

II. Untersuchungsdaten

Im Jahre 1933 machte C.G. Wrenn bei der Untersuchung von Erfolg und Mißerfolg im Hochschulstudium die Feststellung, daß bei der Gruppe, auf die sich seine Erhebung erstreckte, die Unterschiede zwischen den Mittelwerten der zu den diversen Begabungstests erhaltenen Resultate keineswegs signifikant waren. Er stellte die Hypothese auf, daß der Lernerfolg in erster Linie von der Arbeitsmethode abhängig sei¹¹⁰.

Wrenn führte schließlich eine Vergleichsstudie derart durch, daß er zwei Gruppen untersuchte, deren Schulresultate zwar auf gleichem Niveau lagen, bei denen jedoch die Begabungen deutlich differierten, und wies erneut auf die entscheidende Rolle der Lerngewohnheiten hin.

H. C. Carter errechnete seinerseits die Korrelationen zwischen den Arbeitsmethoden — gemessen aufgrund seines Fragebogens (siehe weiter unten) und der Intelligenz (Henmon-Nelson-Test) einerseits, und den Schulerfolgen andererseits:

Korrelation Arbeitsmethode — Intelligenz .36

Korrelation Arbeitsmethode — Schulerfolge .57.

Die Bedeutung derartiger Ergebnisse wird offenbar nicht klar erkannt. Sie könnten ein Hinweis darauf sein, daß in Relation zu dem Test von Carter der Henmon-Nelson-Test nicht bestimmten Aspekten der Intelligenz oder der Persönlichkeit Rechnung trägt, die für den Schulerfolg wesentlich sind.

Aber wie auch immer die Erklärung sein mag, der erhöhte Aussagewert des *Study Methods Survey* von Carter scheint sichergestellt zu sein.

III. Einige Tests

Die Arbeitsgewohnheiten und -methoden eignen sich schlecht für die direkte Beobachtung. Aus diesem Grunde verwendet man bis heute vornehmlich die Fragebogen, Kontrolllisten und *rating scales*.

Nach unserer Meinung könnten die im Folgenden beschriebenen Instrumente leicht verbessert werden; sie erlauben keine hinreichend genaue Analyse der Handlungsschritte des Kindes beim Erlernen eines Unterrichtsstoffes

Um eine brauchbare Validität dieser Tests zu gewährleisten, müßte man darüber hinaus Untersuchungsgruppen bilden, bei denen unter gleichen Ar-

110 Siehe C.G. Wrenn, *Scholarship and Habits of Work* (6th Convention of the Pacific Coast Assoc. of Collegiate Registrars, 1933, S. 26—29).

beitsbedingungen sowie identischen Voraussetzungen der Begabung und der Persönlichkeit lediglich die Schulerfolge variierten. Man weiß, wie schwierig es ist, eine solche Übereinstimmung in den Ausgangsbedingungen zu erzielen.

Study-Habits Inventory, Neubearbeitung 1941, von C. G. Wrenn (Stanford Univ. Press)

Gewichtete Kontrollliste, anwendbar mit Beginn des höheren Schulunterrichts. Der Schüler wird aufgefordert, Fragen zu beantworten, und zwar nicht im Hinblick darauf, was er tun zu müssen glaubt, sondern im Hinblick auf seine tatsächlichen Lerngewohnheiten.

Die folgenden Itembeispiele geben gleichzeitig Aufschluß über die diversen erforschten Gebiete.

	Gewichtete Punktwerte		
	Niemals selten	Manch- mal	Häufig immer
A. <i>Technik im Lesen und Mitschreiben während des Unterrichts</i>			
Entgehen Ihnen im Verlauf des Unterrichts wichtige Ausführungen, während Sie sich Notizen machen?	9	-7	0
B. <i>Konzentration</i>			
Ich kann mich schwer auf das, was ich lerne, konzentrieren. Wenn ich fertig bin, weiß ich nicht mehr, was ich gerade gelesen habe.	8	-5	-7
C. <i>Zeiteinteilung — Zusammenarbeit/Hilfe</i>			
Für bestimmte Aufgaben nehme ich mir zuviel Zeit und für andere wiederum zu wenig.	6	5	-11
Ich lerne am liebsten mit einem anderen zusammen.	5	-4	0
D. <i>Allgemeine Gewohnheiten und Einstellungen</i>			
Während einer Prüfung oder eines Abhörens bilde ich mir in Gedanken meine Antwort, bevor ich zu schreiben beginne.	-13	0	7
Ich kann nur lernen, wenn ich dabei rauche.	7	-5	-7

Die negativen Punktwerte erfordern jeweils eine spezielle Untersuchung und eventuell eine Korrektur der Methode oder Lerngewohnheit. Der Gesamtpunktwert erlaubt den Vergleich mit anderen Schülern.

California Study Methods Survey von H. D. Carter, 1958 (California Test Bureau)

Dieser Test ist für den höheren Schulunterricht bestimmt und soll folgendes untersuchen:

1. die Einstellung gegenüber der Schule,
2. die Lerntechnik,
3. die Planung und die Systematisierung.

Man erstellt drei partielle Punktwerte und einen Gesamtwert. Darüber hinaus erlaubt eine Verifizierungspunktzahl die Ausschaltung eines bestimmten prozentualen Anteils zweifelhafter Antworten (bewußtes Fuschen, Mangel an Aufrichtigkeit).

Itembeispiele:

- | | | |
|---|----|------|
| | Ja | Nein |
| 1. Sind Sie zufrieden mit den Noten, die Sie erhalten? | | |
| 22. Haben Sie den Eindruck, daß Sie imstande sind, alles, unabhängig von dem Schwierigkeitsgrad zu erlernen, wenn Sie sich genügend damit befassen? | | |
| 61. Versuchen Sie, sich während der Erarbeitung einer Lektion vorzustellen, welche Fragen der Lehrer stellen könnte? | | |
| 96. Richten Sie es sich in der Regel so ein, daß Sie sich eine Zeit am Tag für solche Dinge freihalten, die Sie gerne tun? | | |
| 114. Müssen Sie sich häufig im Augenblick eines Examins bei anderen einen Bleistift oder einen Radiergummi ausleihen? | | |
| 128. Versuchen Sie beim Lernen einer Lektion, durch Abfragen derselben selbst zu kontrollieren, inwieweit Sie mit Erfolg gearbeitet haben? | | |

Dieser Test ist in seiner Durchführung leichter und sicherer zu handhaben als der von Wrenn, der in seinem Charakter subjektiver ist.

Brown-Holtzman SSHA, *Survey of Study Habits and Attitudes, Form C* (College Form), 1965 (Psychological Corporation).

Diese ganz neue Bearbeitung ist das Resultat einer 8jährigen Forschungsarbeit.

Der folgende Test erstreckt sich auf die Messung der Gewohnheiten, Notizen aufzunehmen und Nachschlagewerke zu benutzen, Techniken (*skills*) also, die als grundsätzlich bekannt vorausgesetzt werden und daher vielfach gleichermaßen untersucht werden wie die anderen in der Schule erworbenen Kenntnisse und Fähigkeiten. Dennoch ist es verständlich, zumal das methodische Arbeiten vornehmlich die geschulte Handhabung dieser Techniken unterstellt, daß bestimmte Autoren dieses Gebiet unter dem Sachgebiet, das uns beschäftigt, zu erörtern wünschen. Um diese Tendenz zu unterstreichen, weisen wir hier auf den Test von Spitzer hin.

