ein nicht nach Fächern differenzierter naturwissenschaftlicher Unterricht (general science) gegeben. Dabei nimmt der naturwissenschaftliche Unterricht in der Population II zwischen 6 bis 10% (in Frankreich) und 16-24% (Finnland) der gesamten Unterrichtszeit ein. In Population IV ist dieser Anteil in den einzelnen Zweigen bzw. Zügen sehr unterschiedlich und nimmt bei entsprechender Wahl seitens der Schüler bis zu 50% der Unterrichtszeit (England) ein. Dabei kann man in der Population IV zwei Gruppen von Ländern unterscheiden: Länder mit hohen und Länder mit niedrigen Prozentanteilen von Schülern für den naturwissenschaftlichen Unterricht. Länder mit niedrigem Anteil (England, Frankreich, Italien und

Schweden) sind dadurch charakterisiert, daß Schüler sich mit 15 oder 16 Jahren nach naturwissenschaftlichen, sozialwissenschaftlichen, humanwissenschaftlichen Zweigen bzw. Kursen spezialisieren können. Länder mit einem hohen Anteil von Schülern, die naturwissenschaftlichen Unterricht erhalten, sind dadurch charakterisiert, daß in ihnen wenig oder sehr spät differenziert wird, so daß eine völlige Aufgabe der Naturwissenschaften erst kurz vor dem Ende der Sekundarstufe II möglich ist (BRD). Eine Ausnahme stellt der naturwissenschaftliche Unterricht in den USA dar, wo in den Klassen 7 bis 9 ein allgemeiner naturwissenschaftlicher Unterricht (general science) erteilt wird, während in Klasse 10 Biologie, in Klasse 11 Chemie und Klasse 12 Physik angeboten wird, wovon aber nur die Schüler Gebrauch machen müssen, die zwei Credits in Naturwissenschaften für die Zulassung zum College benötigen.

Es ist interessant, daß in der Bundesrepublik wie in Finnland im Rahmen des Kursangebotes in unverhältnismäßig starkem Maße Biologie gewählt wird, was sich dann auch eindeutig in den Leistungen niederschlägt.

3.1.6. Wenn auch generell ein Einfluß der Klassengröße auf die Leistung nicht festgestellt werden konnte (die Bundesrepublik ist eine Ausnahme), so interessiert doch das Lehrer/Schüler-Verhältnis in den einzelnen Populationen. Es ist in Population I in der Bundesrepublik mit 33,2 am höchsten, mit 17,1 in Schweden am niedrigsten. In der Population II liegt die Bundesrepublik ebenfalls mit 27,4 Schülern je Lehrer an der Spitze; Schweden hat ein Verhältnis von 14,2, das französischsprachige Belgien sogar von 11,6. In der Population IV hat die Bundesrepublik einen Mittelwert von 19,9 und wird von den Vereinigten Staaten mit 20,1 und Finnland mit 22,3 übertroffen.

So bestätigt die Erhebung die im Grunde bekannte Tatsache, daß die Bundesrepublik hinsichtlich des Lehrer-Schüler-Verhältnisses am unteren Ende der Skala der industrialisierten Länder steht.

3.2. Schulvariablen

Sie betreffen im wesentlichen das Alter, das Geschlecht und den sozio-kulturellen Lebenskreis, aus dem die Schüler einer Schule kommen.

3.2.1. In der Population I betrug das durchschnittliche Alter in der BRD 125 Monate (3,5)[1], der Anteil der männlichen Schüler 51%.

In der Population II betrug das Durchschnittsalter 173 Monate (3,8), wobei 54% männliche Schüler erfaßt

---

1) In Klammern sind jeweils die Standardabweichungen angegeben.

wurden.

In der Population IV betrug das Durchschnittsalter 231 Monate (10,3), der Anteil der männlichen Schüler 65%.

3.2.2. Über die sozio-kulturelle Herkunft der Schüler gibt Tabelle 5 Auskunft, was den Beruf des Vaters betrifft.

Tabelle 5: Die deutsche Population nach den Berufen der Väter (in %)

| Population | I | II | IV |
|---|---|---|---|
| 1. Akademische und leitende Berufe | 13 | 14 | 49 |
| 2. Angestellte, Handel und Militär | 27 | 32 | 38 |
| 3. Gelernte Arbeiter | 49 | 38 | 7 |
| 4. Ungelernte und Angelernte | 7 | 8 | 1 |
| 5. Ohne Angaben | 4 | 8 | 6 |

Vergleiche mit anderen Ländern sind sehr schwierig, weil die Gesamtzusammensetzung der Bevölkerung hinsichtlich der Berufe nicht bekannt ist. Sie müßte eigentlich bei hinreichender Repräsentanz in der Zusammensetzung der Population I und II deutlich zum Ausdruck kommen. Unsere Tabelle zeigt gewisse Verschiebungen zwischen diesen beiden Populationen in der 2. und 3. Berufsgruppe. Dabei kann nicht angenommen werden, daß diese Verschiebungen die Folge einer Entwicklung in den zwischen den Populationen I und II liegenden vier Jahren sind. Außerdem würde eine in diesen Zahlen zum Ausdruck kommende Entwicklung vom Angestellten zum gelernten Arbeiter der bekannten Tendenz widersprechen. Im Vergleich zu den Zahlen des statistischen Jahrbuches 1973 dürfte die Pop.II in stärkerem Maße für die Berufe der Väter repräsentativ sein als die Pop. I. Das Statistische Jahrbuch der BRD 1973 gibt für April 1971 folgende Verteilung der Erwerbspersonen und Erwerbstätigen nach ihrer Stellung im Beruf:

Selbständige 10,2%, Beamte 5,6%, Angestellte 31,2%, Arbeiter 46,7%, mithelfende Familienangehörige 6,3%.

Deutlich kommt die bekannte Tatsache zum Ausdruck, daß in den Gymnasien eine soziale Selektion stattfindet. Das schlägt sich sowohl in dem hohen Prozentsatz der akademischen und leitenden Berufe wie in der geringen Zahl der Arbeiterkinder nieder. Das Gleiche drückt sich in der für die Population II und IV erfaßten

schulischen Ausbildungsdauer des Vaters und der Mutter aus. Während sie für die Population II, die also in etwa repräsentativ für den Jahrgang sein dürfte, beim Vater 9,1 (3,2) Jahre, für die Mutter 8,4 (2,4) Jahre beträgt, steigt sie für die Population IV für den Vater auf 11,9 (4,1) und für die Mutter auf 10,0 (3,3) Jahre. Dabei ist die Dauer der Ausbildung des Vaters in der Population IV abgesehen von den USA die höchste aller beteiligten Länder (USA 12,1).

3.2.3. Die Angaben im Schülerfragebogen zu verschiedenen Variablen wie Lesegewohnheiten, Fernsehen, Stunden, die für Hausarbeiten verwendet werden, eigene Arbeiten im Bereich der Naturwissenschaften usw., waren im Grunde genommen wenig ergiebig. Eindeutig geht jedoch aus den Antworten hervor, daß in den Schulen der englisch-sprechenden Nationen sehr viel mehr praktische Laboratoriumsarbeiten in den Naturwissenschaften üblich sind, während in den übrigen Ländern Europas eine mehr theoretische Gestaltung des Unterrichts vorherrscht.

Die Erfassung brauchbarer Daten über Formen und Art des Unterrichts, des methodischen Zugangs usw. mit Hilfe von Schüler- und Lehrerfragebogen zu gewinnen, hat sich doch als sehr viel schwieriger erwiesen als man glaubte. Nicht alle erfragten Eingangsvariablen erwiesen sich als brauchbar für die weitere Verarbeitung. Näheres s. 3.5.

3.3. Die Tests und Verhaltensskalen, statistische Werte.

Die Zuverlässigkeit der Tests[1] in den drei Populationen beträgt 0,82. Die entsprechenden Werte für die Bundesrepublik sind 0,79; 0,84; 0,81, sind also durchaus ausreichend für Analysen der Unterschiede auch zwischen den Schülern. Die Teiltests für die einzelnen Fachbereiche haben infolge der geringeren Zahl von Aufgaben naturgemäß eine geringere Zuverlässigkeit. Sie bewegt sich im allgemeinen um den Wert 0,6 und beträgt für die Bundesrepublik in Population I für Chemie und Erdkunde 0,67, für Biologie 0,52, für Physik 0,61. In Population II beträgt sie für Biologie 0,44, für Chemie 0,60, für Physik 0,71, für praktische Fähigkeiten 0,59. In der Population IV beträgt sie für Biologie 0,39, für Chemie 0,59, für Physik 0,70 und für praktische Fähigkeiten 0,51, ist also für Biologie in beiden Populationen verhältnismäßig niedrig.

Die Zuverlässigkeit der nicht kognitiven Meßinstrumente ist im allgemeinen gering; sie sind deshalb für eine weitere Verarbeitung unzureichend. Lediglich die Skalen betreffend das Verhältnis zur Schule, das Interesse für Naturwissenschaften sowie für Motivation (nur für Population I) geben einigermaßen brauchbare

1) Kuder-Richardson 20

Werte. Anstelle der Motivation wurde in den Populationen II und IV nach der Bedeutung der Naturwissenschaften in der gegenwärtigen Welt gefragt, auch die Ergebnisse kommen für eine Auswertung in Frage. Die Zuverlässigkeit liegt zwischen 0,61 und 0,76; die Zuverlässigkeit der Schulmotivationsskala ist geringer, sie beträgt 0,46.

Die Korrelationen zwischen den Teiltests und dem Gesamttest:

Die Korrelationen wurden getrennt für jede Population für jedes Land berechnet, die Mittelwerte sind aus der Tabelle 6 ersichtlich.

Tabelle 6: Seite 18

Die Ergebnisse zeigen, daß die Teiltests tatsächlich verschiedene Dinge messen.

In bezug auf weitere statistische Angaben zu den Meßinstrumenten muß auf die Veröffentlichung Comber/ Keeves verwiesen werden. Im Anhang werden alle erforderlichen statistischen Daten, insbesondere auch über den Schwierigkeitsgrad der einzelnen Items sowie über die Wahl der "Distractors"(der verschiedenen Auswahlantworten) gegeben. Doch soll bereits hier erwähnt werden, daß die Aufgaben grundsätzlich zu schwer waren. Die erreichten Durchschnittswerte liegen international in allen Populationen unter der Hälfte der möglichen Punktzahlen. So wurden international in der Population I im Durchschnitt 16,7 Aufgaben (bei 40 möglichen), in der Population II und IV ein noch geringerer Prozentsatz gelöst. Selbst die international zweithöchste Leistung der Bundesrepublik in der Population IV liegt mit 26,9 Punkten noch unter der Hälfte der möglichen Lösungen. Wenn in Population II von 80 möglichen Lösungen im internationalen Durchschnitt nur 22,3 (11,8)[1] Aufgaben gelöst wurden, so ist das für 15 industrialisierte Länder mit hochentwickelten Schulsystemen für den Test ein insgesamt nicht sehr günstiges Resultat. Trotzdem verteilte sich die Anzahl der gelösten Aufgaben der Gaußschen Kurve entsprechend. Der Schiefheitsgrad für den Gesamttest streut bei der Population II zwischen 0,1 und 0,8, für die BRD beträgt er 0,7.

---

1) ( ) = Standardabweichung

Tabelle 6: Korrelationen zwischen dem Gesamttest und den Teiltests (r=100)

| Sachbereiche | Population I 1 | 2 | 3 | 4 | | Population II 1 | 2 | 3 | 4 | 5 | Population IV 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Gesamttest | - | 84 | 84 | 84 | | - | 73 | 77 | 85 | 76 | - | 71 | 75 | 84 | 74 |
| 2. Biologie | 84 | - | 59 | 54 | | 73 | - | 43 | 51 | 40 | 71 | - | 41 | 45 | 38 |
| 3. Chemie | 84 | 59 | - | 56 | | 77 | 43 | - | 51 | 47 | 75 | 41 | - | 52 | 41 |
| 4. Physik | 84 | 54 | 56 | - | | 85 | 51 | 51 | - | 55 | 84 | 45 | 52 | - | 53 |
| 5. Prakt.Fähigk. | - | - | - | - | | 76 | 40 | 47 | 55 | - | 74 | 38 | 41 | 53 | - |
| **Leistungskategorien** | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 1. Gesamttest | - | 80 | 87 | 81 | 66 | - | 73 | 84 | 71 | 65 | - | 60 | 88 | 72 | 53 |
| 2. Kenntnisse A | 80 | - | 57 | 52 | 41 | 73 | - | 52 | 41 | 41 | 60 | - | 43 | 34 | 25 |
| 3. Verständnis B | 87 | 57 | - | 61 | 46 | 84 | 52 | - | 50 | 45 | 88 | 43 | - | 51 | 39 |
| 4. Anwendung C | 81 | 52 | 61 | - | 43 | 71 | 41 | 50 | - | 40 | 72 | 34 | 51 | - | 26 |
| 5. Höhere Denkprozesse D | 66 | 41 | 46 | 43 | - | 65 | 41 | 45 | 40 | - | 53 | 25 | 39 | 26 | - |

Quelle: Comber/Keeves, Tabelle 5.3

## 3.4. Allgemeine Analyse der Schülerleistungen

### 3.4.1. Die Ergebnisse nach Fächern und Leistungskategorien; die Stellung der Bundesrepublik in der Gesamtuntersuchung

Im folgenden werden zunächst die Ergebnisse nach den einzelnen Fächern getrennt untersucht. Darüber hinaus werden die Ergebnisse in bezug auf die in Anlehnung an die Bloom'sche Taxonomie aufgestellten Leistungskategorien analysiert.

Zum Zwecke des Vergleichs wurden die Rohwerte in Standardwerte umgerechnet, wobei der Durchschnitt der ungewichteten Mittelwerte der Schüler in den industrialisierten Ländern als O-Punkt diente. Die mittlere internationale Standardabweichung wurde aus den Mittelwerten der nationalen Standardabweichungen errechnet und diente als Maßeinheit. Tabelle 7 gibt die gewichteten und korrigierten durchschnittlichen Standardwerte für die Bundesrepublik.

Tabelle 7: Durchschnittswerte für die Untertests (gewichtet und korrigiert) für die BRD nach Fachbereichen und Leistungskategorien (in Standardwerten)

| B e r e i c h e | Pop. I | Pop. II | Pop. IV |
|---|---|---|---|
| Biologie | -0,31 | 0,13 | 0,80 |
| Chemie | -0,16 | -0,15 | 0,23 |
| Physik | -0,10 | 0,22 | 0,52 |
| Prakt.Fähigk. | - | 0,16 | 0,37 |
| **Leistungskategorien** | **Pop. I** | **Pop. II** | **Pop. IV** |
| Kenntnisse A | -0,26 | 0,14 | 0,50 |
| Verständnis B | -0,14 | -0,02 | 0,49 |
| Anwendung C | -0,14 | 0,15 | 0,50 |
| Höhere Prozesse D | -0,19 | 0,06 | 0,47 |
| N | 1742 | 2231 | 1989 |

Quelle: Comber/Keeves, Tab.6.1-6.3

Anmerkung: Korrigierte Werte wegen der Möglichkeit des Ratens bei Auswahlantworten.

Die Tabelle 7 zeigt, daß in der Population I sämtliche Leistungen unter internationalem Durchschnitt liegen, auffälligerweise ist das in besonderem Maße für Biologie der Fall, obgleich man doch annehmen sollte, daß gerade biologische Betrachtungen innerhalb der Heimatkunde angemessene Berücksichtigung finden. Die sehr viel höheren Werte anderer Länder, besonders Japans, aber auch Ungarns, Belgiens und Schwedens sind wahrscheinlich auf das frühzeitige Einsetzen eines Fachunterrichts zurückzuführen. Eindeutig trifft das für Japan zu, wo alle Primarlehrer in den Naturwissenschaften besonders ausgebildet werden.

So ist denn verständlich, daß auch in den Leistungskategorien sämtliche Werte unter dem Mittelwert bleiben, besonders aber auch der Wert für Kenntnisse.