Spitzer Study Skills Test von H. H. Spitzer, 1956 (World Book Co.)

Für den höheren Schulunterricht und den Beginn des Hochschulstudiums umfaßt die Prüfung fünf Subtests:

1. Benutzung des Wörterbuchs: 26 Items

Unter Hinzuziehung einer bestimmten Anzahl von Begriffen, die dem *American College Dictionary* entnommen und in diesem Test wiedergegeben sind, muß der Schüler Fragen beantworten, die sich auf die Bedeutung, die Orthographie, die Herkunft, die Betonung usw. erstrecken.

Beispiel: *Narkolepsie, die, Pathol.*: Zustand, der durch ein unkontrollierbares Schlafbedürfnis oder kurze Schlafattacken zu jeder beliebigen Gelegenheit gekennzeichnet ist.

Frage: Die Bezeichnung *pathol.* in Verbindung mit dem Begriff *Narkolepsie* bedeutet:

- daß es sich um ein Wort handelt, das sich auf Krankheiten bezieht,
- daß eine Beziehung zwischen *Narkolepsie* und *Pathos* besteht,
- daß das Wort *Narkolepsie* aus dem Griechischen stammt,
- daß die Symptome stets in der gleichen Reihenfolge auftreten (Anm. des Übersetzers: Dieser letzte Vorschlag ist eine Anspielung auf die Ähnlichkeit der englischen Wörter *pathos* und *path*, eine Sprachfinesse, die man nicht ins Deutsche übertragen kann).

2. Verwendung von Nachschlagewerken: 26 Items

Überprüfung der Fähigkeit, bibliographische Verzeichnisse, Kataloge, Enzyklopädien, Atlanten, Jahrbücher usw. zu verwenden.

Beispiel: Im Rahmen des kunstgeschichtlichen Unterrichts behandeln Sie die zeitgenössischen Maler, und Sie müssen etwas über den mexikanischen Maler *Diego Rivera* erzählen. Unter welcher Rubrik des bibliographischen Verzeichnisses müssen Sie nachsehen, um am schnellsten zu den gewünschten Informationen zu gelangen?

- Rubrik: Acc-Cou.
- Rubrik: Lit-Min.
- Rubrik: Nab-Pru.
- Rubrik: Qui-Sac.

3. Verstehen von graphischen Darstellungen, Tabellen und Landkarten: 42 Items

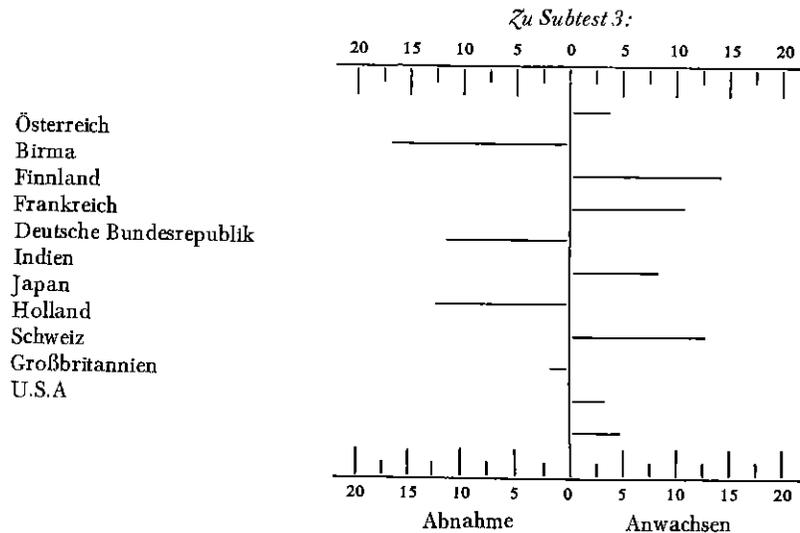
Beispiel: Geben Sie anhand der folgenden graphischen Darstellung an, in welchen Ländern sich die Lebenshaltungskosten am stärksten verändert haben.

(Siehe Abb. S. 176.)

4. Der sachgerechte Umgang mit Nachschlagewerken

Beispiel: Sie möchten wissen, ob *Murphy* — Nordkarolina — an einer Eisenbahnlinie liegt. Welches Nachschlagewerk ist hierfür am geeignetsten:

- Ein Atlas?
- Ein Globus?
- Eine Enzyklopädie?
- Eine Sammlung von Eisenbahnfahrplänen?



Veränderung der Lebenshaltungskosten (in %) Juli 1949 – Juli 1955.

5. Fähigkeit in Aufnahmen von Notizen (Erkennen des Wesentlichen)

Anhand von kurzen Texten, bei denen die jeweils zugrunde liegenden Ideen nummeriert sind, muß der Schüler nach einem festgelegten Schema eine synoptische Übersichtstabelle aufstellen.

7. Die Vorhersage und Bewertung der Leistungsfähigkeit der Lehrer

Ein absolutes Urteil über einen Lehrer abzugeben, ist praktisch unmöglich. Seine Leistungsfähigkeit hängt nicht ausschließlich von seiner Person ab (interne Faktoren), sondern auch von den Schülern, denen er gegenübersteht, sowie von der jeweiligen Situation und den äußeren Umständen, in die er hineingestellt ist – (externe Faktoren).

Der Unterrichtsakt stellt stets eine komplexe Funktion dar, wobei eine Vielzahl von Relationen zwischen dem Unterrichtenden und jedem einzelnen Schüler sowie zwischen dem Lehrer und der gesamten Klasse oder Teilen dieser Gruppe usw. zu berücksichtigen sind. Kurz – die einfachste Beziehungsform: ein Lehrer – ein Schüler läßt schon ein unermeßliches Geflecht von Interaktionen entstehen.

Zudem ist der Wert eines Lehrers relativ und variabel. Der eine Schüler erreicht unter seiner Anleitung gute Erfolge, und mit einem anderen gleichaltrigen Schüler kommt er praktisch zu gar keinem Ergebnis; die Jüngsten sind begeistert von ihm, während die Älteren ihn langweilig finden; sein Unterricht in dem einen Fach ist glänzend, in einem anderen hingegen recht mäßig; an einem Tag hält er eine hinreißende Stunde, am folgenden Tag dagegen nur eine mittelmäßige.

Das *teaching* scheint immer mehr ebenso schwer begreiflich zu sein wie das *learning*, was übrigens nicht voneinander zu trennen ist¹¹¹. Es überrascht daher keineswegs, daß sich bis heute fast alle Versuche der Erstellung von objektiven Untersuchungsverfahren als eine Enttäuschung erwiesen. So haben die den Schulräten vorgeschlagenen Kontrolllisten und *rating scales* häufig nur eine illusorische Sicherheitsgewähr geliefert, weil keineswegs erwiesen war, daß die Qualitäten, auf die sie sich bezogen, gültige Kriterien darstellten; und die Faktorenanalyse in sich ist keine authentische Garantie.