In der Population II liegt die Bundesrepublik bis auf Chemie über den internationalen Mittelwerten, in Physik stärker als in Biologie. Lediglich in Chemie wird der internationale Durchschnitt bei uns nicht erreicht, ein eindeutiges Ergebnis der Stellung der Chemie in unseren Lehrplänen. Japan und Ungarn überraschen im internationalen Vergleich wiederum mit erstaunlich guten Ergebnissen, während Schweden und die USA in dieser Population unter den Durchschnitt zurückfallen. Japan fällt nicht nur durch seine hohen Leistungen insgesamt auf, sondern genau wie in Population I durch relativ hohe Werte für "Verständnis" und "Höheren Denkprozeß". Die BRD liegt hier ungefähr im internationalen Durchschnitt. Auffällig, daß auch die übrigen europäischen Länder (und die USA) nur unerheblich um den internationalen Mittelwert streuen; wahrscheinlich sind die traditionellen europäischen Formen des Naturkundeunterrichts einigermaßen einheitlich. In Population IV ergibt sich für die Bundesrepublik ein auffällig hoher Leistungsstand in Biologie, aber auch in Physik liegt sie noch eine halbe Standardabweichung über dem internationalen Mittelwert. Die Leistungen in Biologie und Chemie gehören zu den höchsten international auftretenden Leistungen überhaupt. Das ist verständlich, da die Bundesrepublik zugleich das Land mit der stärksten Auslese ist, Länder wie Schweden und die USA als Länder mit hohem Schüleranteil in den Abschlußklassen der Sekundarstufe II (USA 75%, Schweden 40%) haben natürlich eine entsprechend niedrigere Durchschnittsleistung. In den Leistungskategorien steht die Bundesrepublik ebenfalls sehr hoch, im Durchschnitt eine halbe Standardabweichung über den internationalen Mittelwerten. Auffällig niedrige Werte haben wieder die Länder mit Gesamtschulsystemen; aber auch die französisch sprechenden Länder liegen trotz ihrer selektiven Systeme

in der Gesamtleistung verhältnismäßig niedrig. Frankreich fällt lediglich durch das gute Abschneiden in den höheren Denkprozessen auf.

3.4.2. Die Abhängigkeit der Leistungen von den Lehrplänen

Die Ergebnisse in den einzelnen Fächern hängen naturgemäß stark von den konkreten Inhalten der Lehrpläne und damit von der Gelegenheit der Schüler ab, den verschiedenen Aufgaben im Unterricht begegnet zu sein. Diese Möglichkeit wurde durch die Lehrer im Lehrerfragebogen in Prozenten für jede Aufgabe geschätzt. Es ist anzunehmen, daß die Leistungen in hohem Grade mit dieser Möglichkeit korrelieren, wobei wiederum anzunehmen ist, daß dies in besonderem Maße im Bereich der Kenntnisse der Fall sein sollte. Man rechnete zur Ermittlung dieses Zusammenhangs die Angaben der Lehrer auf eine Fünf-Punkte-Skala um und erreichte damit für jede Schule einen Zahlenindex für jede Aufgabe, die man mit den internationalen Leistungen korrelieren konnte. Dabei ergab sich erstaunlicherweise in sieben Ländern, daß keine eindeutigen Ergebnisse herauskamen, nur in zwei Ländern wurden brauchbare Korrelationen ermittelt. (s. unter 3.4.4.1, S.24.)

3.4.3. Unterschiede zwischen den Geschlechtern

Aufgrund bisheriger Untersuchungen scheinen große Unterschiede zwischen den Geschlechtern in bezug auf ihre Leistungen in einzelnen Schulfächern zu bestehen. Jetzt ergab sich zum ersten Mal die Möglichkeit, die bisherigen Untersuchungen in größerem Ausmaß für die Naturwissenschaften zu überprüfen.

Der Quotient von Knaben und Mädchen in den drei Populationen beträgt für die Bundesrepublik 1,05-1,07-1,59. Die Berechnung der Unterschiede erfolgte, indem die Differenz zwischen den Leistungen der Knaben und Mädchen durch die mittlere Standardabweichung für alle Länder geteilt und mit 100 multipliziert wurde. Es ergab sich, daß in Population I in den meisten Ländern eine auffällige Konformität insofern besteht, als Knaben jeweils 1/4 einer Standardabweichung mehr leisten als die Mädchen. In der Bundesrepublik hingegen sind die Leistungen der Knaben wie in Finnland und in den Niederlanden wesentlich höher, nämlich 1/3 einer Standardabweichung (33). Ähnlich ist es in der Population II. Während hier die Leistungen der Knaben die der Mädchen im internationalen Vergleich etwa um eine halbe Standardabweichung übertreffen, sind es in der Bundesrepublik fast zwei Drittel einer Standardabweichung (62); in Japan ist diese Differenz noch größer (74). In Population IV vergrößern sich im allgemeinen die Unterschiede. Sie be-

tragen in der Bundesrepublik 3/4 einer Standardabweichung (75), sind jedoch in anderen Ländern z.T. noch erheblich größer. Ob dieses international relativ konstante Ergebnis an der Form des Unterrichts bzw. an den Inhalten der Lehrpläne liegt oder aber geschlechtsbedingt ist, ist nicht eindeutig festzustellen. Geht man den Unterschieden in den Teiltests nach, so ergibt sich folgendes Bild:

Tabelle 8: Leistungsunterschiede der Geschlechter in den Teiltests und Leistungskategorien in der BRD (in Standardwerten)

| | Pop. I | Pop. II | Pop. IV |
|---|---|---|---|
| Biologie | 12 | 15 | 38 |
| Chemie | 21 | 32 | 35 |
| Physik | 53 | 77 | 88 |
| Prakt.Fähigk. | - | 60 | 66 |
| Kenntnisse A | 21 | 30 | 41 |
| Verständnis B | 35 | 63 | 64 |
| Anwendung C | 20 | 31 | 49 |
| Höhere Prozesse D | 27 | 32 | 38 |

Quelle: Comber/Keeves, Tab. 6.15-6.17

Das Ergebnis ist eindeutig: Die Leistungsunterschiede werden größer von Population zu Population in sämtlichen Teiltests. Sie sind am größten in Physik und am geringsten in Biologie. Auffällig sind auch die Unterschiede im Bereich des Verständnisses und in den praktischen Fähigkeiten.

Diese Unterschiede sind erstaunlich; sie finden sich bereits in Pop.I, wo der Einfluß der Schule auf den Erwerb naturwissenschaftlicher Kenntnisse gering ist und obgleich in den Lehrplänen kaum Unterschiede in der Erziehung der Geschlechter bestehen. Diese unterschiedliche Leistung spiegelt sich naturgemäß in den Angaben über das Interesse bei den Mädchen und Knaben wider. Auf alle Fälle dürfte hier im Hinblick auf die Rolle der Naturwissenschaften in der Mädchenbildung ein wesentliches Problem aufgeworfen worden sein, das nicht nur auf die BRD beschränkt ist.

Der Einfluß gemeinsamer Erziehung von Knaben und Mädchen:

Bereits in der Mathematikerhebung ergaben sich Leistungsunterschiede zwischen Knaben und Mädchen in Kodukationsschulen und eingeschlechtlichen Schulen, die

von den allgemeinen Unterschieden zwischen Knaben und Mädchen abwichen. Da bei den Zehnjährigen heute allgemein Koedukation besteht, konnten soche Unterschiede lediglich für die Population II und IV ermittelt werden: In Population II ist die Leistung der Knaben in eingeschlechtlichen Schulen um sehr viel mehr besser als in den Koedukationsschulen (in 8 von 12 Ländern). In der Bundesrepublik ist es umgekehrt, der Unterschied beträgt 59 in Koedukationsschulen gegenüber 35 in gleichgeschlechtlichen Schulen. Bei Population IV hingegen bestehen überall größere Unterschiede zwischen Knaben und Mädchen in Koedukationsschulen als in gleichgeschlechtlichen Schulen: In der Bundesrepublik sogar 97 gegenüber 47. Die Tendenz in der Bundesrepublik ist also einheitlich. Die Gründe dafür sind allerdings nicht ersichtlich. Sie können jedenfalls aus den der Untersuchung zugrunde liegenden Daten nicht erschlossen werden. Wir können vielleicht vermuten, daß bei dem großen Übergewicht männlicher Lehrer in den Naturwissenschaften, das mit größeren Erfolgen verbunden ist, die Anlage des Unterrichts in Koedukationsschulen stärker auf die Interessen der Jungen als die der Mädchen ausgerichtet ist, während sich die Lehrerinnen in reinen Mädchenschulen stärker auf die Interessen und Bedürfnisse der Mädchen einstellen.

### 3.4.4. Leistungsunterschiede zwischen den Ländern und die ihnen zugrunde liegenden Faktoren

In bezug auf diese Fragestellung glaubt die Untersuchung einen besonderen Beitrag zur Vergleichenden Erziehungswissenschaft leisten zu können, da hier mit Hilfe statistischer Verfahren neue Einsichten über das Zusammenwirken von Faktoren am Unterrichtserfolg zu gewinnen sind.

#### 3.4.4.1. Die BRD im Vergleich der absoluten Meßwerte

Der Versuch, sozio-ökonomische Unterschiede zwischen den industrialisierten Ländern durch eine Reihe von Daten, die im Länderfragebogen ermittelt werden sollten, quantitativ so zu erfassen, daß sie als Variable in die statistische Auswertung eingehen können, ist nicht gelungen. Deshalb beschränkt sich die Untersuchung auf die Herstellung von Beziehungen zwischen sozio-kulturellen Faktoren und der Leistung. Vor der Darstellung der Leistungsunterschiede soll jedoch ausdrücklich darauf hingewiesen werden, daß es den Beteiligten nicht darauf ankam, Vergleiche im Sinne einer "Olympiade" durchzuführen, sondern vielmehr hoffte man, den spezifischen Zusammenhang bestimmter Faktoren mit der Leistung in besserer Weise zu erfassen als es bei Studien, die innerhalb eines Landes durchgeführt wurden, möglich ist.

Bei allen Vergleichen müssen die Leistungen im Zusammenhang zweier Variablen beurteilt werden:

a) der Gelegenheit, den Inhalten der Aufgaben im Unterricht begegnet zu sein, und

b) dem Prozentanteil der Schüler eines Jahrgangs, der sich am Ende der Sekundarstufe II in den Schulen befindet (holding power).

Diese Werte werden einander in Tabelle 9 zugeordnet.

Tabelle 9: Korrigierte Punktwerte für die BRD, Gelegenheit zu lernen und Prozentsatz der in den Schulen befindlichen Schüler

| | Pop.I | i.M.[1] | Pop.II | i.M. | Pop. IV | i.M. |
|---|---|---|---|---|---|---|
| Punktzahl | 14,9 (7,4) | 16,7 (7,9) | 23,7 (11,5) | 22,3 (11,8) | 26,9 (8,9) | 20,9 (9,9) |
| Lernmöglichkeit[2] | 1,67 | 2,06 | 1,82 | 1,99 | 3,15 | 2,56 |
| Haltekraft[3] | 99 | | 94 | | 9 | |

Quelle: Comber/Keeves, Tab. 7.2

1) i.M. = internationales Mittel

2) Die Prozentangaben der Lehrer werden in 4 Werten ausgedrückt (0-25, 26-50, 51-75, 76-100%) und gemittelt. Höchstwert wäre also =4

3) Holding power, ausgedrückt als Prozentanteil eines Jahrgangs, der sich in den herangezogenen Schulen befand.

Wie bereits erwähnt, waren die Aufgaben grundsätzlich zu schwer. Die internationalen Durchschnittswerte (i.M.) für alle Populationen liegen weit unter der Hälfte der möglichen Punktzahl. Selbst die höchsten nationalen Durchschnittswerte liegen noch unter der Hälfte der möglichen Punktzahl (bis auf Japan in Population I).

Die Lernmöglichkeit liegt im internationalen Durchschnitt für Pop. I und II etwa bei 2, steigt allerdings für Population IV auf 2,56. Für die Bundesrepublik wird sie von den Lehrern für die Populationen I und II geringer eingeschätzt, erreicht in Population IV allerdings mit 3,15 unter allen Ländern den höchsten Grad. Das bedeutet nach

Meinung der Lehrer, daß die Schüler der BRD bei mehr als 75% aller Aufgaben die Möglichkeit hatten, den Sachverhalt vorher im Unterricht kennenzulernen. Stellt man Durchschnittsleistungen der Beteiligten in ihrer Abhängigkeit von der Lernmöglichkeit graphisch dar, wie es in Figur 7.1. (S.161) in der Veröffentlichung von Comber und Keeves geschehen ist, ergeben sich für die Population I keinerlei Zusammenhänge. In Population II hingegen besteht ein deutlicher positiver Zusammenhang für die Länder, in denen Naturwissenschaften systematisch als Fach unterrichtet werden. In der Population IV ist dieser Zusammenhang weniger deutlich. Es bilden sich zwei Ländergruppen heraus, bei denen jeweils dieser Zusammenhang deutlich heraustritt, ohne daß Gründe im einzelnen ersichtlich werden. Der Zusammenhang zwischen dem, was in der Schule gelehrt wird, und der Leistung muß fernerhin gesehen werden im Hinblick auf die Zeit, die den Naturwissenschaften im Lehrplan gewidmet ist, und zwar nicht nur im Jahr der Erhebung, sondern auch in den vorhergehenden Jahren. Dieses Gewicht kommt sicherlich in den Schätzungen der Lehrer nur unzureichend zum Ausdruck. Der Prozentsatz derjenigen Schüler eines Jahrgangs, die die Schule zur Zeit der Untersuchung noch besuchen (holding power) beeinflußt die Leistungen naturgemäß in noch stärkerem Maße, zumal, wenn es sich um Ausleseschulen handelt. In den entwickelten Ländern sind in der Population II noch annähernd 100% aller Schüler in der Schule (Ausnahmen: Ungarn und Italien, wo eine Schulpflicht nur bis zum 14. Lebensjahr besteht), in Population IV bestehen jedoch gewaltige Unterschiede. Wie schon erwähnt, betragen die Extremwerte 9% (BRD) und 75 % (USA). So ist es nicht verwunderlich, wenn die BRD mit 26,9 Punkten (nach Neuseeland) an der Spitze der Leistungsrangordnung steht, während die USA mit 13,7 Punkten am Ende derselben liegen. Die BRD hat also bei einem Vergleich der absoluten Zahlen außer dem Vorteil, daß der größte Anteil an Schulen Gelegenheit hatte, die herangezogenen Inhalte einmal zu lernen, den Vorteil, daß die Leistung durch eine hochausgelesene kleine Elite erzielt wurde. Tatsächlich sind diese Vergleiche also wenig aussagefähig.

Ehe wir gültigere Formen des Vergleichs darstellen, wollen wir einen Blick auf die Streuung der Leistung werfen.

### 3.4.4.2. Die Streuung der Leistungen

Die Länder unterscheiden sich erheblich hinsichtlich ihrer Variabilität. Sie entspricht in der BRD in den Populationen I und II in etwa dem internationalen Durchschnitt; die Standardabweichung beträgt in Population IV jedoch nur 8,9 gegenüber

9,9 im internationalen Durchschnitt (s.Tab.9). Das konnte wohl auch erwartet werden, da es sich um eine hochausgelesene Gruppe handelt. Eine Analyse dieser Varianz brachte international folgende Ergebnisse:

1. Es besteht ein ausgesprochener Zusammenhang zwischen den Durchschnittsleistungen und den Standardabweichungen, die Rangkorrelation r= 0,86. (Die Schiefe der Verteilung läßt den Schluß zu, daß die Leistungen keineswegs am unteren Ende der Verteilungsskala massiert auftreten.)
2. Es besteht ein starker Zusammenhang zwischen den Standardabweichungen und dem Grad der Homogenität der Tests, wie sie durch eine biseriale Korrelation ermittelt wurde. Die Korrelation zwischen diesen Variablen beträgt 0,94 d.h. also, je größer die Homogenität eines Tests in einem Land, desto größer ist die entsprechende Standardabweichung.
3. Da die Testergebnisse allgemein sehr niedrig liegen, sind in vielen Ländern viele Aufgaben unter die 20%-Grenze des Schwierigkeitsgrades gefallen. 20% richtiger Antworten werden jedoch bei 5 Auswahlantworten bereits durch bloßes Raten erreicht. Die Rangkorrelation zwischen der Standardabweichung und der Zahl schwieriger Aufgaben beträgt tatsächlich 0,91, d.h., je größer die Zahl der schwierigen Aufgaben, desto kleiner ist die Standardabweichung.
4. In bezug auf die Schulorganisation und andere im Bereiche des Unterrichts liegende Unterschiede zwischen den Ländern ließen sich keine klaren Ergebnisse gewinnen. Die Schulorganisation insbesondere wirkte sich in diesen Standardabweichungen nicht aus. Ebenfalls ließ sich kein Zusammenhang der Varianz mit den unterschiedlichen Möglichkeiten der Schüler, den einzelnen Aufgaben bereits begegnet zu sein, finden.

3.4.4.3. Der Leistungszuwachs wurde mit Hilfe von Anker-Aufgaben für die Populationen I und II bzw. II und IV in einem besonderen Berechnungsverfahren von einer Population zur nächsten ermittelt (Comber/Keeves, S. 167 - 173). Dieser Zuwachs konnte auf diese Weise in seiner relativen Größe auf einer für die Populationen gemeinsamen Skala abgelesen werden. Er beträgt für die BRD ausgedrückt in Standardwerten:

Pop. I → Pop. II 1.23 (-100/0.23),

Pop. II → Pop. IV 1.93 (0.23/2.16),

Pop. I → Pop. IV 3.16

Dieser Zuwachs muß natürlich im Zusammenhang mit dem Grad der Selektion in der Sekundarstufe II beurteilt werden. Er ist keinesfalls ausschließlich Ausdruck für die Güte des naturwissenschaftlichen Unterrichts. Die Korrelation des Zuwachses zur Selektivität der Systeme ist eindeutig, r = 0,76 (Pop.I bis IV) bzw. r = 0,77 (Pop.II bis Pop.IV). Ebenso eindeutig ist die Beziehung zu einem Berufsindex: sie beträgt von Pop. I zu Pop. IV r = 0,87.

### 3.4.5. Die Bundesrepublik im Vergleich der relativen Leistungen

Bei allen bisherigen Vergleichen stellte sich die unterschiedliche "Haltekraft" der Schulsysteme als Störfaktor ein. Dieser läßt sich nur umgehen, indem man Spitzengruppen in ihrer Leistungsfähigkeit vergleicht, die jedoch prozentual nicht größer sein dürfen als der Prozentsatz von Schülern einer Population in den Abschlußklassen der Sekundarstufe des am stärksten auslesenden Landes, d.h.also der BRD mit 9%. Die Frage lautet also, wie sieht es mit der Leistungsfähigkeit der besten 9% in den beteiligten Ländern aus? Mit anderen Worten: Bleiben Neuseeland und die Bundesrepublik in ihrer Spitzenstellung, wenn sie mit den 9% Besten der anderen Länder verglichen werden?

Um das eindeutige Ergebnis vorwegzunehmen: Ein geringerer Grad von Auslese führt keineswegs dazu, daß die Leistungen einer entsprechenden Spitzengruppe geringer werden, im Gegenteil, die BRD steht bei diesem Vergleich am unteren Ende der Rangskala. Tabelle 10 gibt die Werte für die 1% - 5%-und 9%-Gruppen.

Tabelle 10: Seite 28

Bei den oberen 9% fällt also die Bundesrepublik vom Rangplatz 2 mit dem französisch-sprachigen Belgien auf den vorletzten Platz. Nur Italien hat noch einen niedrigeren Rangplatz in dieser Spitzengruppe. Ein ganz ähnliches Resultat erhält man beim Vergleich der oberen 5 und 1%.

Im groben gesehen fallen die industrialisierten Länder in drei Leistungsgruppen, innerhalb welcher die Unterschiede bei den 1% und 5% Höchstleistungsfähigen gering sind. Die erste Gruppe umfaßt die Commonwealth-Länder: Australien, England, Neuseeland und Schottland mit Punktwerten über 50. Die zweite Gruppe setzt sich aus den nordeuropäischen Ländern und den USA zusammen, an deren Spitze mit 49,5 Punkten Schweden und an deren Ende mit 45 Punkten die BRD liegt. Die dritte Gruppe wird von den romanisch

Tabelle 10: Werte und Standardabweichungen für gleiche Prozentanteile der Altersgruppe (Pop.IV)

| L a n d | % i.d. Schule | N | ganze Stichprobe | beste 1% | beste 5% | beste 9% |
|---|---|---|---|---|---|---|
| Neuseeland | 13 | 1676 | 30,8(12,6) | 52,8(2,8) | 43,5(5,9) | 36,8(9,0) |
| England | 20 | 2181 | 24,4(12,4) | 51,6(3,2) | 41,6(6,5) | 35,5(8,5) |
| Australien | 29 | 4194 | 26,1(11,5) | 51,5(3,2) | 44,0(4,7) | 39,9(5,9) |
| Schottland | 17 | 1321 | 24,4(12,9) | 50,7(3,8) | 40,6(6,4) | 34,4(8,7) |
| Schweden | 45 | 2754 | 20,1(10,9) | 49,5(3,4) | 41,2(5,3) | 37,0(6,2) |
| Ungarn | 28 | 2828 | 24,0( 9,6) | 48,0(3,8) | 39,0(5,4) | 35,0(6,1) |
| Niederlande | 13 | 1138 | 24,4(12,0) | 47,1(3,6) | 37,2(6,5) | 30,3(9,4) |
| Finnland | 21 | 1725 | 20,8(10,5) | 46,0(4,1) | 35,7(6,4) | 30,7(7,4) |
| USA | 75 | 2514 | 14,2( 9,9) | 45,8(2,8) | 36,8(5,5) | 33,1(5,9) |
| BRD | 9 | 1989 | 28,4( 9,6) | 45,0(4,1) | 35,3(6,2) | 28,4(9,6) |
| Frankreich | 29 | 3523 | 19,1( 9,1) | 40,5(3,5) | 33,3(4,4) | 29,9(5,1) |
| Belgien (fr.) | 47 | 941 | 16,0( 8,3) | 36,2(2,0) | 30,9(3,1) | 28,4(3,7) |
| Italien | 16 | 15719 | 16,5( 9,2) | 38,2(4,7) | 27,4(6,5) | 22,7(7,3) |
| Belgien (fl.) | 47 | 467 | 18,1( 8,5) | 39,8(3,7) | 33,0(4,0) | 30,5(4,2) |
| Durchschnitt | | | 22,0 | 45,9 | 37,1 | 32,3 |

Quelle: Comber/Keeves, Tab.7.6

sprechenden Ländern mit Leistungen zwischen 36,2 (französisch sprechendes Belgien) und 40,5 (Frankreich) gebildet.

Es ergibt sich also wie bereits in der Mathematikstudie, daß hohe Selektivität keine entsprechend hohen Leistungen der ausgelesenen Gruppe garantiert, liegen doch Länder mit Gesamtschulsystemen wie die USA und Ungarn in den Spitzengruppen - trotz oder wegen ihrer Gesamtschulsysteme - leistungsmäßig sehr viel höher als z.B. die BRD. Wenn andererseits Länder mit selektiven Systemen (England, Neuseeland) ebenfalls innerhalb der Spitzengruppen liegen, läßt sich in bezug auf die Förderung von Hochbegabten in unterschiedlichen Systemen keine klare Aussage machen. Aber eines ist eindeutig: In gesamtschulartigen Systemen können Spitzenbegabungen zum mindesten in gleichem Maße gefördert werden wie in selektiven Systemen. Die Gründe hierfür sind also offensichtlich nicht in den Systemen als solchen, sondern in inneren schulorganisatorischen und Lehrplanfaktoren zu suchen.

Ein anderer Ansatz, die relative Leistungsfähigkeit von Schulsystemen zu vergleichen, ist gegeben, wenn man aufgrund festgelegter internationaler Standards ermittelt, welcher Anteil der Altersgruppe diese Standards erreicht. Es läßt sich dann vergleichen, wieviel Prozent eines Schülerjahrgangs in die betreffenden Perzentilränge fallen. Diese Untersuchung ist allerdings nur dann sinnvoll, wenn sich noch sämtliche Schüler oder gleiche Prozentanteile eines Jahrgangs zur Zeit der Untersuchung in der Schule befinden. Sie wird sinnlos, wenn die Zielpopulation unterschiedlich ausgewählt wurde, wie es in der vorliegenden Untersuchung für die Population IV der Fall war. Deshalb ist der Versuch, der in der Veröffentlichung von Comber und Keeves gemacht wird, den Prozentanteil in den einzelnen Perzentilrängen für die Population IV dadurch zu gewinnen, daß man eine Normalverteilung für die Gesamtpopulation zugrunde legt und daraus die Werte ableitet, unbrauchbar, zumal,wenn dabei für die Bundesrepublik die nicht zutreffende Annahme gemacht wird, daß alle nicht getesteten Schüler vorzeitige Abgänger sind und deshalb unter dem 25.Perzentil liegen. Für die Bundesrepublik ist das Ergebnis, wie es von Comber und Keeves in Tabelle 7.8 (S.180) dargestellt wird, nur insofern interessant, als es abermals ein Bild davon gibt, wie gering im Grunde die Leistung unserer hochselektiven Schulen in den Naturwissenschaften ist: Nur gut drei Viertel aller in der Population IV herangezogenen Schüler fallen international in den Perzentilrang 50.

Fassen wir das Ergebnis der Analyse der Schülerleistungen zusammen, so läßt sich feststellen, daß

1. in der BRD die Leistungen in der Population I in sämtlichen durch die Tests erfaßten Bereiche unter dem internationalen Durchschnitt liegen, am weitesten in Biologie,

2. in Population II die BRD mit Platz 5 (23,7 Punkte) der Rangordnung einen guten Mittelplatz hat, hinter Japan (31,2 Punkte), Ungarn (29,1 Punkte), Australien (24,6 Punkte) und Neuseeland (24,2 Punkte), wobei die Leistungen in Chemie unter dem internationalen Mittelwert liegen,

3. in Population IV die BRD an zweiter Stelle liegt (mit 26,9 Punkten hinter Neuseeland mit 29,0 Punkten), daß die BRD jedoch bei Berücksichtigung der Selektivität an das untere Ende der Rangskala abgleitet (auffällig gut sind die Leistungen in Biologie und Physik),

4. die Schulsysteme die Unterschiede in der Leistung nicht erklären. Länder mit Gesamtschulen haben einen gleich hohen Prozentsatz leistungsfähiger Schüler aufzuweisen wie Länder mit selektiven Systemen.

Betrachten wir im folgenden die Ergebnisse der Regressionsanalysen, um die Faktoren zu ermitteln, die für die Leistungsunterschiede maßgebend sind. Dabei werden zunächst die Unterschiede zwischen den Schulen untersucht.

## 3.5. Unterschiede zwischen den Schulen und die ihnen zugrunde liegenden Faktoren

Groß angelegte Untersuchungen wie das "Projekt Talent", der "Equality of Educational Opportunities Survey" sowie der "Plowden National Survey" haben immer wieder ergeben, daß der Einfluß der Schule auf die Schulleistungen gegenüber dem Gewicht außerschulischer Faktoren als verhältnismäßig gering angesehen werden muß, zum mindesten, was die Unterschiede zwischen den verschiedenen Schulen betraf. Deshalb sollte in dieser Untersuchung besonderes Gewicht auf die Erfassung von Faktoren gelegt werden, die aller Wahrscheinlichkeit nach von Bedeutung für den Einfluß der Schule und des Unterrichts auf die Leistung sind. Dabei wurde folgende Hypothese aufgestellt: Der Einfluß der Schule und der Lehrer auf die Leistungen in Fächern, die, wie z.B. die Naturwissenschaften, typische Schulfächer sind, wird größer sein, als in bisherigen Untersuchungen festgestellt wurde. Es wurde außerdem angenommen, daß der Einfluß auf den drei erfaßten Altersstufen unterschiedlich groß sei. Es durfte ferner angenommen werden, daß infolge des Klumpeneffekts der ge-

messene Einfluß der Schule auf die Leistungen größer ist, wenn die Schule und nicht der Lehrer als Einheit für die Untersuchung gewählt wird.

### 3.5.1. Das methodische Vorgehen

Die Prädiktoren wurden mit Hilfe dreier Instrumente erfaßt:

1. der Schulfragebogen, die von den Schulleitern ausgefüllt wurden,
2. der Lehrerfragebogen, die von allen Lehrern, die an einer Schule naturwissenschaftlichen Unterricht erteilten, ausgefüllt wurden,und schließlich
3. der Schülerfragebogen.

Dabei wurden aus diesen Daten Durchschnittswerte für die einzelnen Schulen ermittelt, wobei man hoffen konnte, daß in den Lehrer- und Schülerfragebogen sich nicht-systematische Fehler ausgleichen.

Die Analyse der Daten erfolgte mit Hilfe von Korrelationen und darauf folgenden Regressionsanalysen (s.o. S.11).

Da Regressionsanalysen mit der Varianz zwischen den Prädiktoren sowohl wie der Varianz der Kriterienvariablen arbeiten, ist es nur möglich, solche Prädiktoren für die Analyse heranzuziehen, die eine ausreichende Streuung aufzeigen. Das Kausalmodell, das der Analyse zugrunde liegt, verlangt ein schrittweises Fortschreiten von Regressionsanalyse zu Regressionsanalyse, wobei die Ergebnisse der ersten Analyse in die zweite eingehen. Als erster Schritt hat sich dabei aufgrund der Korrelationen der meßbare Unterschied in der dem Schuleinfluß vorausgehenden Gruppe von Faktoren (Home Background) empfohlen, in die auch das Alter und das Geschlecht der Schüler einzubeziehen sind. Im zweiten Schritt wurden dann die Art der Schule bzw. das vom Schüler verfolgte Lehrprogramm (Zweig, Kurs) in die Analyse eingebracht (Faktorengruppen 2 und 3). Viertens wurden schließlich restliche Variablen einbezogen, die sich als von Einfluß auf den gegenwärtigen Bildungsstand erwiesen haben (Faktorengruppe 4). Jeder dieser Schritte klärt einen bestimmten Anteil der Varianz und hinterläßt einen ungeklärten Rest, der mit jedem weiteren Schritt geringer wird, ohne jedoch völlig aufgehoben zu werden, da unmöglich alle für die Leistung maßgeblichen Faktoren in einer solchen Analyse erfaßt bzw. berücksichtigt werden können.

Hier kann im einzelnen nicht auf Verfahren eingegangen werden, die über den Weg der Korrelation jeder einzelnen erfaßten Variablen mit der Leistung zur Aufstellung der Faktorengruppen führten.

Die erste Gruppe von Variablen, von uns sozio-kultureller Schulindex genannt, ("school-handicap-score") ergibt einen Wert für jede Schule, in den die Berufe der Väter, ihre Bildungsgänge, die Bildungsgänge der Mütter, der Gebrauch von Wörterbüchern bzw. die Zahl der Bücher im Hause und die Familiengröße eingehen. Getrennt bearbeitet, aber zur Faktorengruppe gehörig, werden das Alter und das Geschlecht der Schüler als Faktoren, auf die die Schule keinen Einfluß hat, in ihrem Gewicht berücksichtigt.

Die zweite Gruppe von Variablen ist die der Schul- oder Unterrichtsorganisation. In der Bundesrepublik ist sie für die Population II durch die Unterschiede unserer drei Schularten, die in die Untersuchung einbezogen wurden, gegeben, für die Population IV durch die Schulzweige. Dabei muß darauf hingewiesen werden, daß die international für das "Programm" aufgestellten Kategorien sich in der BRD nur unzureichend anwenden ließen, was im übrigen auch für andere Länder galt, weil sie im wesentlichen an die verschiedenen Zweige (tracks) der meisten amerikanischen senior-high-schools angelehnt waren. Infolgedessen war es notwendig, Gruppierungen zu bilden, die lediglich für ein Land bzw. wenige Länder Gültigkeit haben.

In der dritten Gruppe von Faktoren wurden die Lernbedingungen der Schule zusammengefaßt, wie der Gebrauch von audio-visuellen-Hilfen, die Klassengröße, die Tatsache, daß besondere Stunden für den naturwissenschaftlichen Unterricht vorgesehen waren bzw. ob der Unterricht nach Fächern getrennt oder als allgemeine Naturwissenschaft gegeben wurde, ob ein Lehrbuch vorhanden war, ob Schülerversuche im Unterricht durchgeführt, ob und in welchem Ausmaß Hausarbeiten für den naturwissenschaftlichen Unterricht gemacht wurden, die Zahl der männlichen Lehrer innerhalb des Lehrerkollegiums, die Gelegenheit, die in den Tests vorgesehenen Aufgaben im Unterricht zu lernen sowie die unterschiedliche Ausbildung der Lehrer in Naturwissenschaften.
Zur Faktorengruppe 4 waren unterschiedliche Variablen wie das Verhältnis zur Schule, die Lernmotivation, das Interesse für Naturwissenschaften, die Meinung über die Rolle, die die Naturwissenschaften in der modernen Welt spielen, Berufserwartungen, die Stunden, die täglich vor dem Fernsehgerät zugebracht wurden, wie die Stunden, die für freiwilliges Lesen verwendet wurden, zusammengefaßt.