Dennoch wäre es nicht gerechtfertigt, wollte man die uns gegenwärtig zur Verfügung stehenden technischen Hilfsmittel schlechterdings ablehnen. Diese Untersuchungsverfahren lassen sich vervollkommen, und in der Hand gut ausgebildeter Beobachter erlauben sie Sondierungen, die in einem nicht zu übersehenden Maße geeignet erscheinen, gewisse Meinungen richtigzustellen. Überdies sind die Fortschritte in der pädagogischen Forschung zur Zeit indiskutabel.

Schließlich erfordern die unternommenen Versuche ein Bemühen zur Klärung: Um die Verhaltensweisen eines Erziehers zu messen, muß man sie zuvor untersucht haben!

Ohne den Anspruch zu erheben, das Problem erschöpfend zu behandeln, wollen wir im Folgenden die Tendenz der objektiven Forschung aufzeigen. Mehrere amerikanische Untersuchungsverfahren, die hier zur Diskussion stehen, sind nicht unmittelbar auf europäische Verhältnisse übertragbar, sie können jedoch eine Ausgangsgrundlage für neue Arbeiten darstellen.

Unter einer Reihe anderer Vorschläge scheint uns die von H. E. Mitzel¹¹² vorgenommene Klassifizierung der allgemeinen Kriterien am meisten befriedigend zu sein. Er unterscheidet:

111 Siehe: N. L. Gage, Herausgeber, Handbook of Research on Teaching, Chicago, Rand McNally & Co., 1963, 1218 Seiten. Dieses Werk weist auf den Umfang des noch zu erforschenden Arbeitsgebietes hin.

112 H. E. Mitzel, Teacher Effectiveness (Enc. of Educ. Res., op. cit., S. 1482–1484).

1. *Die Kriterien aufgrund des Ergebnisses:* Man mißt die Qualifikation des Lehrers an den Auswirkungen seines Unterrichts.
2. *Die Kriterien der mutmaßlichen Erwartung:* Nach allgemeinem und sachlichem Ermessen ist anzunehmen, daß gewisse Züge oder Fähigkeiten es erlauben, einen guten Unterricht vorauszusagen. So unterstellt man z. B., daß ein hochintelligenter Lehrer sich als ein besserer Pädagoge erweist als ein weniger begabter Lehrer.
3. *Die Kriterien aufgrund der Unterrichtsverfahren oder -funktionen:* Man sucht nach Gesichtspunkten der Verhaltensweisen von Lehrern und Schülern, die in Verbindung mit der Wirksamkeit des Unterrichts in Erscheinung treten. So scheint es beispielsweise, daß ein Lehrer, der klar und deutlich erklärt und seine Schüler vielfach ermuntert, größere Erfolgchancen hat als ein zerstreuter und nörgelnder Lehrer.

I. Die Kriterien aufgrund der Ergebnisse

Ein Unterricht findet seine Wertbestätigung in den fruchtbaren Auswirkungen auf die Schüler, und in Übereinstimmung mit Remmers und Gage kann man die Hypothese aufstellen, daß man „denjenigen Erzieher als den besten bezeichnen kann, dem es gelingt, bei seinen Schülern ein optimales Maß an wünschenswerten Veränderungen zu bewirken“¹¹³.

Dennoch läßt sich dieses Höchstmaß an Veränderungen leichter wünschen als tatsächlich erreichen.

Handelt es sich um Schulerfolge?¹¹⁴ Sie sind ziemlich leicht meßbar, hängen allerdings nicht ausschließlich von dem Wert des Lehrers ab. Die Leistungsfähigkeit der Schüler, ihre Gesundheit, ihre vorausgegangene Ausbildung spielen hier mit, ohne daß man den prozentualen Anteil dieser Einflußfaktoren genau bestimmen könnte.

Handelt es sich um eine Verbesserung der Arbeitsmethode, des logischen Denkens, um einen Anreiz zu intellektueller Neugierde oder auch um eine größere geistige Aufnahmebereitschaft?

Handelt es sich schließlich um eine Persönlichkeitsentfaltung, um ein besseres inneres Gleichgewicht oder um eine ständig zunehmende Sicherheit?

Alle diese Effekte und noch manch andere sollten gleichzeitig gemessen wer-

¹¹³ H. Remmers u. N. Gage, *Educational Measurement and Evaluation*, New York, Harper, 1955, S. 479.

¹¹⁴ Siehe in diesem Zusammenhang: L. Heil u. C. Washburne, *Brooklyn College Research in Teacher Effectiveness* (*Journal of Ed. Res.*, Nr. 55, Mai 1962, S. 347–351).

den. Zudem müßte man auch zahlreichen internen und externen Faktoren Rechnung tragen, auf die wir weiter oben bereits hingewiesen haben.

Die Vielfalt der genauer zu untersuchenden Aspekte¹¹⁵ ist eine Erklärung dafür, daß nur wenige qualifizierte Forscher den Versuch unternommen haben, bei der Beurteilung der Erzieher das Ergebnis ihres Unterrichts als Kriterium zu nehmen, und abgesehen von einer Serie von Arbeiten, deren naive Vereinfachung des Problems sie ihrer wissenschaftlichen Bedeutung beraubt, verfügen wir in dieser Richtung bisher noch über keine bedeutenden Untersuchungen.

Die in der Psychologie und in der Pädagogik erzielten Erfolge, die verfeinerte Form der experimentellen Schemata und die Verwendung der elektronischen Datenverarbeitung lassen jedoch für die nahe Zukunft auf ein schnelles Voranschreiten hoffen.

II. Die Kriterien der mutmaßlichen Erwartung

A. Allgemeines

Wie der Name bereits andeutet, werden diese Kriterien vornehmlich zur Vorhersage der Leistungsfähigkeit der Lehrer angewendet. Im allgemeinen werden die folgenden charakteristischen Merkmale untersucht¹¹⁶.

1. Konstitutionelle Voraussetzungen

Eine gute Gesundheit, eine entsprechende äußere Erscheinung und eine sympathische Stimme scheinen in erster Linie die wünschenswerten Qualifikationen zu sein. Evans weist mit Recht darauf hin, daß das Kriterium „Gesundheit“ im allgemeinen nur den Verantwortlichen des Unterrichts vorbehalten bleibt, während die Schüler (vor allem die Mädchen) die äußere Erscheinung und die Stimme registrieren. Die beiden letzten Kriterien werden jedenfalls in den amerikanischen Untersuchungen häufig erwähnt.

Tatsächlich aber scheinen, abgesehen von einem Minimum an offenbar notwendigen Voraussetzungen, die physischen Merkmale keinen so bedeutenden Einfluß auf den Unterrichtserfolg auszuüben.

¹¹⁵ Erinnern wir uns, daß, wenn man die Interaktionen berücksichtigt, zwei experimentelle Faktoren, die sich auf fünf charakteristische Merkmale beziehen, schon $2^5 = 32$ Beobachtungen erfordern.

¹¹⁶ Zur eingehenden Erörterung dieser Kriterien siehe: K. M. Evans, *Research on Teaching Ability* (*Educational Research*, Bd. I, Nr. 3, Juni 1959).