3.5.2. Ergebnisse

Wir geben in Tabelle 11 eine Übersicht über den ermittelten Beitrag der einzelnen Gruppen von Variablen zur Erklärung der Varianz zwischen den Schulen in der BRD.

Tabelle 11: Beitrag der Gruppen von Variablen zur Varianz zwischen den Schulen (in %)

| Gruppen der Variablen | Pop.I | i.M.[1] | Pop.II | i.M. | Pop.IV | i.M. |
|---|---|---|---|---|---|---|
| 1. Sozio-kultureller Schulindex | 30 | 29 | 46 | 37 | 0 | 14 |
| Alter der Schüler | - | - | 1 | - | - | - |
| Geschlecht der Schüler | - | - | 6 | 6 | 37 | 26 |
| Summe 1 | 30 | 33 | 53 | 45 | 48 | 44 |
| 2. Schultyp bzw. Zweig | 3 | 2 | 3 | 6 | 12 | 8 |
| 3. Lernbedingungen | 34 | 15 | 26 | 15 | 10 | 18 |
| 4. Sonstige Variablen | 1 | 6 | 3 | 4 | 2 | 4 |
| Summe 1-4 | 68 | 55 | 85 | 69 | 72 | 75 |
| Zahl der Schulen | 68 | | 83 | | 80 | |

Quellen: Comber/Keeves, Tab. 8.2; 8.4; 8.6

1) i.M. = internationaler Mittelwert

Population I: In den Gruppen 1 und 2 weichen die Werte für die Bundesrepublik nicht wesentlich von den internationalen Mittelwerten ab. Doch dürfte von Bedeutung sein, daß bereits in der Grundschule 30% der Varianz allein durch den sozio-kulturellen Schulindex bedingt sind.

Erstaunlich ist jedoch - im Gegensatz zum internationalen Mittelwert - der große Teil von Varianz, der durch die Lernbedingungen erklärt wird (34%). Innerhalb dieser Variablengruppe erweist sich international die Klassengröße von unterschiedlicher Bedeutung. Der Regressionskoeffizient beträgt in der BRD 0,18; in fünf weiteren Ländern besteht ebenfalls ein positiver Zusammenhang in dem Sinne, daß bei kleineren Klassen höhere Leistungen zu erwarten sind, während in fünf anderen Ländern die Leistung

um so größer ist, je größer die Klasse ist. Erwähnenswert ist vielleicht noch, daß die Zahl der Stunden, die für Fernsehen angegeben wird, negativ mit der Leistung korreliert.

Wenn auch mit Hilfe der Regressionsanalyse insgesamt 68% der Leistungsvarianz durch die einbezogenen Prädiktoren aufgehellt werden konnten, so ermöglicht das Ergebnis trotzdem nicht, die unterschiedlichen Lernbedingungen, die zu einem Drittel die vorhandene Varianz bedingen, im einzelnen so aufzuklären, als daß daraus eindeutige Hinweise für die Gestaltung des naturwissenschaftlichen Unterrichts in der Primarschule gewonnen werden könnten. Positiv korrelieren mit dem Unterrichtserfolg (in der Reihe der Reihenfolge der Korrelation): der Schultyp (groß, mittel, klein), die Klassengröße, Fachunterricht in den Naturwissenschaften, die Jahrgangsklasse.
Population II: Auffällig ist zunächst, daß der soziokulturelle Schulindex mit 46% fast die Hälfte der Varianz erklärt. Das ist verständlich, weil noch die gesamte Population erfaßt wird, aufgeteilt nach den drei Schularten, wobei, wie wir wissen, sich eine soziale Auslese in den Schularten bemerkbar macht. Auch das Geschlecht der Schüler, das in der Pop.I noch ohne Bedeutung war, trägt hier bereits mit 6% zur Varianz bei, so daß insgesamt 53% derselben durch außerschulische Bedingungen bestimmt sind. Daß der Schultyp bzw. Schulzweig nur wenig zur Aufklärung der Varianz beiträgt, ergibt sich zum Teil aus der Tatsache, daß die Leistungsunterschiede zwischen den Realschulen und den Gymnasien nicht sehr groß sind (s.Figur 3, S.49). Diese ganze Gruppe von Variablen ist auch international gesehen unergiebig geblieben, weil für sie keine eindeutigen allgemein gültigen Kategorien aufgestellt werden konnten.

Wiederum sind es die Lernbedingungen, die für die Varianz innerhalb der BRD in erheblichem Maße ins Gewicht fallen. International gesehen sind die Unterschiede in dieser Faktorengruppe so groß (zwischen 3 und 44%, daß man sich fragen muß, ob es tatsächlich gelungen ist, die Faktoren so zu skalieren, daß sie zu brauchbaren Ergebnissen in der Regressionsanalyse geführt haben. Positiv korrelieren mit dem Unterrichtserfolg (nach der Größe der Korrelationen): Hausaufgaben in den Naturwissenschaften, Schulklima, Prozentsatz männlicher Lehrer. Das Geschlecht der Lehrer spielt in den Populationen II und IV insofern eine Rolle, als männliche Lehrer durchgehend bessere Leistungen erzielen als weibliche.
Immerhin bleibt es bemerkenswert, daß 85% der Varianz durch die in die Analyse eingegangenen Faktoren auf-

geklärt werden.
Population IV: Hierbei fällt sofort auf, daß der sozio-kulturelle Schulindex praktisch für eine Erklärung der Varianz völlig wegfällt, während das Geschlecht der Schüler von größtem Gewicht ist (37%). Sollte man tatsächlich annehmen können, daß die soziale Zusammensetzung der Schülerschaft auf der Oberstufe der Gymnasien so einheitlich ist, daß diesbezüglich keine großen Unterschiede zwischen den Schulen bestehen? Daß der Schulzweig ein höheres Maß an Varianz aufklärt als in Population II, dürfte einleuchtend sein. Erstaunlich ist nur, wie geringfügig die Lernbedingungen auf die Varianz einwirken; das könnte den Schluß zulassen, daß die Lernbedingungen für den naturwissenschaftlichen Unterricht auf der Oberstufe der Gymnasien so einheitlich sind, daß durch die Analyse keine Unterschiede aufgezeigt werden konnten. Es muß jedoch berücksichtigt werden, daß Unterschiede der Lernbedingungen, die durch den Schulzweig bedingt sind, bereits in die Faktorengruppe 2 eingegangen sind. Doch darf nicht übersehen werden, daß sie in etwa das gleiche Gewicht haben wie die Unterschiede des Schulzweiges. Welche Variablen in die restliche, ungeklärte Varianz von 28% eingehen, ist schwer zu sagen. Es dürften sich in ihr sowohl unterschiedliche Begabungen der Schüler wie unterschiedliche Tüchtigkeit von Lehrern sowie die Ausstattung der Schulen niederschlagen. Aufgrund der ermittelten Korrelationen hatten den stärksten Zusammenhang mit der Leistung das Geschlecht der Lehrer sowie der Umfang der gegenwärtigen Studien in den Naturwissenschaften. Letzteres dürfte ohne weiteres verständlich sein.

Geht man anschließend der Frage nach, ob die eingangs aufgestellte Hypothese, daß der Einfluß der Schule auf die Leistungen in den Naturwissenschaften größer sei als bisher bei solchen Untersuchungen festgestellt werden konnte, so läßt sich international, besonders aber für die Bundesrepublik, sagen, daß dieser Einfluß, wie er in den Faktorengruppen 2 und 3 zum Ausdruck kommt, erheblich ist, beträgt der Beitrag zur Varianz doch für die einzelnen Populationen 37, 29 bzw. 22%, nimmt also-im Gegensatz zur internationalen Tendenz- mit zunehmendem Alter der Schüler ab, während man doch annehmen sollte, daß er in der Grundschule am geringsten ist. Um so mehr weist das Ergebnis darauf hin, von wie ungeheurem Einfluß die unterschiedlichen Lernbedingungen auf einen erfolgreichen naturwissenschaftlichen Unterricht, insbesondere in der Grundschule, sind. Das Gefühl, das viele Eltern haben, daß nämlich der Bildungsweg ihres Kindes mit der Wahl der Grundschule, auf die sie keinen Einfluß haben, entscheidend geprägt wird,

läßt sich zum mindesten für die Naturwissenschaften aus den vorliegenden Ergebnissen stützen. Daß dieser Einfluß auf der Unterstufe sogar größer ist als der sozio-kulturelle Schulindex, bestätigt die aufgestellte Hypothese für die Grundschule in entscheidender Weise.
Innerhalb der Lernbedingungen spielen, international gesehen, das Ausmaß und die Möglichkeit, systematisch Naturwissenschaften zu lernen, die größte Rolle. Systematischer naturwissenschaftlicher Unterricht, die Zeit, die diesem Unterricht gewidmet ist, der Aufbau des naturwissenschaftlichen Unterrichts durch die Schuljahre hindurch, die Zeit, die auf Hausarbeiten insgesamt sowie auf Hausarbeiten für den naturwissenschaftlichen Unterricht gerichtet ist, und schließlich die Inhalte des Lehrplans sind die wichtigsten Faktoren. Hier liegen also die größten Einflußmöglichkeiten auf die Effektivität des naturwissenschaftlichen Unterrichts.

### 3.6. Unterschiede zwischen den Schülern und die ihnen zugrunde liegenden Faktoren

Bei dieser Analyse wurde das gleiche Verfahren angewendet wie bei der Untersuchung der Faktoren, die für die Unterschiede zwischen den Schulen bedeutsam sind: Berechnung der Korrelationen zwischen den Prädiktoren und der Leistung und anschließende Regressionsanalysen mit Gruppen dieser Prädiktoren.
Für diese Gruppierung wurde eine größere Zahl, wenn auch im wesentlichen die gleichen Variablen benutzt, wie in 3.5.1. beschrieben.
Tabelle 12 gibt eine Zusammenstellung des Gewichts der einzelnen Faktorengruppen; sie ist unmittelbar mit Tabelle 11 vergleichbar.

Tabelle 12: Seite 37

Die Analyse der Varianz zwischen den Schülern wurde für die jeweiligen Populationen für die gesamte internationale Stichprobe durchgeführt. Da hier der Klumpeneffekt fortfiel, war anzunehmen, daß ein geringerer Teil an Varianz aufgeklärt würde, was tatsächlich auch der Fall ist. Für die Populationen in der BRD wurden somit nur 27, 34 und 39% ermittelt, was in etwa dem internationalen Mittel entspricht.
Wesentlich neue Einsichten ergab diese Analyse nicht; sie ergibt die gleichen relativen Gewichte für die Faktorengruppen, wenn auch die Prozentanteile im einzelnen geringer ausfallen.
In der Population IV ergibt sich hinsichtlich des Alters der Schüler, daß jüngere Schüler einheitlich bessere Leistungen erzielen, was wahrscheinlich dadurch zu erklären ist, daß ältere Schüler z.T. weniger leistungsfähige Wiederholer sind.

Tabelle 12: Beitrag der Gruppen von Variablen zur Varianz zwischen den Schülern (in%)

| Gruppen der Variablen | Pop.I | i.M. | Pop.II | i.M. | Pop.IV | i.M. |
|---|---|---|---|---|---|---|
| 1. Sozio-kultureller Schulindex | 4 | 8 | 10 | 10 | 0 | 2 |
| Alter der Schüler | - | - | - | - | 1 | |
| Geschlecht der Schüler | 2 | 2 | 8 | 4 | 12 | 11 |
| Summe 1 | 8 | 11 | 18 | 16 | 13 | 14 |
| 2. Schultyp bzw. Zweig | 2 | 1 | 2 | 6 | 12 | 7 |
| 3. Lernbedingungen | 10 | 8 | 14 | 9 | 8 | 12 |
| 4. Sonstige Variablen | 4 | 6 | 0 | 5 | 6 | 5 |
| Summe 1-4 | 27 | 24 | 34 | 36 | 39 | 44 |

Quellen: Comber/Keeves, Tab. 9.2; 9.4; 9.6

Abschließend darf man wohl feststellen, daß zur Klärung der Varianz zwischen den Schülern vor allem der Faktor herangezogen werden müßte, der gerade in diesen Untersuchungen nicht erfaßt wurde: Unterschiedliche Begabung bzw. schulische Leistungsfähigkeit der Schüler sowie ihre unterschiedliche Motivation.

## 4. Der naturwissenschaftliche Unterricht in der Bundesrepublik

Vorbemerkung:

Wir können den folgenden Analysen nur die Zahlen zugrunde legen, die uns im Zusammenhang der internationalen Analyse für die BRD zugegangen sind. Eine eingehendere Analyse der nationalen Ergebnisse verbot sich aus finanziellen Gründen, aber auch, weil sich infolge mancher Unzulänglichkeiten eine solche weitergehende Analyse für die Bundesrepublik kaum lohnte.

### 4.1. Die Stichprobe

Tabelle 13 gibt einen Überblik über die Stichprobe und zugleich die korrigierten Mittelwerte für die Leistung in den Naturwissenschaften und im Wortschatztest aufgrund der vom Computer ausgeworfenen Univariatenliste. (Die Zahlen in den verschiedenen in dieser Darstellung wiedergegebenen Tabellen weichen zum Teil deshalb geringfügig voneinander ab, weil für verschiedene Analysen je nach Vollständigkeit der Unterlagen verschiedene Zahlen von Schulen bzw. Schülern in die Berechnungen eingegangen sind).

Tabelle 13: Seiten 39 und 40

Wie bereits erwähnt, konnte trotz der doppelt gezogenen Schullisten infolge vieler Absagen weder in allen Bundesländern die wünschenswerte Zahl von Schulen erreicht werden, noch konnten je Schule 30 Schüler in die Untersuchung einbezogen werden. Letzteres gilt vor allem für die Populationen I und IV. Zum Teil fiel die Hälfte, in einzelnen Ländern mehr als die Hälfte, der angeschriebenen Schulen aus. Das hatte seinen Grund wahrscheinlich u.a. in der Tatsache, daß besonders in kleinen Ländern mit Universitäten und Pädagogischen Hochschulen die Schulen stark mit Schulversuchen und Begleituntersuchungen belastet sind.

Wenn somit eine für die gesamte Bundesrepublik noch repräsentative Stichprobe herauskam, so sind die Stichproben für einige Länder zu klein, als daß sie auch für diese Länder als repräsentativ angesehen werden könnten. Schon dadurch, daß bei der Wahl von

Tabelle 13: Zahl der Schulen und Schüler und Testergebnisse ( Naturw. und Wortschatz) nach Bundesländern und Populationen (Standardabweichungen in Klammern)

Population I:

| L a n d | Zahl d. Schulen | Zahl d. Schüler | korrig. Mittelwerte Naturw. | korrig. Mittelw. Wortschatz |
|---|---|---|---|---|
| Nordrh.-Westf. | 17 | 523 | 13.0(7.1) | 12.1(8.8) |
| Bayern | 16 | 336 | 18.6(7.6) | 16.9(9.3) |
| Baden-Wtg | 14 | 285 | 14.5(7.8) | 11.3(9.4) |
| Nieders. | 5 | 157 | 14.6(6.7) | 12.9(9.0) |
| Hessen | 8 | 154 | 14.1(6.4) | 14.4(7.6) |
| Rheinl.-Pfalz | 6 | 126 | 15.0(6.0) | 14.6(8.4) |
| Schlesw.-Holst. | 3 | 22 | 13.8(6.5) | 13.7(8.1) |
| Saarland | 1 | 41 | 15.3(5.4) | 15.9(7.0) |
| Hamburg | 1 | 42 | 14.0(6.6) | 12.4(6.0) |
| Berlin | 1 | 36 | 16.2(7.5) | 13.3(8.3) |
| Bremen | 1 | 20 | 14.2(7.3) | 9.6(9.7) |
| BRD | 73 | 1742 | 14.9(7.4) | 13.4(9.0) |

Rangkorrelation Naturw./Wortschatz r= 0,79[1)]

| Population II: | VS. | RS. | G. | | | |
|---|---|---|---|---|---|---|
| Nordrh.-Westf. | 11 | 3 | 6 | 608 | 22.7(10.9) | 15.9(9.6) |
| Bayern | 2 | 3 | 3 | 249 | 27.3( 9.6) | 20.1(8.7) |
| Baden-Wtg. | 14 | 1 | 4 | 447 | 22.1(10.2) | 14.9(9.8) |
| Nieders. | 12 | 3 | 2 | 387 | 21.6(11.3) | 12.1(9.6) |
| Hessen | 5 | 2 | 1 | 208 | 23.1( 9.8) | 16.5(8.9) |
| Rheinl.-Pfalz | 3 | 0 | 2 | 125 | 34.3(17.4) | 21.8(10.1) |
| Schlesw.-Holst. | 3 | 1 | 1 | 93 | 24.8(11.3) | 15.5(8.4) |
| Saarland | 3 | 0 | 1 | 69 | 22.8(9.3 ) | 15.3(9.4) |
| Hamburg | 0 | 1 | 1 | 33 | 25.4(9.6) | 19.2(9.1) |
| Berlin | 0 | 1 | 0 | 13 | 26.0(11,9) | 13.7(8.0) |
| Bremen | 0 | 0 | 0 | 0 | --- | --- |
| BRD | 53 | 15 | 21 | 2232 | 23.7 (11.5) | 15.9(9.8) |

Rangkorrelation Naturw./Wortschatz r= 0.89[1)]

Population IV: Seite 40

Population IV:

| L a n d | Zahl d. Schulen | Zahl d. Schüler | korrig. Mittelwerte Naturw. | korrig. Mittelw. Wortschatz |
|---|---|---|---|---|
| Nordrh.-Westf. | 19 | 553 | 26.5(8.6) | 24.5(6.4) |
| Bayern | 14 | 333 | 30.5(8.8) | 27.4(6.7) |
| Baden-Wtg. | 14 | 309 | 26.8(8.6) | 24.4(5.6) |
| Nieders. | 9 | 254 | 24.8(9.3) | 22.1(7.0) |
| Hessen | 8 | 232 | 25.9(8.6) | 22.7(6.3) |
| Rheinl.-Pfalz | 7 | 129 | 26.3(8.7) | 25.7(6.2) |
| Schlesw.-Holst. | 4 | 102 | 26.1(8.5) | 22.1(6.2) |
| Saarland | 3 | 26 | 25.3(8.6) | 28.6(5.3) |
| Hamburg | 2 | 52 | 28.7(8.7) | 25.7(4.9) |
| Berlin | 0 | 0 | --- | --- |
| Bremen | 0 | 0 | --- | --- |
| BRD | 80 | 1990 | 26.9(8.9) | 24.5(6.6) |

Rangkorrelation Naturw./Wortschatz r = 0.82[1)]

Quelle: Liste der univariat analysierten Variablen

1) nur Länder mit mehr als 90 Schülern.

Ersatzschulen darauf geachtet werden mußte, die gleichen Strata zu wählen, mußte zum Teil auf andere Länder ausgewichen werden. In der Population IV haben wir uns von vornherein auf die drei überlieferten Schulzweige beschränkt und neuere Zweige (Wirtschaftsgymnasien, sozialwissenschaftliche Zweige, Gymnasien für Frauenbildung usw.) nicht berücksichtigt. Das war auch nicht erforderlich, weil nicht die Absicht bestand, einen nationalen Survey mit der Untersuchung durchzuführen.

Diese"Stichprobenfehler" haben zur Folge, daß eine Auswertung für die BRD eine Reihe von Einschränkungen erfahren muß. Wir haben für die Berechnung von Korrelationen alle Länder ausgelassen, von denen weniger als fünf Schulen oder weniger als 100 Schüler (in Pop. II 90 Schüler) in die Untersuchung einbezogen wurden.

In Pop. II fällt auf, daß z.B. für Bayern nur eine geringe Zahl von Volksschulen herangezogen wurde, so daß sich daraus die relativ hohe Durchschnittsleistung für das Land Bayern ohne weiteres erklärt. Einzelne Schulen (und z.T. das betreffende Land) mußten deshalb für die weitere Analyse aus der Stichprobe ausgeschlossen werden, weil offensichtlich in Einzelfällen nicht sorgfältig administriert wurde. So sind z.B. in einer Volksschulklasse der Population II so hohe Leistungen erzielt worden, daß sie sämtliche andere Schulen einschließlich der Gymnasien weit in den Schatten stellt. Daraus erklärt sich in Rheinland-Pfalz (bei nur drei Volksschulen in der Population II) der übermäßig hohe Durchschnittswert sowohl im Wortschatztest wie in den Naturwissenschaften. Auch die große Streuung, die aus den übrigen Werten herausfällt, läßt sofort den Verdacht aufkommen, daß hier außergewöhnliche Verhältnisse vorliegen.

## 4.2. Die außerschulischen Variablen

4.2.1. Bereits im internationalen Vergleich wurde deutlich, ein wie großer Einfluß durch die Faktorengruppe (1)- sozio-kultureller Schulindex einschließlich Alter und Geschlecht der Schüler - auf die Schulleistungen ausgeübt wird. Sie erklärte in der Bundesrepublik für die Population I 30%, für die Population II 53% und für die Population IV 48% der Varianz. Leider liegt weder für die einzelnen Schulen noch für die Länder der Bundesrepublik dieser sozio-kulturelle Schulindex vor. So konnten wir die Abhängigkeit der Leistungen lediglich von den Berufen der Väter ermitteln. Wir haben zu diesem Zweck die Durchschnittsleistungen der Länder mit dem Prozentanteil von Kindern in den unteren und oberen drei Berufsgruppen ( von insges. 9)korreliert und die Ergebnisse in Tabelle 14 dargestellt.

Tabelle 14: Rangkorrelationen zwischen den Durchschnittsleistungen (Naturw.) der Länder und den unteren und oberen Berufsgruppen der Väter.

| | Berufsgruppen | |
|---|---|---|
| | 0 - 3 | 7 - 9 |
| Population I | 0,03 | 0,60 |
| Population II | 0,43 | 1,00 |
| Population IV | - 0,04 | 0,37 |

Es ergibt sich die erstaunliche Tatsache, daß die Leistungen lediglich in Population II in dem Sinne mit den Berufen der Väter korrelieren, als eine größere Anzahl von Kindern aus Familien mit Vätern unterer Berufsgruppen eine niedrigere Durchschnittsleistung zur Folge hat. In Population I und IV ist diese Korrelation praktisch gleich null. Die Korrelation der Leistung mit der Herkunft aus oberen Berufsgruppen ergibt eindeutig positive Korrelationen, auffällig hoch wiederum für die Population II. Die niedrigere Korrelation für Population IV ist aus bereits genannten Gründen verständlich. Somit kann man verallgemeinernd sagen, daß der Einfluß der Sozialstruktur auf die Leistungen in der Schule stärker durch den Anteil der Kinder aus Familien gehobener Berufe als durch die Kinder aus Familien unterer Berufsgruppen bestimmt wird.

In Tabelle 15 geben wir eine Übersicht über die Zusammensetzung der Populationen in den verschiedenen Ländern nach den Berufsgruppen der Väter.

Tabelle 15: Seite 43 a

Bei Betrachtung der Tabelle fällt auf, daß bereits in der Population I erhebliche Unterschiede zwischen den Ländern hinsichtlich der Zusammensetzung der Schüler bestehen. Der Anteil der Schüler aus den Berufsgruppen 0-3 schwankt zwischen 48,9 und 70,6%, entsprechend bei den Berufsgruppen 7-9 zwischen 5,4 und 15,9%. In Nordrhein-Westfalen, das in der Population I und II den höchsten Anteil an Kindern aus unteren Berufsgruppen hat, trifft diese Situation in der Population IV nicht mehr zu. Diese Änderung wird bewirkt durch die eindeutig soziale Auslese in den Gymnasien, die eine Angleichung an die relativ einheitliche Situation in den übrigen Ländern in der Population IV zur Folge hat.

4.2.2. Was die Zusammensetzung der Schüler nach der schulischen Begabung, der zweiten Faktorengruppe, auf die die Schule keinen Einfluß hat, anbelangt, wurde zu ihrer groben Ermittlung ein Wortschatztest gewählt.

Ein Vergleich der Leistungen in beiden Bereichen bestätigt die Tatsache, daß der Wortschatztest wie mit anderen Leistungen, so auch mit den naturwissenschaftlichen Leistungen, hoch korreliert, so daß man ihn mit einer gewissen Vorsicht benutzen kann, um die allgemeine schulische Leistungsfähigkeit der Schüler auszudrücken.

In Tabelle 13, in der die beiden Werte gegenübergestellt sind, haben wir bereits die ermittelten Korrelationen wiedergegeben: Sie betragen für Population I r= 0,79, für Population II r = 0,89 und für Population IV r = 0,82.
Diese Korrelationen wären sicherlich noch höher, wenn die Schülerleistung und nicht die durchschnittliche Leistung der Länder als Einheit für die Berechnung zugrunde gelegt worden wäre.

Genauso wie die Länder (und die Schulen) sich hinsichtlich der sozio-kulturellen Zusammensetzung ihrer Schüler unterscheiden, genauso unterscheiden sie sich hinsichtlich der Zusammensetzung ihrer Schüler in bezug auf die allgemeine schulische Leistungsfähigkeit, wie sie mit dem Wortschatztest erfaßt wird. So wie Nordrhein-Westfalen hinsichtlich der sozialen Zusammensetzung seiner Schüler in den Populationen I und II am unteren Ende der Rangskala liegt, so auch hinsichtlich der durchschnittlichen Leistungen im Wortschatztest. Dem entspricht in Population I verständlicherweise die niedrige Durchschnittsleistung in Naturwissenschaften, die gerade in dieser Population weitgehend durch außerschulische Faktoren bestimmt sein dürfte.
Die Durchschnittsleistung eines Landes in Naturwissenschaften und damit die Leistung jeder Schule dürfen also jeweils nur unter Berücksichtigung des sozio-kulturellen Index (einschl. Geschlecht der Schüler) wie der Leistungen im Wortschatztest beurteilt werden.

## 4.3. Korrelationen zwischen den Populationen

Bei der Betrachtung der durchschnittlichen Leistungen der einzelnen Länder ergibt sich die Frage, wie weit die Leistungsunterschiede in allen Populationen konstant sind.

Die Leistungen in den Naturwissenschaften korrelieren zwischen der Population I und II mit r = 0,43, zwischen den Populationen II und IV mit r = 0,32 und zwischen den Populationen I und IV mit r = 0,26.
Die Korrelation ist also positiv, nimmt jedoch mit zunehmendem Alter der Schüler ab.Mit anderen Worten:
Die Leistungsunterschiede zwischen den Ländern sind bis zu einem gewissen Grade auf allen Stufen konstant; in Ländern, in denen relativ gute Leistungen am Ende

Tabelle 15: Anteil der Berufsgruppen an der Zusammensetzung der Schüler in den Ländern der BRD nach Populationen (in %)

| | Population I | | | | Population II | | | | Population IV | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| L ä n d e r | 0-3 | 4-6 | 7-9 | N | 0-3 | 4-6 | 7-9 | N | 0-3 | 4-6 | 7-9 | N |
| Nordrh.-Westf. | 70,6 | 26.9 | 2.6 | 508 | 59.0 | 31.6 | 9.4 | 607 | 10.1 | 48.1 | 41.9 | 548 |
| Bayern | 59.7 | 30.6 | 9.6 | 327 | 47.7 | 37.9 | 14.4 | 249 | 10.0 | 52.1 | 38.0 | 325 |
| Baden-Wtg | 48.9 | 38.1 | 15.9 | 266 | 51.3 | 39.7 | 9.1 | 435 | 11.4 | 45.7 | 42.8 | 306 |
| Nieders. | 55.7 | 37.2 | 7.2 | 156 | 50.4 | 43.5 | 6.0 | 371 | 8.6 | 52.3 | 40.2 | 253 |
| Hessen | 51.2 | 43.3 | 5.4 | 149 | 54.4 | 35.3 | 10.3 | 207 | 16.7 | 46.0 | 37.3 | 217 |
| Rheinl.-Pfalz | 61.2 | 24.6 | 14.4 | 126 | 46.6 | 46.5 | 7.0 | 124 | 7.5 | 50.4 | 41.9 | 127 |
| Schlesw.-Holst. | 44.5 | 42.7 | 13.1 | 22 | 45.9 | 39.2 | 14.9 | 94 | 12.6 | 49.8 | 37.6 | 100 |
| Saarland | 63.6 | 36.4 | 0.0 | 41 | 63.3 | 29.2 | 6.9 | 68 | 16.1 | 57.7 | 26.0 | 25 |
| Hamburg | 35.3 | 52.9 | 11.8 | 42 | 90.5 | 9.5 | 0.0 | 13 | 12.2 | 50.5 | 37.3 | 52 |
| Berlin | 37.2 | 51.4 | 11.4 | 36 | 45.9 | 44.2 | 10.0 | 33 | - | - | - | 0 |
| Bremen | 54.1 | 24.3 | 21.6 | 20 | - | - | - | 0 | - | - | - | 0 |

Quelle: Liste der univariat analysierten Variablen

der Grundschule vorhanden sind, ist dies auch in den übrigen Populationen der Fall, wenngleich aus verständlichen Gründen der Zusammenhang nicht sehr stark sein kann und mit zunehmendem Alter der Schüler abnimmt.

## 4.4. Analyse der Ergebnisse für die drei Populationen

Soweit aufgrund des vorliegenden Materials möglich, sollen im folgenden einige Analysen getrennt nach Populationen durchgeführt werden. In den Figuren 1 und 2 geben wir die Verteilungskurven für die Gesamtleistung sowie die Leistungen in den Teiltests, getrennt nach Populationen. Es ergibt sich daraus, daß trotz der Schwierigkeit des Gesamttests die Verteilung auch in sämtlichen Teiltests nur unwesentlich von einer Normalverteilung abweicht (größte Schiefe -0,4 bzw. +0,4).

Figuren 1 und 2: Seiten 45 und 46

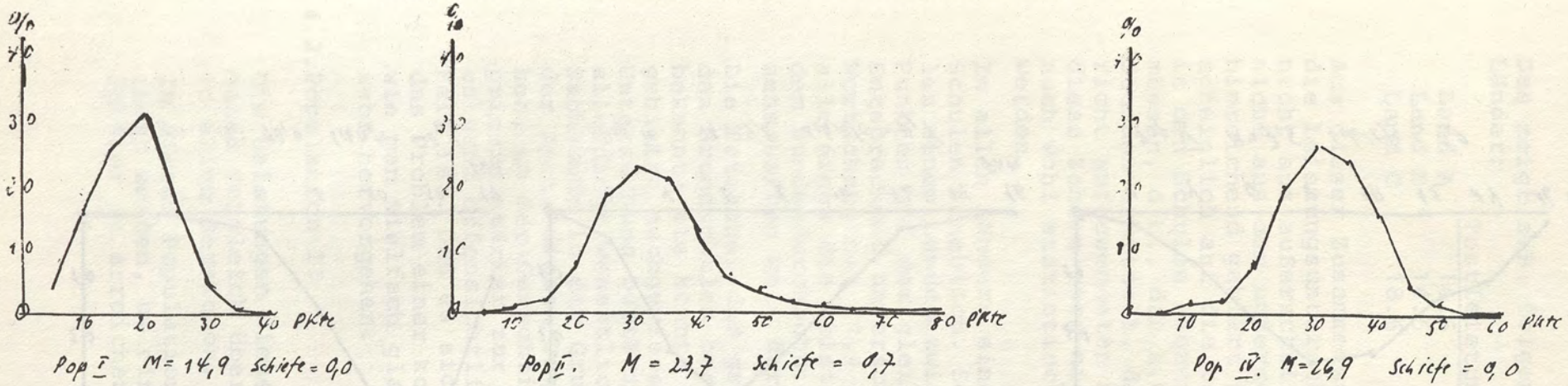
### 4.4.1. Population I

Wir sahen bereits im internationalen Vergleich, daß die Leistungen am Ende der Grundschule in der Bundesrepublik sehr wesentlich unter dem internationalen Durchschnitt liegen - wahrscheinlich wegen des Fehlens systematischen Fachunterrichts.