2. Intelligenz und Kenntnisse

- a) Resultate der Begabungs- und Wissenstests, wobei die Tests, die sich auf die Psychologie und die Pädagogik erstrecken, mit einbegriffen sind¹¹⁷.
- b) Schulergebnisse, insbesondere solche, die aus den Hausaufgaben und vorbereitenden Übungen direkt auf den Unterricht hin resultieren.
- c) Spezialisierte Berufskennntnisse und Bemühungen um Vervollkommnung (Arbeitstagungen, bzw. Praktika, Fortbildungskurse usw.).

Diese Faktoren üben gewiß einen Einfluß aus, aber auch dieser ist noch nicht klar definiert. Somit schwankt die Korrelation zwischen den Schulresultaten und den der Unterrichtspraktik zugeordneten Notenwerten zwischen 0,2 und 0,6 (Evans). Ryans bemerkt hierzu, daß die Erzieher, die ihr Studium mit Auszeichnung bestanden haben, im allgemeinen den anderen überlegen sind, mit Ausnahme im Hinblick auf Faktoren emotionaler Stabilität.

3. Einstellungen und Interessen

Der Bereich der Einstellungen ist bislang noch wenig erforscht; was das Interesse am Unterrichten betrifft, so scheint dies kein besonders sicheres Kriterium der Vorhersage zu sein.

Die Beurteilung der Lehrer mit verschiedenartigen Interessen steht unter einem günstigen Vorzeichen. Sie haben im allgemeinen eine bessere Auffassungsgabe, sind interessanter und geeigneter; ihre Unterrichtsführung (die Schüler lassen sich von ihnen lieber disziplinarisch beeinflussen) ist freundlicher (Ryans).

4. Emotionale und soziale Anpassung

Es gibt keinen „Universal-Lehrertyp“. Leichte innere Unausgeglichheiten beeinträchtigen noch nicht unbedingt die Qualität des Unterrichts. Auf jeden Fall steht fest, daß die Persönlichkeit des Erziehers auf den Schüler einen direkten und manchmal tiefgreifenden Einfluß ausübt¹¹⁸.

117 Der amerikanische *Educational Testing Service* (Princeton) bietet jedes Jahr eine speziell für diesen Zweck entworfene Testbatterie an: National Teacher Examination (NTE).

118 Sie beeinflusst auch seine Schulresultate. Christensen hat festgestellt, daß eine signifikante Relation zwischen der vom Lehrer bekundeten Sympathie, zwischen dem Grad menschlicher Kontakte und den Fortschritten im Wortschatz und im Rechnen besteht. Siehe: *Journal of Ed. Psychology*, Juni 1960, Nr. 51, S. 169–174.

5. Ansehen und Berufsstatus

B. Einige Tests

Die Zahl der Tests, die sich auf Kriterien der Vorhersage erstrecken, ist nicht sehr groß. Wir halten es für sinnvoll, die wichtigsten ziemlich ausführlich darzulegen, weil sie allgemein wenig bekannt sind.

Zwei Vorbemerkungen erscheinen jedoch angebracht:

1. Man darf einen klassischen Einwand nicht übersehen: Eine „korrekte“ Beantwortung solcher Tests gibt keineswegs einen sicheren Hinweis auf die eigentlichen Verhaltensweisen.
2. Keines der angeführten Untersuchungsverfahren scheint vollkommen zu sein, sei es, daß die Konstruktion von mittelmäßiger Qualität ist, sei es, daß die Eichungen oder die Validitätsuntersuchungen unzureichend sind.

1. *How I Teach* von I. Kelley und K. Perkins (Educational Testing Bureau), 1941

Skalen zur Messung der theoretischen und praktischen Kenntnisse der Lehrer auf dem Gebiet der Psychologie aufgrund ihrer Urteilsfähigkeit in bezug auf bestimmte Handlungen oder Situationen.

Beispiele:

Einem Schüler, der lügt, mit Strafe zu drohen:

- ist bestimmt gut;
- ist vermutlich gut;
- hat einen zweifelhaften Wert;
- ist vermutlich unangebracht;
- ist bestimmt unangebracht.

Einem Schüler, der sich in der Klasse schlecht beträgt, zusätzliche Hausaufgaben zu geben:

- die gleichen fünf Beantwortungsmöglichkeiten.

2. *Educational Aptitude Test* von T. Hunt und J. Fox (Center for Psychological Service), 1940

Dieses Verfahren wurde ausgearbeitet, um die Studierenden bei ihrem Eintritt in das Pädagogische Institut der George Washington University zu testen. Die Korrelation mit den im Laufe der Studienzeit erzielten Lernergebnissen soll bei .51 und mehr (was uns nichts über die Leistungsfähigkeit des Probanden im eigentlichen Unterricht vorauszusagen erlaubt) liegen. Die Prozentrangnormen geben keine genaue Auskunft über die untersuchten Populationsschichten.

Der Test umfaßt sechs Subtests :

1. *Wortschatz* (Wörter gleicher und entgegengesetzter Bedeutung)

2. *Test zur Messung des Urteilsvermögens* im Hinblick auf die pädagogische Forschung.

Beispiel: Sie wollen eine Untersuchung anstellen über das Problem: „Erleichtert ein vorausgegangener Lateinunterricht das spätere Erlernen der französischen Sprache?“ Vierhundert Schüler der höheren Schule, die seit einem Jahr französischen Unterricht gehabt haben, können an der Untersuchung teilnehmen. Streichen Sie unter den folgenden Punkten fünf an, die Ihnen für die Durchführung der Arbeit am wichtigsten erscheinen:

- Über eine Schülergruppe verfügen, die im 1. Jahr Französischunterricht hat;
 - Über eine Schülergruppe verfügen, die Lateinunterricht hat;
 - Über die am Ende des Jahres für alle Schüler erteilten Französischnoten verfügen;
 - Den Prozentsatz derjenigen Schüler kennen, die Französisch und Latein gelernt haben;
 - Den Leistungsstand aller Schüler in allen Unterrichtsfächern kennen;
 - Über eine Aufstellung verfügen, auf der alle Schüler eingetragen sind, die Latein gelernt haben;
 - Über eine Aufstellung verfügen, die alle Sprachen angibt, die die Schüler gelernt haben;
- usw.

3. *Logisches Denken*

Beispiel: Geben Sie an, ob die Schlußfolgerung richtig oder falsch ist:

Viele Studierende, die beim Hochschulstudium Schwierigkeiten haben, sind „slow readers“ (langsame Leser).

Dieser Student ist ein „slow reader“, folglich kann er kein erfolgreiches Hochschulstudium absolvieren.

4. *Information* (Richtig-Falsch)

Beispiel: Sich in Gegenwart seiner Kameraden über einen Schüler lustig zu machen, ist eine Form der Bestrafung, von der abzuraten ist.

5. *Textverständnis*

6. *Rechnen:* Einige Rechenarten, die in Richtung auf die Pädagogische Psychologie tendieren.

Beispiel: Errechnung eines Intelligenzquotienten anhand der Formel.

3. *Teaching Aptitude Test* von F. Moss, T. Hunt und F. Wallace, gekürzte Ausgabe (Center for Psychological Service), 1927

Test zur Prüfung von Studierenden in Pädagogik und zur Lehrerauswahl. Sehr allgemeine Normen (Medianwert und Quartile) werden erstellt für die Pädagogischen Institute der Universität, für Pädagogische Hochschulen, erfahrene Pädagogen und für Schüler der Oberstufe des höheren Schulwesens.

1. *Beurteilung von Schulsituationen:* Was ist zu tun?

Beispiel: Ein Kind ist sehr schüchtern und äußerst befangen, wenn es eine Aufgabe vorzutragen muß. Der Lehrer soll:

- das Kind jedes Mal, wenn es etwas vorträgt, ermutigen;
- das Kind dazu zwingen, vor seinen Kameraden vorzutragen;
- dem Kind erlauben, nach der Schule allein vorzutragen;
- dem Kind erlauben, seine Aufgabe vorzutragen, ohne seinen Platz zu verlassen.

2. *Urteilsvermögen und Information über pädagogische Probleme* (Richtig-Falsch)

Beispiel: Alle Kinder bringen von Geburt die gleichen geistigen Voraussetzungen mit;

Unaufmerksamkeit ist häufig ein Zeichen von schlechter Gesundheit.

3. *Verständnis und Gedächtnis*

Aus dem Gedächtnis auf 20 Fragen antworten (Richtig-Falsch), die sich auf einen vor dem Subtest 1 und 2 gelesenen Text beziehen.

Beispiel: Man beruft sich auf die Psychoanalyse, um Sprachfehler zu behandeln, die auf emotionale Störungen zurückzuführen sind.

Jemand, der an Agraphie (Schreibunfähigkeit) leidet, begreift nicht, was andere ihm sagen.

4. *Minnesota Teacher Attitude Inventory (MTAI)* von W. Cook, C. Leeds und R. Callis (Psychological Corporation), 1951¹¹⁹

Der MTAI ist bestrebt, die Einstellungen zu messen, die es erlauben, das Lehrer-Schüler-Verhältnis sowie indirekt den Gewinn vorauszusagen, den der zukünftige Pädagoge aus dem Unterricht ziehen wird.

Die Konstruktion dieses Tests ist offenbar stark beeinflusst worden von dem Werk von T. W. Adorno u.a., *The Authoritarian Personality* (New York, Harper, 1950). Tatsächlich räumt er den Fragen, die zur Klärung der Absicherungstendenzen geeignet erscheinen, weiten Raum ein.

Der Test umfaßt 150 Items, wobei der Proband jeweils angibt:

sehr einverstanden — einverstanden — unentschieden — nicht einverstanden — ganz und gar nicht einverstanden.

Beispiel:

- Heutzutage läßt man zu viele Kinder eigenmächtig handeln;
- ein Lehrer kann dem, was die Schüler sagen, keinen großen Glauben schenken;
- man soll einen Schüler, der an seinen Nägeln kaut, auslachen;
- ein Schüler hat das Recht, sich offen darüber auszusprechen, wenn er mit dem Lehrer nicht einer Meinung ist.

119 Zu diesem Test siehe: K. M. Evans, An Examination of the MTAI (British Journal of Ed. Psychol., 28, 1958, S. 253–257) und K. M. Evans: The MTAI (Educational Research, VIII, Nr. 2, 1966, S. 134–141).

5. *Cartoon Situation Test* von E. Shapiro, B. Biber und P. Minuchin¹²⁰

Dieser von Rosenzweig inspirierte Projektive Test dient der Erfolgsvoraussage der Lehrer im Unterrichten. Besonders die Trickfilmaufnahmen, bei denen der Proband einen Kommentar geben muß, erlauben die Gewandtheit im Ausdruck, die Art der Reaktion gegenüber den gestellten Aufgaben, die Art der Identifizierung mit den dargestellten Personen, die Wahrnehmung der autoritären Rolle, die psychologische Sensibilität, die Aggressivität und die Gesellschaftsfähigkeit zu messen. Bisher stehen nur vorläufige Resultate zur Verfügung; sie erscheinen jedoch ermutigend.

6. *Draw-A-Teacher Technique*

Dieser andere Projektive Test wurde von dem Amt für Erziehungsforschung der *Municipal Colleges* in New York eingesetzt, um zu untersuchen, wie sich die Lehrer und die Studenten der Pädagogik den Unterricht vorstellen.

Er besteht darin, daß man ein Blatt Papier mit folgender Instruktion vorlegt: „Zeichnen Sie in den unten dafür vorgesehenen Raum einen Lehrer in seiner Klasse zusammen mit seinen Schülern. Entwerfen Sie das Bild möglichst vollständig. Ihre künstlerischen Talente spielen dabei gar keine Rolle; zeichnen Sie es so gut, wie Sie können.“

Travers¹²¹ zeigt zwei treffliche auf diese Art erhaltene Entwürfe. Nach unserer Kenntnis verfügt man bisher weder über eingehende Untersuchungen dieser Technik noch über ein Schema der Inhaltsanalyse.

III. Die Kriterien aufgrund der Unterrichtsverfahren oder der Unterrichtsfunktionen

In diesem Abschnitt unterscheiden wir: A. Die Beurteilung des allgemeinen Klassenklimas; B. Die Beobachtung und das Messen der charakteristischen Verhaltensweisen von Lehrern und Schülern; C. Die Beurteilung der Pädagogen durch die Schüler.

A. *Beurteilung des allgemeinen Klassenklimas*

Es werden vor allem zwei Techniken zur Bestimmung des Klassenklimas

120 Siehe: The Cartoon Situation Test: a semi-structured technique for assessing aspects of personality pertinent to the teaching process (Journal of Projective Techniques, Nr. 20, 1957, S. 172-184).

121 R. Travers, An Introduction to Educational Research, New York, Macmillan, 1958, S. 219.

eingesetzt: die kontrollierte Beobachtung und die Beurteilungsskala. Diese beiden Techniken sind von Wrightstone¹²² angewandt worden.

1. *Kontrollierte Beobachtung*

Gut ausgebildete Beobachter halten sich während einer festgesetzten Dauer in den Klassen auf (*time sampling*). Es werden alle Tätigkeiten des Lehrers sowie der Schüler notiert, oftmals geschieht dies mit Hilfe eines Codesystems. Außerdem werden Stichproben der Verhaltensweise stichwortartig beschrieben. Schließlich analysiert man die Beobachtungen, beispielsweise nach folgenden Gesichtspunkten:

- a) Initiative;
- b) Zusammenarbeit;
- c) Kritikübung;
- d) *Leadership*;
- e) Arbeitsmethode.

2. *Pupil Teacher Rapport Scale*¹²³

Diese *rating scales* erstrecken sich auf folgende charakteristische Merkmale: Art, Grad und Qualität der Lehrer-Schüler-Interaktionen; Interesse, Vergnügen, emotionales Verhalten, Anordnungen oder Vorschläge des Erziehers; physische Spannkraft der Gruppe, emotionales Verhalten der Schüler.