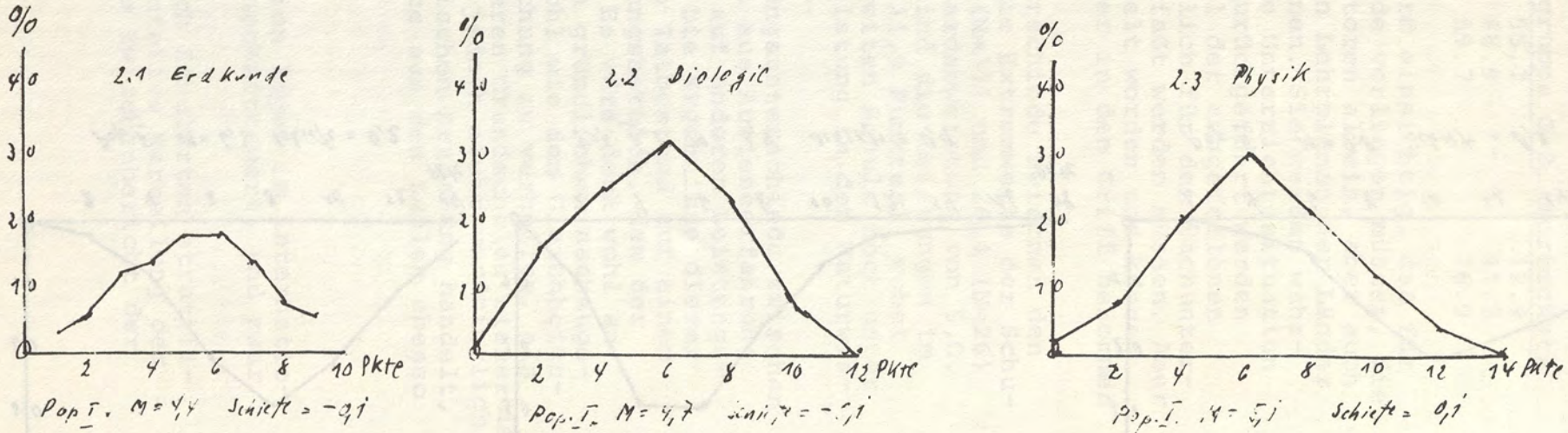
Die geringe Zahl der richtigen Lösungen macht es uninteressant, den Leistungsunterschieden nachzugehen. Erwähnenswert ist lediglich, daß es Schulen gibt, in denen nicht die Physik an erster Stelle liegt, sondern die Erdkunde oder die Biologie.

Im Ländervergleich fällt auf, daß das Land Bayern positiv, das Land Nordrhein-Westfalen negativ herausragen. Das kann nicht ausschließlich aus der sozialen Zusammensetzung der Schülerschaft erklärt werden, wenngleich auch beim Wortschatztest auffällt, daß Bayern einen auffällig hohen Mittelwert erreicht. Es gibt jedoch Länder, wie Baden-Württemberg, die mit dem niedrigsten Anteil der Schüler aus der Berufsgruppe 0-3 und dem gleichzeitig höchsten Anteil aus der Berufsgruppe 7-9 im Wortschatz auffällig niedrige Leistungen zeigen, aber in den naturwissenschaftlichen Leistungen unter dem Bundesdurchschnitt liegen. Länder, die praktisch gleiche Leistungen aufweisen (Baden-Württemberg / Niedersachsen), zeigen andererseits eine völlig unterschiedliche soziale Zusammensetzung der Schülerschaft.

Figur 1: Verteilungskurven - Gesamtleistung

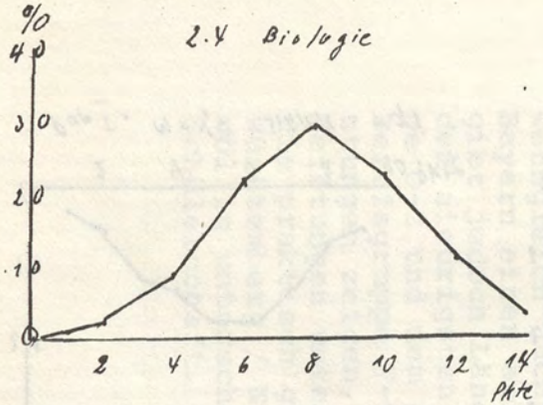



Figur 2: Verteilungskurven - Teiltests



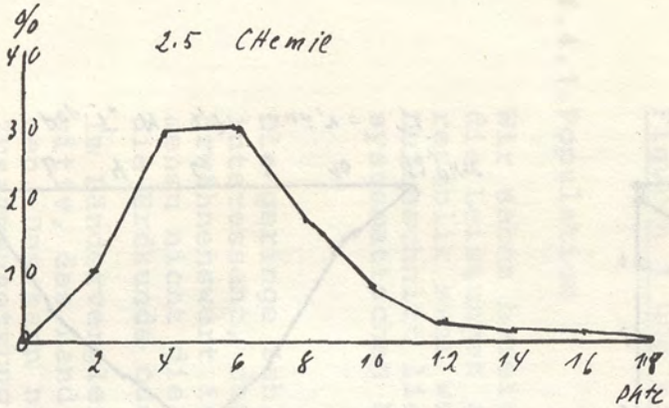
Fortsetzung: Teiltests

2.4 Biologie



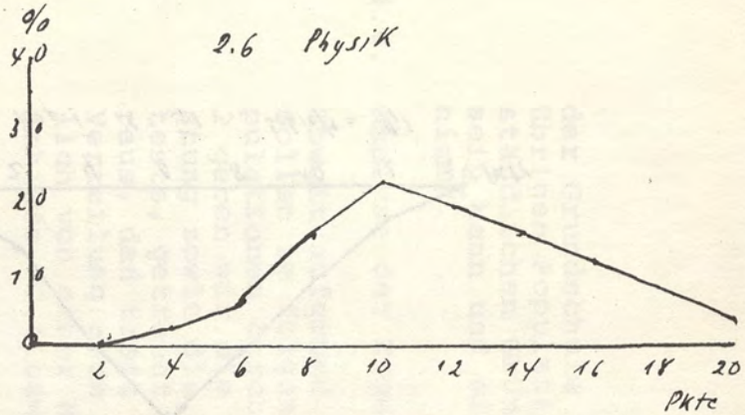

Pop. II. M = 6,3 Schiefe = 0,2

2.5 CHemie



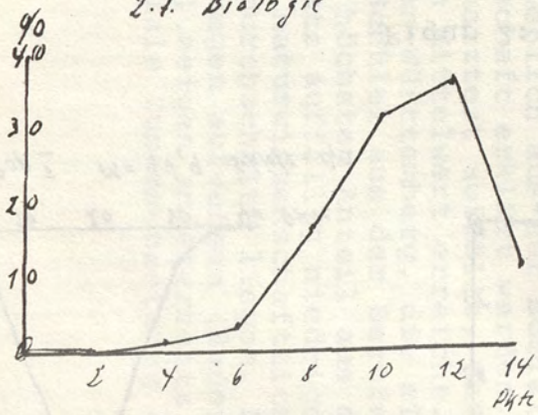

Pop. II. M = 3,4 Schiefe = 0,9

2.6 Physik



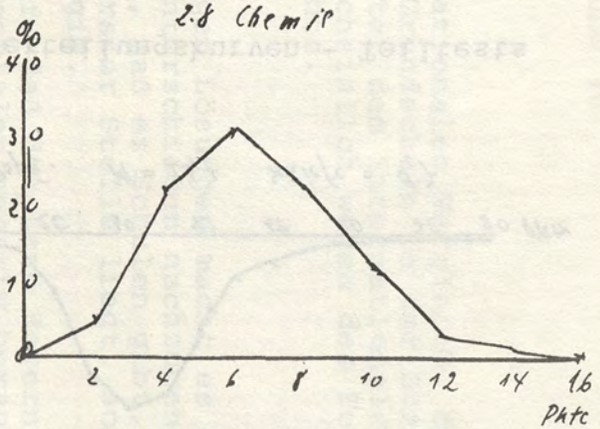

Pop. II. M = 9,8 Schiefe = 0,3

2.7. Biologie



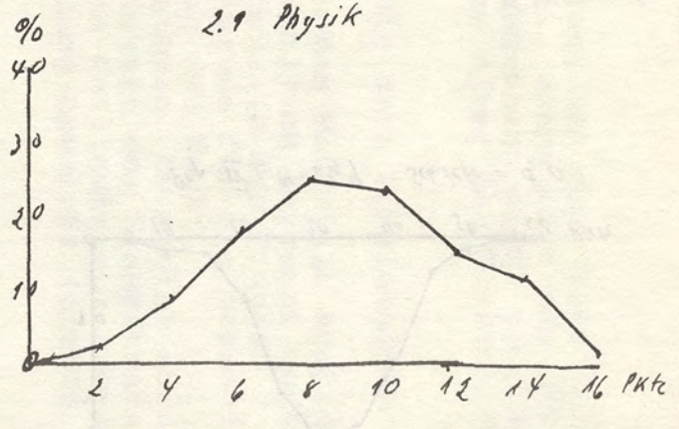

Pop. IV. M = 10,0 Schiefe = -0,4

2.8 Chemie




Pop. IV. M = 4,8 Schiefe = 0,4

2.9 Physik




Pop. IV. M = 7,5 Schiefe = 0,0

Das zeigt z.B. folgende Gegenüberstellung dreier Länder:

| | Testleistung | Berufsgruppe O-3 | Wortschatz |
|---|---|---|---|
| Land A | 14.6 | 55.7 | 12.9 |
| Land B | 14.5 | 48.9 | 11.3 |
| Land C | 18.6 | 59.7 | 16.9 |

Aus dieser Zusammenstellung wird einsichtig, daß für die Leistungsunterschiede Gründe vorliegen müssen, die nicht aus außerschulischen Faktoren allein, aber auch nicht aus den unterschiedlichen Lehrplänen der Länder hinreichend geklärt werden können. Sie werden wahrscheinlich auf die tatsächliche Unterrichtssituation in den Schulen dieser Länder zurückgeführt werden müssen, d.h., daß z.B. die Zahl der ausgefallenen Stunden ermittelt, die tatsächlich für den Sachunterricht aufgewendeten Stunden erfaßt werden müssen. Aber diese Zahlen sind nicht ermittelt worden und können auch wohl statistisch nur schwer in den Griff bekommen werden.

In allen Ländern sind die Unterschiede zwischen den Schulen erheblich. So liegen die Extremwerte der Schulen eines Landes zwischen 5,4 (N=17) und 26,4 (N=26) Punkten bei der gleichen Standardabweichung von 5,0. Entsprechend unterschiedlich sind die Leistungen im Wortschatztest (1,4 gegenüber 11,9 Punkten), wobei allerdings die Leistung der zweiten Schule noch unter dem Bundesdurchschnitt, die Leistung in den Naturwissenschaften weit darüber liegt.

Die Tatsache, daß große Leistungsunterschiede zwischen den Grundschulen bestehen, ist aus Ausleseerfahrung bekannt; sie könnte ebensogut auf anderen Leistungsgebieten nachgewiesen werden. Die Ergebnisse dieser Untersuchung bestätigen diesen Tatbestand auf einem allerdings wesentlichen Leistungsbereich, dem der Sachkunde in der Grundschule. Es wäre doch wohl an der Zeit, diesen Unterschieden gründlicher nachzugehen, um der Gerechtigkeit sowohl wie dem Gleichheitsprinzip stärker zur Verwirklichung zu verhelfen, als es ganz offensichtlich in unseren Grundschulen bisher der Fall ist. Daß es sich hierbei jedoch nicht vordringlich um das Problem einer kompensatorischen Erziehung handelt, wie man vielfach glaubt, dürfte aus den Zahlen ebenso sehr hervorgehen.

### 4.4.2. Population II

Die Leistungen dieser Population lagen im internationalen Vergleich über dem Durchschnittswert, und zwar in allen Bereichen.

In dieser Population mußte nach Schularten stratifiziert werden, um eine repräsentative Verteilung der Schüler zu erreichen. Nach dem Zwischenbericht der

Bund-Länder-Kommission verteilten sich die dreizehnjährigen Schüler 1969 folgendermaßen auf die Schularten (in Prozent):

Hauptschulen 56,8; Realschulen 18,0 Gymnasien 20,5 Sonderschulen 4,9.

Unter Auslassung der Sonderschulen kommen damit auf je 100 Hauptschüler : 32 Realschüler und
36 Gymnasiasten.

In unserer Erhebung müßten sich 2 232 Schüler demgemäß folgendermaßen verteilen:

1 329 Hauptschüler,
425 Realschüler,
478 Gymnasiasten.

Die tatsächliche Verteilung betrug jedoch:

1 151 Hauptschüler,
396 Realschüler,
691 Gymnasiasten,

d.h., auf 100 Hauptschüler kommen 34 Realschüler und
66 Gymnasiasten.

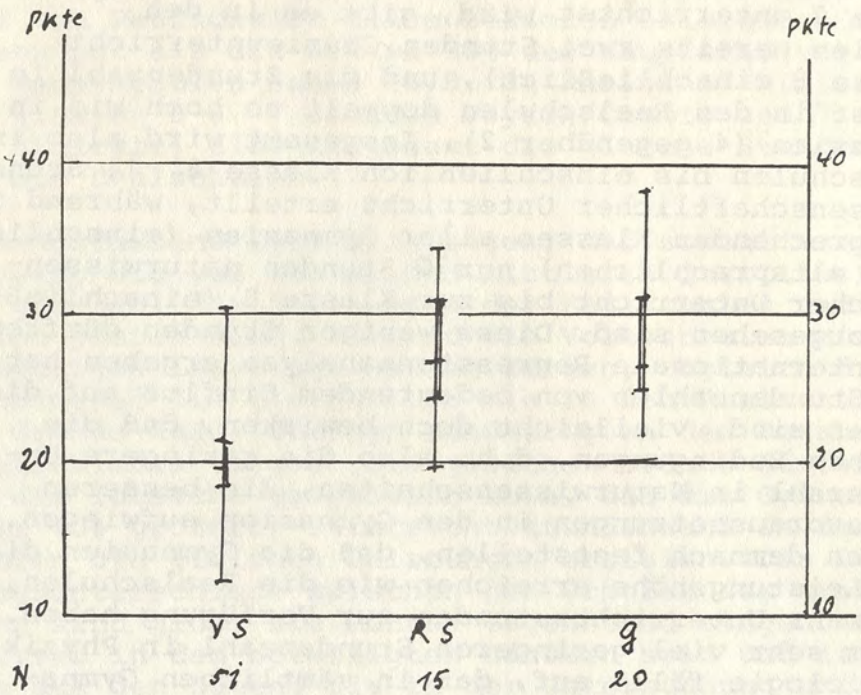
Das Übergewicht an Gymnasiasten konnte für den internationalen Vergleich durch Gewichtung ausgeglichen werden. Für die BRD verbieten sich jedoch Ländervergleiche, zumal ein den Ländern entsprechendes Verhältnis zwischen den Schülern nicht innegehalten werden konnte. In Tabelle 16 geben wir die Ergebnisse für die Population II in Form von Quartilen wieder:

Tabelle 16: Leistungsverteilung der Schulen (in korrigierten Punktwerten) in Population II nach Quartilen, getrennt nach Schularten

| | Hauptschulen (N = 51) | Realschulen (N = 15) | Gymnasien (N = 20) |
|---|---|---|---|
| HW | 30.2 | 34.3 | 38.2 |
| $Q_3$ | 21.5 | 31.0 | 31.2 |
| Md | 19.8 | 26.7 | 26.5 |
| $Q_1$ | 18.5 | 24.3 | 25.1 |
| NW | 12.1 | 19.6 | 22.2 |

Noch besser als die Zahlen gibt die graphische Darstellung (Figur 3) ein Bild der Leistungssituation.

Figur 3: Die Leistungen der Volksschulen, Realschulen und Gymnasien in Quartilwerten




Die Realschulen und Gymnasien liegen etwa leistungsgleich weit über den Leistungen der Hauptschulen, die lediglich mit den oberen 50% ($Q_3$ und $Q_4$) in die Leistungsbereiche der weiterführenden Schulen hineinreichen. 75% der Hauptschulen bleiben unter den Leistungen der Gymnasien. Auffällig ist ferner, daß die mittleren 50% der Hauptschulen ($Q_2$ und $Q_3$) relativ eng beieinander liegen und nur um drei Punkte streuen, während die mittleren 50% der weiterführenden Schulen sich leistungsmäßig über die doppelte Punktzahl erstrecken.