Beispiel: Emotionales Verhalten des Lehrers:

	<i>Beurteilung</i>
1. Aggressiv (in Opposition, spöttisch zu den Schülern)
2. Reizbar (gereizter Ton, wenn er sich an die Schüler wendet)
3. Tolerant (bemüht sich, jegliche Gereiztheit zu vermeiden)
4. Liebenswert, aber reserviert
5. Herzlich und sympathisch

Unter den neueren Untersuchungen sind zu erwähnen:

3. *Observation Schedule and Record (OScAR)* von D. Medley und H. Mitzel¹²⁴

Der OScAR ist ein Beobachtungsplan, der sich in vier Abschnitte aufteilt:

- a) Tätigkeiten des Lehrers und der Schüler,
- b) Soziometrie,
- c) Benutzte Lehr- und Lernmittel,
- d) Zeichen, die dazu dienen, die symptomatischen Gesichtspunkte, des Klassenklimas zu registrieren.

122 J. Wrightstone, J. Justman u. I. Robbins, Evaluation in Modern Education, New York, American Book Co., 1956, S. 424 ff.

123 J. Wrightstone, Measuring the Social Climate of Class-Room (Journal of Educational Research, Januar 1951, S. 341 ff).

124 Siehe: A technique for Measuring Class-Room Behavior (Journal of Educational Psychology, April 1958, S. 86-92).

Die Auswertung führt zu vierzehn möglichen Ergebnisvariablen. Diese Variablen sind jedoch nicht klar genug definiert, und die Faktorenanalyse zeigt an, daß der OScaR nur über drei Dimensionen sichere Informationen gibt, und zwar über das soziale und emotionale Klassenklima, den mehr oder weniger verbalen Aspekt des Lernens und darüber, in welchem Grade die Sozialstruktur um den Pädagogen zentriert ist.

In diesem Zusammenhang weisen wir noch auf die von Withall¹²⁵ durchgeführte Untersuchung hin, der ein Klassifizierungsschema des sprachlichen Ausdrucksvermögens des Lehrers auf einem Kontinuum vorschlägt, das sich von der Feststellung des sprachlichen Ausdrucks beim Schüler bis hin zum Einfluß der sprachlichen Ausdrucksweise des Lehrers erstreckt. Aufgrund dieser Daten wird ein Index des sozio-emotionalen Klassenklimas errechnet.

B. Beobachtung charakteristischer Verhaltensweisen im Unterricht

Auf diesem Gebiet sind in letzter Zeit Fortschritte erzielt worden, die nach unserer Meinung die größte Beachtung verdienen. Wir weisen hin auf ein gut bekanntes Untersuchungsverfahren (Skala von Torgenson) und analysieren eingehend zwei bedeutende Untersuchungen der letzten Jahre, und zwar die von M. Hughes sowie die von D. Ryans.

Unter dieser Rubrik sollte man auch der Technik der kritischen Ereignisse (*critical incidents*) einen Platz einräumen, die wir bereits beschrieben haben¹²⁶.

1. The Torgenson Diagnostic Teacher Rating Scale of Instructional Activities (Public School Publ. Co.)

Diese rating scale, die der Inspektion dient, umfaßt 18 Items zur Erforschung der folgenden Gesichtspunkte: Diskussion mit den Schülern, Kenntnis der Kinder, Übungsmaterial zur Leistungssicherung und Individualisierung des Unterrichts, Technik des Zensierens, kritische Einstellung, Aufmerksamkeit, Motivation, Aktivität der Schüler usw.

Beispiel: Diskussion mit den Schülern.

- A. Wendet sich nur an die besten Schüler.
- B. Die Mehrzahl der Schüler beteiligt sich am Gespräch.
- C. Die Mehrzahl der Schüler zeigt kein Interesse an der Diskussion.
- D. Der Lehrer wirkt hemmend auf die Unterhaltung bzw. die Fragen.
- E. Wenig Gedankenaustausch mit den Schülern.

125 J. C. Withall, Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms (J. of Exp. Educ., Nr. 17, 1949, S. 347 ff).

126 Siehe S. 83 ff.

2. Eine objektive Auswertungsmethode der Qualität des Unterrichts in der Volksschule von Marie Hughes et al.¹²⁷

M. Hughes stellt den Grundsatz auf, daß die einzig wirkungskräftige Form, Kriterien zur Beurteilung des pädagogischen Wertes der Erzieher zu erhalten, die genaue Kenntnis dessen darstellt, was in der Klasse vorgeht. Wie sieht die Aktivität der Lehrer tatsächlich aus? Wie spielt sich das Leben des Kindes in der Schule ab? Bisher stehen noch keine technischen Mittel zur Verfügung, die es erlauben, während einer Unterrichtsstunde die Verhaltensweisen aller Schüler einer Klasse gleichzeitig zu verfolgen^{127a}. Es scheint aber durchaus möglich, daß die Art, wie der Lehrer seinen Unterricht abhält, die Teilnahme der Schüler am Unterricht, ihre mehr oder weniger große Mitbestimmung und sogar den Denkprozeß bei ihrer Arbeit bestimmt. Zudem hängt das Verhalten der Kinder untereinander in großem Maße von der Art ab, wie der Erzieher sie individuell behandelt.

Darum konzentriert M. Hughes ihre ganze Aufmerksamkeit auf den Lehrer:

- a) Welche Unterschiede lassen sich zwischen den Verhaltensweisen (Funktionen) der als gut angesehenen und anderen Erziehern diagnostizieren?
- b) Welches sind die vorherrschenden oder die statistisch am häufigsten auftretenden Unterrichtsmethoden bei den guten Lehrern im Verhältnis zu den weniger guten?
- c) Welche Verhaltensunterschiede treten in verschiedenartigen Situationen zutage (z. B. in den einzelnen Unterrichtsstunden: Lesen, Geschichte, Rechnen, Werkarbeit)?
- d) Welche Verhaltensweisen des Erziehers scheinen die Schüler zu höherer geistiger Mitarbeit, zu einer persönlichen Teilnahme sowie zur schöpferischen Tätigkeit anzuregen?
- e) Wie sieht das Modell eines guten Unterrichts aus?

Zwei Jahre lang beobachtete man die Verhaltensweise einer großen Zahl von Erziehern in den verschiedenen Schulstufen. Schließlich wurde ein Plan entwickelt, der die tatsächlich ausgeübten Verfahrensweisen analysierte¹²⁸.

Nachdem dieser Plan entwickelt worden war, wählte M. Hughes 35 von ihren Vorgesetzten als gut bezeichnete Erzieher aus, und die systematische Beobachtung führte zu folgenden Resultaten:

127 M. Hughes, Means for the Assessment of the Quality of Teaching in Elementary Schools, Salt Lake City, Univ. of Utah, 1959, 400 S.

Die Auswertungsmethode von M. Hughes ist jetzt von G. de Landsheere im Laboratorium für Pädagogische Forschung der Universität Lüttich weiterentwickelt worden. Aus den sechs ursprünglichen Kategorien sind nur zwei praktisch unverändert geblieben. Die Lütticher Forscher arbeiten jetzt mit neun Kategorien.

127a Anm. d. dtsh. Bearbeiters: Durch die Unterrichtsmittschau (Fernsehen) ist seit neuester Zeit eine solche techn. Möglichkeit geschaffen worden, um in Bild u. Ton eine Unterrichtsstunde genau festzuhalten u. beliebig oft vorzuführen.

128 Dieser Plan findet sich auf S. (223-224).

Verteilung der Gesamtzahl der Unterrichtshandlungen: 35 Erzieher – 90 Beobachtungsmi-
nuten

Gesamt		Kontroll- ierende Hand- lungen	An- ord- nungen	Hel- fende Hand- lungen	Ent- wick- lung	Persön- liche Beant- wor- tungen	Handlungen der Affektivität	
							+	-
Gesamt- zahl der Hand- lungen	26385	12402	738	1918	4131	1289	3243	2664
%	100	47	3	7	16	5	12	10
Durch- schnitts- zahl pro Erzieher	753,9	354,3	21,1	54,8	118	36,8	92,7	76,1

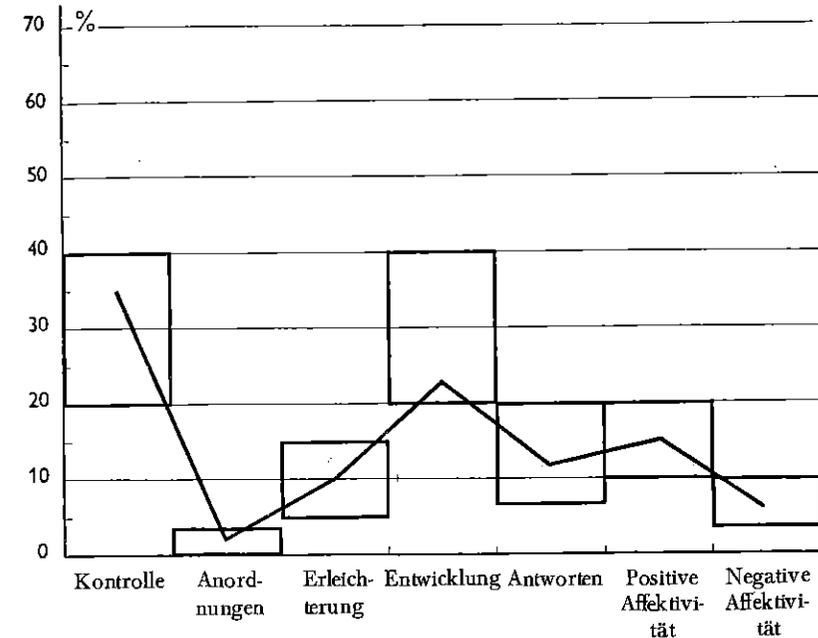
Nach einigen ergänzenden Untersuchungen schlagen M. Hughes u. a. als Modell
eines guten Unterrichts vor:

Kontrollierende Handlungen	20 bis 40 %
Den Unter. bestimmende Aktionen (Anordnungen)	1 bis 3 %
Helfende Handlungen	5 bis 15 %
Entwickelnde Unterrichtsaktionen	20 bis 40 %
Persönliche Beantwortungen	8 bis 20 %
Handl. positiver Affektivität	10 bis 20 %
Handl. negativer Affektivität	3 bis 10 %

Um den Wert eines Erziehers, verglichen mit dem angenommenen Modell zu beur-
teilen, genügt es, das Profil der Resultate in die graphische Darstellung auf S. 189 zu
übertragen:

In sich selbst ist das von M. Hughes vorgeschlagene Modell offensichtlich anfecht-
bar; es beruht – wie gesagt – auf der Beobachtung von Erziehern, die von ihren
Rektoren oder Schulräten als besonders qualifiziert bezeichnet wurden. Aber die von
diesen Vorgesetzten benutzten Kriterien sind uns nicht bekannt. Außerdem steht ein
solches Modell nicht nur unter dem Einfluß unseres aktuellen Kenntnisstandes auf
dem Gebiet der Pädagogik und der Psychologie, sondern auch unter dem unserer
Erziehungsphilosophie.

Dennoch hat die Ausarbeitung eines Modells von der Art des vorliegenden den Vor-
zug, einen präzisen Vorschlag zu bringen, und es ist wegbereitend für ergänzende
Untersuchungen, die aller Wahrscheinlichkeit nach sehr instruktiv ausfallen werden.
Von gut ausgebildeten Beobachtern und Kodierungsspezialisten angewandt, erlaubt
das System von Marie Hughes darüber hinaus eine objektive Beschreibung der
Tätigkeit des Unterrichtenden, und es ist in dieser Hinsicht eines der besten, das bis
heute entwickelt wurde.



Das auf das Modell übertragene Profil stellt den Durchschnittswert der Beobach-
tungen innerhalb von 30 Minuten dar, die dem Modell am nächsten liegen.

3. Teacher Characteristics Schedule and Classroom Observation Record von D. G. Ryans¹²⁹

Wir beschreiben kurz zwei Untersuchungsverfahren, die aus einer koordinierten
Untersuchungsreihe hervorgegangen sind, und die sicherlich die umfangreichste
wissenschaftliche Untersuchung darstellt, die man bisher über die Verhaltensweise
der Lehrer in ihrer Klasse, ihre Wertsysteme und ihre kognitiven und emotionalen
Merkmale angestellt hat.

Es wurden dabei die nachstehend angeführten Ziele verfolgt:

- Entwicklung von Techniken, die eine zuverlässige Beurteilung der Verhaltens-
weise in der Klasse erlauben.
- Bestimmung der Musterbeispiele (*patterns*) für die Verhaltensweise der Lehrer
(vor allem durch Faktorenanalyse).

¹²⁹ D. G. Ryans, Characteristics of Teachers, Washington, American Council on Educa-
tion, 1960, 416 Seiten.

c) Schaffung eines Untersuchungsinstrumentariums, das es mit Hilfe der aufgedeckten Fakten erlaubt, das Verhalten der Lehrer in der Klasse vorauszusagen und ihre persönlichen und sozialen Merkmale zu bestimmen.

d) Vergleich bestimmter Gruppen von Lehrern (Volksschullehrer, Lehrer des höheren Schuldienstes, verheiratete, ledige Lehrer usw.).

Es wurden mehr als 6000 Erzieher aus 1700 verschiedenen Schulen geprüft. Eine der bedeutenden Schlußfolgerungen aus der exakten Datenanalyse ist die, daß das grundsätzliche Verhalten des Lehrers in der Klasse durch 3 Dichotomien (Gegensätze) dargestellt werden kann:

- Liebenswürdig — distanziert, reserviert;
- Systematisch — unorganisiert;
- Phantasievoll — ohne Inspiration.

Der *Classroom Observation Record* umfaßt 22 bipolare Beobachtungen: 18 über das Verhalten der Erzieher und 4 über das der Schüler (apathisch — lebhaft, parteiisch — gerecht usw.).

Jede Beobachtung wird in eine 6 stufige *rating scale* eingetragen.

Das *Teacher Characteristics Schedule* umfaßt 300 Items (Mehrfachwahl oder *Check list*), die sich auf die Einstellungen und die Meinungen der Erzieher erstrecken, die mit den anhand des *Classroom Observation Record* beobachteten Verhaltensweisen zu korrelieren scheinen.