Der Medi nwert der Hauptschulen liegt genau auf der Höhe der Realschule mit der niedrigsten Leistung.

Die Realschulen und Gymnasien unterscheiden sich lediglich dadurch, daß das untere Quartil der Realschulen niedriger liegt als das entsprechende Quartil der Gymnasien, und daß das obere Quartil sich nicht so weit hinauf erstreckt.

Dieser annähernde Gleichstand der Leistungen in den Gymnasien widerspricht der Tatsache, daß die Sozialstruktur einen erheblichen Einfluß auf die Leistungen hat, liegt doch die Sozialstruktur der Gymnasien günstiger als die der Realschulen. Die Lehrpläne können nur zum Teil zur Klärung herangezogen werden. Die Stundentafeln zeigen für Biologie für die Klassen 5-8 in

beiden Schultypen die gleiche Zahl von 8 Stunden, so daß für die Leistungen in Biologie die gleichen Voraussetzungen gegeben sind. Anders ist es in Chemie und Physik. Während in den Gymnasien Chemie erst ab Klasse 9 unterrichtet wird, gibt es in den Realschulen bereits zwei Stunden Chemieunterricht (Klasse 8 einschließlich), und die Stundenzahl in Physik ist in den Realschulen doppelt so hoch wie in den Gymnasien (4 gegenüber 2). Insgesamt wird also in den Realschulen bis einschließlich Klasse 8 14 Stunden naturwissenschaftlicher Unterricht erteilt, während in den entsprechenden Klassen aller Gymnasien (einschließlich der altsprachlichen) nur 10 Stunden naturwissenschaftlicher Unterricht bis zur Klasse 8 (einschließlich ) vorgesehen sind. Diese wenigen Stunden dürften, da die internationale Regressionsanalyse ergeben hat, daß die Stundenzahlen von bedeutendem Einfluß auf die Leistungen sind, vielleicht doch bewirken, daß die schulischen Bedingungen, d.h. also die geringere Unterrichtszahl in Naturwissenschaften, die besseren Leistungsvoraussetzungen in den Gymnasien aufwiegen. Wir können demnach feststellen, daß die Gymnasien die gleiche Leistungshöhe erreichen wie die Realschulen, die 40% mehr Unterrichtsstunden zur Verfügung haben. Trotz der sehr viel geringeren Stundenzahl in Physik als in Biologie fällt auf, daß in sämtlichen Gymnasien (und Realschulen) die Leistungen in Physik sehr viel höher liegen als in der Biologie, was durchaus dem Ergebnis in der BRD insgesamt entspricht, zumal das gleiche auch für die Hauptschulen gilt.

Hinsichtlich etwaiger Leistungsunterschiede zwischen großen und kleinen Schulsystemen lassen sich keine eindeutigen Aussagen machen. Es gibt Schulen sowohl mit schwachen wie mit guten Leistungen in allen Größen. Die Zahl der Aufgaben in Physik ist größer als die in den beiden anderen Bereichen (22 Aufgaben gegenüber 19), aber bei dem Prozentsatz der richtigen Lösungen kann dieser Unterschied nicht die erheblich besseren Leistungen in Physik erklären. Hier müssen starke außerschulische Interessen eine Rolle spielen. Leider ist es nicht möglich, die Unterschiede in den Leistungen der Knaben und der Mädchen zu verfolgen; wir wissen nur aus der internationalen Analyse, daß die Knaben einen erheblichen Leistungsvorsprung haben.

Wenn Ländervergleiche auch aus den angegebenen Gründen nicht möglich sind, so fällt doch auf, daß insgesamt die Streuungen in der Population II größer sind als in der Population I, eine Folge der Unterschiede zwischen den Schularten. Das war aufgrund des Vergleichs der Quartilwerte und der allgemein höheren Durchschnittswerte zu erwarten.

Offen bleibt die Frage, wie es zu erklären ist, daß immerhin die Hälfte der beteiligten Volksschulen in die Leistungsbereiche der Realschulen und Gymnasien hineinragen, oder anders formuliert, warum eine große Zahl von Realschulen keine besseren Leistungen aufzuweisen hat als die besten 50% der Hauptschulen. Die Hauptschulen haben (z.B. in Rheinland-Pfalz) in den Klassen 5-8 7 Stunden Physik/Chemie und 7 Stunden Biologie, also insgesamt die gleiche Stundenzahl wie die Realschulen.

Zusammenfassend läßt sich sagen, daß sowohl vom Medianwert wie von den beiden mittleren Quartilwerten her gesehen, die Leistungen der Realschulen und Gymnasien gleich sind, während die Volksschulen einen eindeutigen Leistungsabstand zeigen, obgleich die Stundenzahl die gleiche ist wie in den Realschulen. Die Gründe dafür dürften eindeutig in den Begabungsunterschieden liegen. Die gleichen Gründe dürften auch für die Tatsache maßgeblich sein, daß die Gymnasien trotz 40% größerer Unterrichtsstundenzahl in den Realschulen die gleichen Leistungen erzielen. Die Leistungsunterschiede zwischen den Schulen der drei Schularten sind groß, sie müßten Gegenstand eingehender Analysen in den beteiligten Ländern sein und lassen sich aus den Daten, die in der Untersuchung erfaßt wurden, nicht erklären.

### 4.4.3. Population IV

Bei dieser Population handelt es sich im Gegensatz zu den Populationen I und II um eine ausgelesene Gruppe, die 9% eines Jahrgangs erfaßt. Die verhältnismäßig große Homogenität der Population läßt deshalb gültige Vergleiche auch zwischen den Ländern zu. Die einzige Fehlerquelle könnte darin liegen, daß in verschiedenen Ländern die Anteile der in die Untersuchung einbezogenen Schulzweige nicht dem zahlenmäßigen Verhältnis dieser Zweige in den einzelnen Ländern entsprechen. Ein Ländervergleich wird deshalb interessant, weil im Zuge der Zulassungsbedingungen zu den Universitäten nicht nur die Frage nach einer unterschiedlichen Benotung, sondern vor allem auch nach einer unterschiedlichen Leistungsfähigkeit der Gymnasien in den verschiedenen Ländern aktuell geworden ist.

Was zunächst die Stundenzahl für die naturwissenschaftlichen Fächer anbelangt, so sind die Unterschiede zwischen altsprachlichen und neusprachlichen Zweigen sehr gering, doch bei entsprechender Fächerwahl wird in den altsprachlichen Gymnasien eine höhere Stundenzahl in Naturwissenschaften erreicht als in den neusprachlichen Gymnasien.

Größer ist die Stundenzahl für Naturwissenschaften in den mathematisch-naturwissenschaftlichen Gymnasien, allerdings fast ausschließlich in Physik, wo sie mehr als doppelt so hoch liegt wie in den beiden anderen Schultypen. Wenn es auch nicht möglich ist, zu ermitteln, wie hoch der Anteil der Schulen in einzelnen Leistungsbereichen in den einbezogenen Schulen ist, so darf doch vermutet werden, daß bei einer Relation von 47,8% neusprachlichen und 12,6% altsprachlichen und 37,1% mathematisch-naturwissenschaftlichen Gymnasien in der BRD die Zahl der in den in die Untersuchung einbezogenen Schulen erteilten Stunden in Physik zumindest so groß, wenn nicht größer ist als die der durchschnittlich erteilten Stunden in Biologie. (Biologie 11-12 Stunden in Gymnasien insgesamt, Physik in alt- und neusprachlichen Gymnasien 8-9 (+3 bei entsprechender Fächerwahl), in mathematisch-naturwissenschaftlichen Gymnasien 20). Trotzdem sind die Leistungen, wie übrigens auch in anderen Ländern, in Biologie erheblich besser als in Physik, und zwar für alle Bundesländer gleichmäßig. Wahrscheinlich ist dafür aber auch die Anlage des Tests verantwortlich zu machen, zieht man nämlich die Gelegenheit, dem Inhalt der Testaufgabe im Unterricht begegnet zu sein, wie sie von Lehrern geschätzt wurde, heran, so ergibt sich in Biologie (Population IV)weitaus die größte Möglichkeit (fast eine Standardabweichung) gegenüber einer negativen halben Standardabweichung in Chemie und einer Fünftel-Standardabweichung in Physik (errechnet aufgrund der durchschnittlichen Möglichkeit innerhalb der Bundesrepublik). Somit entsprächen die errechneten Standardabweichungen von 0,95 (Biologie), -5,50(Chemie) und 0,18(Physik) durchaus der Rangfolge der Leistungshöhe in diesen verschiedenen Bereichen. (Die Ergebnisse der Population II spiegeln ebenfalls diese Möglichkeiten wider, wobei die Höhe bei Physik 1,34, bei Biologie 0,84 und bei Chemie -1,12 beträgt.)

Die größere Homogenität der Population IV drückt sich auch in den Leistungen im Wortschatztest aus (s.Tab.13): Die Standardabweichung beträgt für die BRD 6,6 Punkte, während sie in der Population II 9,8 Punkte beträgt. Berechnet man die Standardabweichung aber als Prozentsatz des Mittelwertes, so wird die Streuung sehr viel deutlicher sichtbar. Sie beträgt: in Population I 73,9%, in Pop.II 61,6%, in Pop.IV 18,5% des Mittelwertes.
Die entsprechenden Werte für die Leistungen im naturwissenschaftlichen Test zeigen keinen so großen Rückgang der Streuung: Population I 49,7%, Pop.II 48,5%, Pop.IV 33,0%.
Es bestehen also trotz einer relativ homogenen allgemeinen schulischen Leistungsfähigkeit erhebliche Leistungsunterschiede in den naturwissenschaftlichen Leistungstests zwischen den Ländern. Bei einem Durch-

schnittswert von 26,9 Punkten und einer Standardabweichung von 8,9 Punkten fällt zunächst auf, daß die Standardabweichungen innerhalb der einzelnen Länder annähernd gleich sind, nur Niedersachsen hat eine größere Standardabweichung, d.h., die Leistungen sind in diesem Land am wenigsten homogen. Dabei ist Niedersachsen zugleich das Land mit dem niedrigsten Durchschnitt. Die Extreme stellen die Schulen mit einem durchschnittlichen Punktwert von 16,4 (7,4) und 29,9 (9,4) dar. Um einen Gesamtüberblick zu geben, stellen wir in Figur 4 für sieben Länder die Leistungsverteilung der Schulen graphisch dar.

Figur 4: Seite 54

Aus dieser Darstellung geht eindeutig hervor, wie unterschiedlich die Leistungssituation in den einzelnen Gymnasien ist und wie niedrig die Leistungen der schwächsten Schüler tatsächlich sind. Diese Darstellung besagt aber auch, daß es wenig Sinn hat, einzelnen Ländern bei den Zulassungen zu den Universitäten einen Bonus oder Malus zuzuschreiben, sondern daß vielmehr ein solcher Bonus oder Malus sinnvollerweise auf die einzelnen Schulen zu verteilen wäre: Ist es doch unmöglich, daß sich solche Leistungsunterschiede zwischen den Schulen nicht auch auf die Zeugnisgebung auswirken.

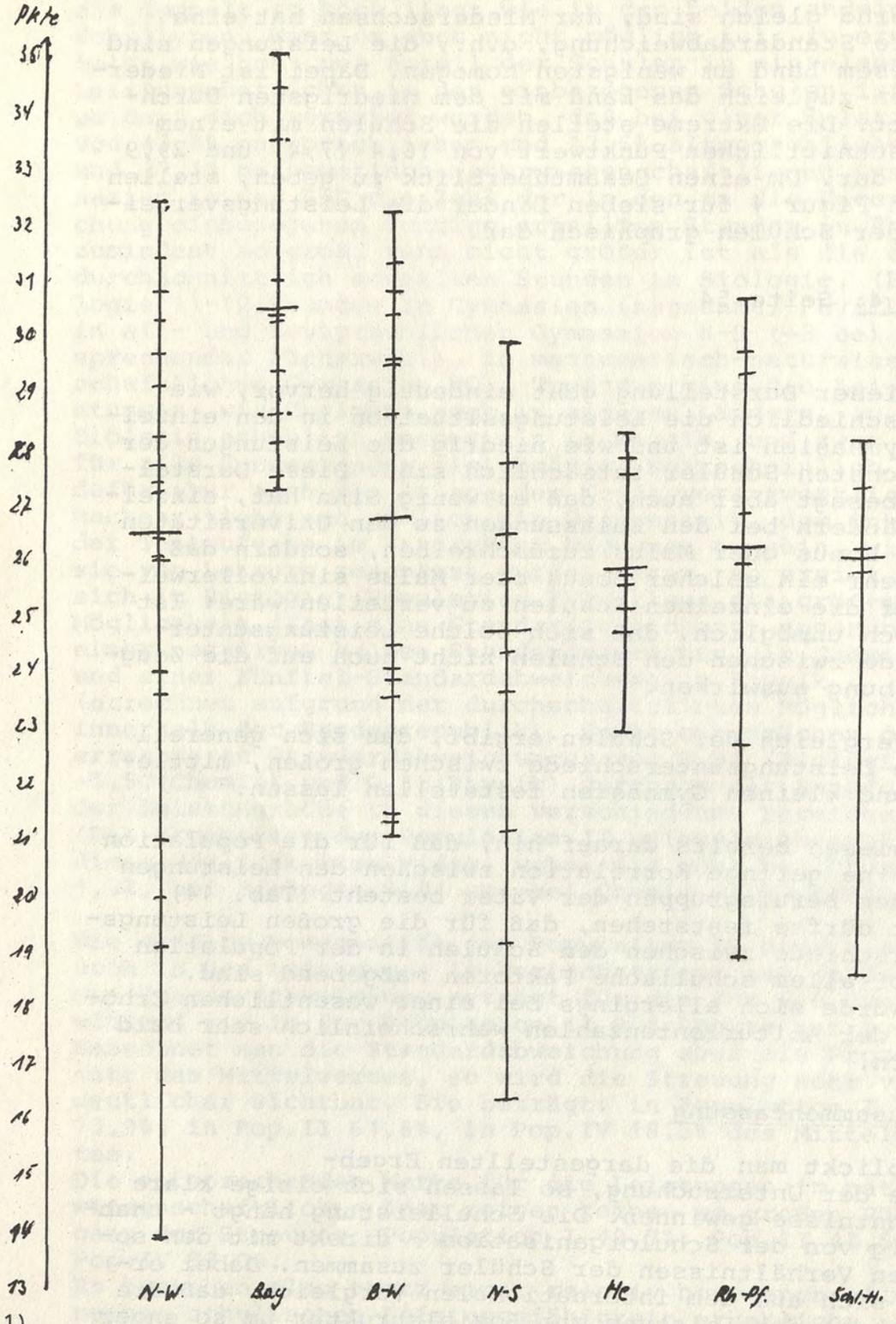
Ein Vergleich der Schulen ergibt, daß sich generell keine Leistungsunterschiede zwischen großen, mittleren und kleinen Gymnasien feststellen lassen.

Wir wiesen bereits darauf hin, daß für die Population nur eine geringe Korrelation zwischen den Leistungen und den Berufsgruppen der Väter besteht (Tab. 14). Damit dürfte feststehen, daß für die großen Leistungsunterschiede zwischen den Schulen in der Population IV vor allem schulische Faktoren maßgebend sind. Das würde sich allerdings bei einer wesentlichen Erhöhung der Abiturientenzahlen wahrscheinlich sehr bald ändern.

## 5. Zusammenfassung

Überblickt man die dargestellten Ergebnisse der Untersuchung, so lassen sich einige klare Erkenntnisse gewinnen: Die Schulleistung hängt - unabhängig von der Schulorganisation - direkt mit den sozialen Verhältnissen der Schüler zusammen. Dabei ergibt sich aus dem internationalen Vergleich, daß die Beziehung der Leistung zur Sozialstruktur um so enger ist, je selektiver das System ist. Die Unterschiede zwischen den Schulen sind mit zunehmendem Alter der Schüler in immer stärkerem Maße durch schulische Fak-

Figur 4: Leistungen der Schulen (Pop.IV), getrennt nach Bundesländern[1)]




1) Jeder Querstrich stellt die Durchschnittsleistung einer Schule dar; der längere Querstrich den Länderdurchschnitt. Ein Punkt eine weitere Schule mit der gleichen Punktzahl.

toren bedingt. Als wesentlich erwiesen sich dabei der Grad der Selektivität, der Anteil der Naturwissenschaften an der Gesamtstundenzahl, der Umfang der Hausaufgaben, die Qualität der Lehrerbildung (auf der Oberstufe), die Übereinstimmung der Testaufgaben mit den Lehrplänen der Schulen sowie das Geschlecht der Lehrer.

Ein geringerer Grad an Selektivität der Schulsysteme bedeutet nicht, daß die Hochleistungsfähigen in ihrer Leistung beeinträchtigt werden; der Prozentsatz der Leistungsfähigen ist unabhängig von der Organisationsform der Schulsysteme.

Die Geschlechtsunterschiede in der Leistung sind groß, beträgt doch die Leistung der Mädchen mit 18 Jahren 30% weniger als die der Knaben, wobei ihre Leistungen in Biologie günstiger sind als in den anderen Fächern.

Die Klassengrößen stehen nicht in positiver Korrelation mit der Leistung. Japan mit der höchsten Schülerzahl je Klasse (87% der Klassen haben mehr als 30 Schüler) hat z.B. bessere Leistungen als die Bundesrepublik mit nur 38% Klassen mit mehr als 30 Schülern.

Bei der Analyse der Lehrpläne ergab sich, daß zwischen den Ländern zwar Unterschiede hinsichtlich des Einsetzens und der Stundenzahl für den naturwissenschaftlichen Unterricht bestehen, daß aber hinsichtlich der Lerninhalte eine weitgehende Übereinstimmung besteht. Wenn sich trotzdem die Tests in allen Nationen als zu schwer erwiesen, so wahrscheinlich deshalb, weil die in den Lehrplänen aufgestellten Anforderungen zwar weitgehend übereinstimmen, aber keineswegs erreicht werden. Der Unterrichtserfolg bleibt also überall hinter den Erwartungen zurück.
Aufgrund der Ergebnisse ergibt sich für die Bundesrepublik im einzelnen:
Wenn auch allgemein die Stundenzahlen für den naturwissenschaftlichen Unterricht positiv mit der Leistung korrelieren, so zeigt Population II, daß das nur für gleiche Schultypen gilt, nicht aber für den Vergleich der Realschulen mit den Gymnasien. Auch die großen Leistungsunterschiede zwischen den Gymnasien in der Population IV können durch unterschiedliche Stundenzahlen nicht hinreichend erklärt werden. Es dürfte sich lohnen, den Unterschieden im einzelnen nachzugehen, zumal anzunehmen ist, daß solche Leistungsunterschiede nicht nur im Bereich der Naturwissenschaften bestehen, wie die Mathematikuntersuchung bereits einprägsam gezeigt hat. Da diese Leistungsunterschiede innerhalb der Länder bestehen und sehr viel größer sind als die Unterschiede in den Durchschnittswerten der einzelnen Länder, dürfte zumindest von den Naturwissenschaften her gesehen die Einrichtung eines Bonus bzw. Malus pauschal für einzelne Länder in keiner Weise gerechtfertigt sein.

Die im internationalen Vergleich sichtbar gewordenen relativ ungünstigen Ergebnisse des naturwissenschaftlichen Unterrichts in den Populationen I und II sollten Anlaß zu grundsätzlichen Überlegungen geben.

In der Population I dürfte eine stärkere Strukturierung des Sachunterrichts, wie sie sich bereits in neueren Lehrplänen anbahnt, durchaus wünschenswert sein. Auch die Einführung von Frachgruppenlehrern, wie sie in größeren Grundschulen möglich und vorgesehen ist, dürfte einer Hebung des Leistungsstandards in den Naturwissenschaften förderlich sein; denn man kann wohl kaum den Umfang und die Strukturierung der Arbeit im Sachunterricht dem jeweiligen Interesse des einzelnen Klassenlehrers überlassen, wie man früher geglaubt hat. Besonders die Ergebnisse in Japan zeigen, daß bei entsprechender Anlage des Unterrichts und bei entsprechender Vorbereitung der Lehrer ausgesprochen hohe Leistungen erzielt werden können.

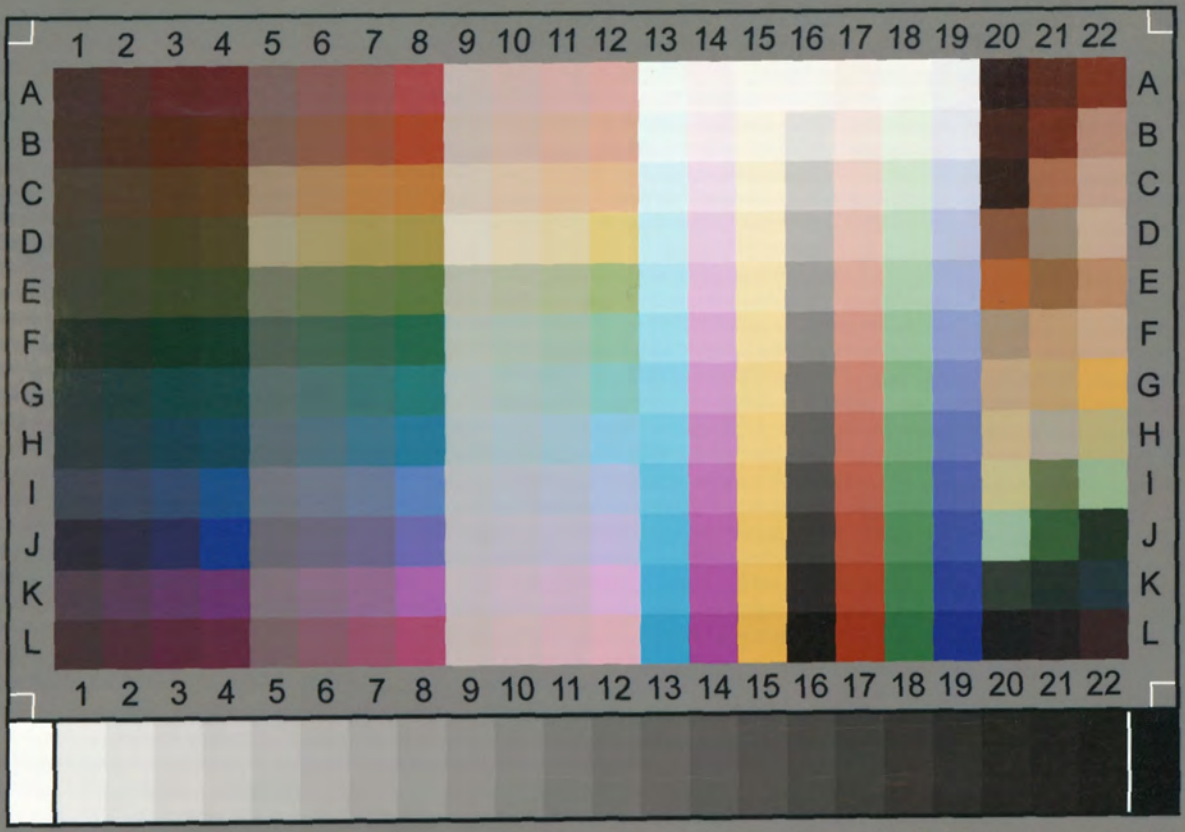
Wie weit die im internationalen Vergleich ungünstigen Ergebnisse der Population IV Anlaß zu einer Änderung unserer Lehrpläne, sei es hinsichtlich des Gewichts einzelner Fachgebiete, sei es hinsichtlich der Revision unserer Lehrpläne im Hinblick auf die schlechten Leistungen der Mädchen, geben sollten,muß den Verantwortlichen überlassen bleiben. Vielleicht ergibt sich bei den diesbezüglichen Überlegungen , daß es gar nicht wünschenswert ist, in allen Schultypen das Gewicht der Naturwissenschaften innerhalb der Gesamtstundenzahl zu verändern. Es dürfte jedoch sicher sein, daß bei uns die relativ schlechte Durchschnittsleistung nicht darin begründet ist, daß eine zu frühe Wahlfreiheit, wie z.B.in England, die Teilnahme einer größeren Zahl von Schülern am naturwissenschaftlichen Unterricht auf der Oberstufe verhindert.

In bezug auf die Anlage und Durchführung der Untersuchung ist kritisch anzumerken, daß nur das, was meßbar ist, in die Untersuchung eingeht, daß alle anderen Faktoren unzureichend oder überhaupt nicht erfaßt werden bzw. erfaßt werden können. Von daher hat die Methode als Methode vergleichender erziehungswissenschaftlicher Untersuchungen nur zusätzliche Bedeutung, kann aber keineswegs als einzig brauchbare Methode angesehen werden. Gerade für eine vertiefte Interpretation der Ergebnisse sind die traditionellen Methoden der vergleichenden Erziehungswissenschaft dringend notwendig.

Die 400 Prädiktoren, die in der Untersuchung berücksichtigt wurden, sind zum Teil für detaillierte Untersuchungen zu grob, weil sie für die große Zahl der teilnehmenden Länder passen mußten, um eindeutig skaliert werden zu können. Es dürfte sich empfehlen,

bei zukünftigen Untersuchungen nicht Länder mit einer zu unterschiedlichen Bevölkerungs- und Schulstruktur in die gleiche Untersuchung einzubeziehen, da dann immer die Gefahr besteht, daß keine einheitlichen Kategorien für die Prädiktoren aufzustellen sind.
Ferner ist festzustellen: Die Werte für die Prädiktoren wurden zum großen Teil aus den Antworten der Schülerfragebogen gewonnen. Damit sind naturgemäß Zweifel an der Objektivität ihrer Beantwortung gegeben. Aussagen über die Art des Unterrichts, über den Umfang von experimentellen Übungen, über das Klassen- und Schulklima sind zu sehr subjektiv bestimmt; sie können nur durch die große Zahl einigermaßen gerechtfertigt werden.

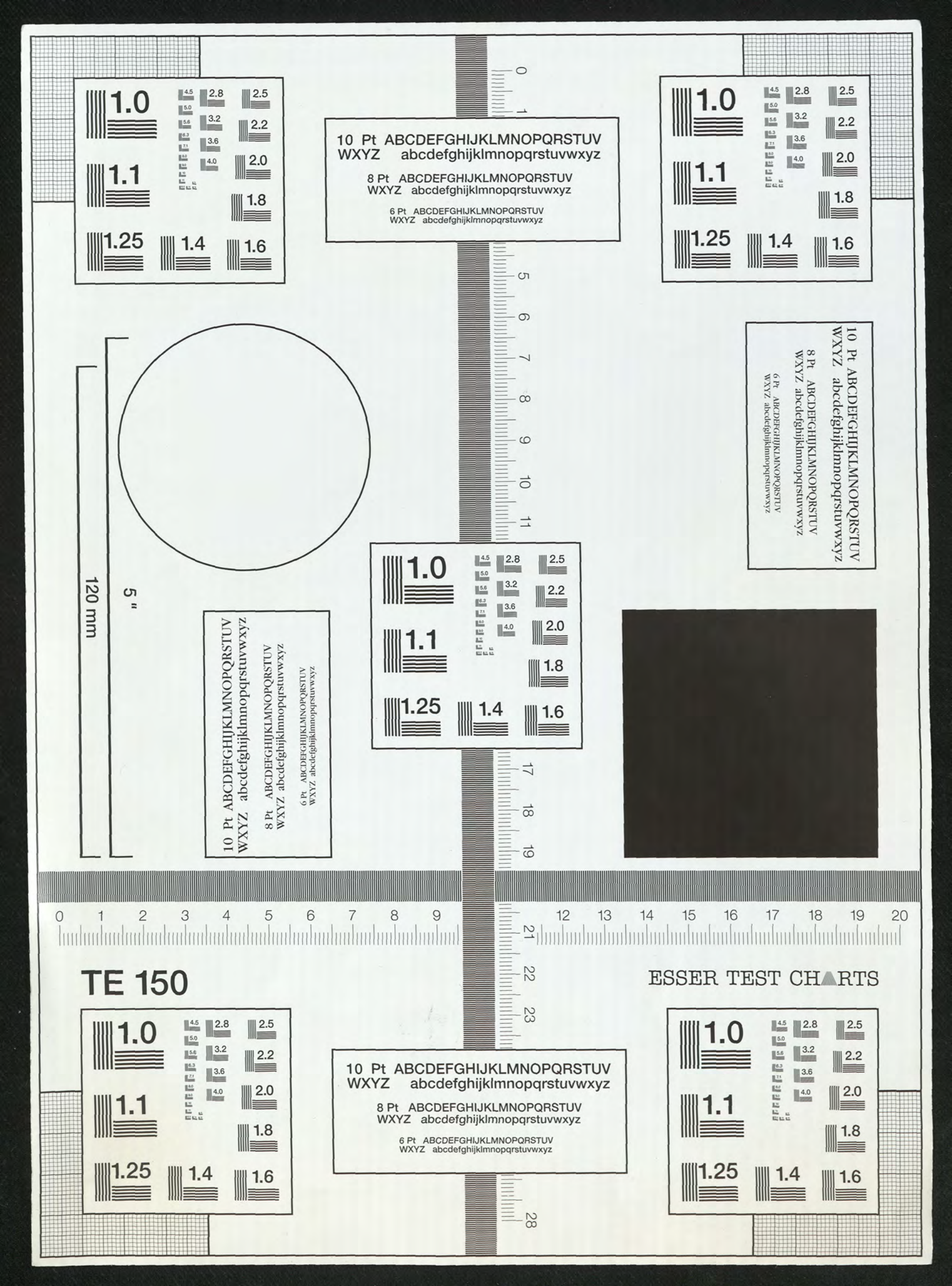
Schließlich ist zu fragen, ob eine Untersuchung dieses Ausmaßes, die sich auf 22 Länder und 6 Fächer zugleich erstreckt, nicht eine zu aufwendige Organisation verlangt. Wenn sich sämtliche Kosten der vorliegenden Untersuchung auf rund 5 Millionen US Dollar belaufen (wovon ein großer Teil noch zu stabilen Dollarkursen zur Verfügung gestellt wurde), wobei allerdings die nationalen Ausgaben, die für die Bundesrepublik DM 230.000,-- betragen, eingeschlossen sind, so muß man sich fragen, ob die Ergebnisse in einem rechten Verhältnis zu den Kosten stehen. Es hat sich eine kostenaufwendige internationale Organisation neben den bestehenden Organisationen wie UNESCO, OECD und EUROPARAT mit ihren Forschungseinrichtungen entwikkelt, die sicherlich auf die Dauer nicht von Beiträgen einzelner Länderregierungen und von Stiftungen getragen werden kann, sondern von allen beteiligten Ländern mitfinanziert werden muß. Das allerdings dürfte dann wahrscheinlich das Ende dieser Arbeit bedeuten. Es ist zu fragen, ob nicht der große Ertrag dieser Anstrengungen in 22 Ländern, der ja nicht nur in den veröffentlichten bzw. zur Veröffentlichung vorgesehenen 9 Bänden niedergelegt ist, sondern aus einer Fülle von Erfahrungen besteht, die für künftige internationale Untersuchungen in keinem Fall verlorengehen sollten, in anderer Weise gesichert werden kann. Dabei wäre vor allem an Untersuchungen zwischen weniger Ländern zu denken, die in bezug auf die jeweilige Fragestellung besonders ausgewählt werden müßten. Bestehende, finanziell gesicherte Forschungsinstitute großer internationaler Organisationen könnten Träger solcher Untersuchungen sein.
Die Ziele der I.E.A., eine möglichst erdballumspannende, unabhängige, vergleichende Untersuchung durchzuführen, sind vielleicht doch zu hoch gesteckt gewesen.




IT8.7/2-1993
2013:02

Printed on Kodak Professional Paper - Made by Wolf Faust (www.coloraid.de)

Charge: R130227


1.0
1.1
1.25
1.4
1.6
2.8
2.5
3.2
2.2
3.6
2.0
4.0
1.8
4.5
5.0
5.6
6.3
7.1
8.0
10 Pt ABCDEFGHIJKLMNOPQRSTUV
WXYZ abcdefghijklmnopqrstuvwxyz
8 Pt ABCDEFGHIJKLMNOPQRSTUV
WXYZ abcdefghijklmnopqrstuvwxyz
6 Pt ABCDEFGHIJKLMNOPQRSTUV
WXYZ abcdefghijklmnopqrstuvwxyz
120 mm
5 "
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 28
TE 150
ESSER TEST CHARTS