Beispiel: Was halten Sie von folgender Behauptung: „Die Mehrzahl der Schüler arbeitet nach bestem Vermögen.“

Absolut nicht einverstanden — nicht einverstanden — ich weiß nicht — einverstanden — absolut einverstanden.

Ryans hat vielleicht mehr als jeder andere das ungeheure Ausmaß der durchzuführenden Arbeiten zur Messung und Voraussage der Verhaltensweise der Lehrer ins Licht gerückt. Seine Untersuchungen, die er mit einem beträchtlichen Aufwand an technischen Mitteln durchgeführt hat, haben länger als 10 Jahre gedauert.

C. Die Beurteilung der Lehrer durch die Schüler

Die Schüler beurteilen ihre Lehrer bekanntlich mit einer Treffsicherheit, die in den seltensten Fällen versagt. Auch haben etliche Untersucher Untersuchungsverfahren entwickelt, die eine Meinungsanalyse der Schüler über ihre Lehrer erlauben.

Gewiß fehlt es hierbei nicht an Einwendungen von Seiten der Lehrer. Remmers und Gage¹³⁰ haben sich zweifellos am eingehendsten mit der Diskussion des aufgeworfenen Problems befaßt. Die in der Folge aufgeführte Tabelle gibt einen Überblick über ihre Argumente.

130 Op. cit., S. 492—497.

Einwendungen

1. Die Schüler sind inkompetent.
2. Das ist Demagogie. Der beste Lehrer ist nicht unbedingt derjenige, der den Schülern gefällt.
3. Die von den Schülern in Schlagworten geäußerten Meinungen sind wertlos.
4. Die Meinungen können beeinflusst werden durch den Umfang des vom Lehrer gegebenen Arbeitspensums, durch das Interesse, das ein Schüler an einem Fach zeigt oder durch die Schwierigkeiten, denen er begegnet.
5. Man provoziert Eifersucht unter den Kollegen und entmutigt sie.
6. Die Schüler werden ihre Lehrer weniger respektieren.

Widerlegungen

1. Selbst wenn dies zutrifft, spielt ihre Einstellung dennoch eine wichtige Rolle beim Lernprozeß.
2. Die beste Erziehung ist die demokratische.
3. Die statistische Kontrolle hat das Gegenteil erwiesen.
4. Die Korrelationen zeigen, daß dies nicht der Fall ist (es besteht lediglich ein Unterschied zwischen Jungen und Mädchen).
5. Wenn man diese Reaktionen befürchtet, sollte allein der interessierte Lehrer die Resultate kennen. Aber dieses Problem scheint kaum aufzutreten.
6. Dies wäre noch zu beweisen, und die Untersuchungen von Bowman scheinen sogar das Gegenteil anzuzeigen.
7. In jedem Fall beurteilen die Schüler ihre Lehrer. Wichtig wäre, zu wissen, ob die Erzieher Wert darauf legen, die geäußerten Beurteilungen zu erfahren.

Man hat den Eindruck, daß das im amerikanischen Unterrichtswesen herrschende soziale Klima viel eher eine offene Kritik von Seiten der Schüler erlaubt als im Rahmen westeuropäischer Schulverhältnisse. Indessen erscheint uns das System jedoch nach einer angemessenen psychologischen Vorbereitung anwendbar, zumindest im experimentellen Stadium und in gut definierten Abgrenzungen. Es folgen nun einige standardisierte Tests, die es dem Schüler erlauben, ihre Meinungen zu äußern.

1. *The Purdue Rating Scale for Instruction* von H. Remmers und D. Eliot (Purdue University), 1950

Beurteilungsskalen in bezug auf folgende Aspekte: Interesse des Erziehers an dem, was er unterrichtet, Sympathie für die Schüler, Integrität, die Art, neuen Unterrichtsstoff darzubieten, Taktgefühl und Sinn für Humor, physische Gesichtspunkte, Fähigkeit, Lerneifer bei den Schülern zu wecken.

Beispiele:

- Integrität — Absolut gerecht und unparteiisch gegen jedermann
- Beweist häufig, daß er besondere Günstlinge hat
- Vielfach parteiisch
- Toleranz — Erkennt die verschiedenen Standpunkte an
- Hat gewisse Vorurteile, bleibt jedoch gewöhnlich tolerant
- Intolerant, duldet keinen Widerspruch

2. The Diagnostic Teacher Rating Scale von Tschechtelin, Amatora und Remmers (Purdue University), 1940

Diese nach der Technik von Thurstone konstruierten Skalen sind für Schüler der letzten Volksschulklassen oder für die Unterstufe der höheren Schule bestimmt. Es wurden folgende charakteristische Merkmale näher untersucht: Achtung vor dem Lehrer, seine Fähigkeit zu erklären, seine Liebenswürdigkeit, seine Integrität, seine Disziplin, das Arbeitspensum, das er in der Klasse verlangt, und die Hausaufgaben, die er stellt.

Beispiel: Achtung vor der Lehrerin

1. Ich habe sie lieber als andere.
2. Sie hat Sinn für Humor.
3. Sie hält die Klasse gut in Zucht.
4. Sie ist hübsch.
5. Sie lächelt niemals.
6. Sie ist nicht höflich.
7. Sie ist zu verdrießlich.

3. The Bryan-Yntema Rating scale ¹³¹

Diese für den höheren Schulunterricht konstruierte Skala umfaßt zehn Items mit geschlossener und drei Fragen mit offener Antwort.

- a) Auf folgende Fragen antworten, und zwar mit: ausgezeichnet (1), gut (2), genügend (3), ausreichend (4), ungenügend (5).

Für jedes Item wird jede einzelne Beurteilung kurz erläutert (Beispiel: Item 1: genügend = im allgemeinen nett und liebenswürdig, berücksichtigt jedoch nicht immer den Standpunkt des Schülers).

1. Finden Sie diesen Lehrer sympathisch ?
 2. Sorgt er für gute Disziplin ?
 3. Sind die Noten, die er verteilt, gerecht ?
 4. Erklärt er neuen Unterrichtsstoff klar und deutlich ?
- usw.

131 In: A Manual on the Evaluation of Students Reactions in Secondary Schools, Kalamazov (Mich.), Western State Teacher College, 1939.

- b) Die drei Fragen mit offener Antwort erlauben es, die einzelnen Standpunkte zu präzisieren.

1. Auf welche Frage zwischen 1 und 8 haben Sie die ungünstigste Antwort gegeben ? Erklären Sie kurz, warum.
2. Geben Sie bitte ein oder zwei Dinge an, die Sie an diesem Lehrer besonders schätzen.
3. Hat dieser Lehrer, abgesehen von dem oben Gesagten, Angewohnheiten, die Sie nicht mögen ? Welche ?

IV. Schlußfolgerung

Wir haben bereits zu Beginn darauf hingewiesen, daß die uns für die objektive Voraussage und Beurteilung der Leistungsfähigkeit der Erzieher zur Verfügung stehenden Mittel bisher noch sehr begrenzt sind.

Wir stimmen mit Ryans überein, der zu folgendem Schluß kommt¹³²:

1. Die exakte Beurteilung der Leistungsfähigkeit ist illusorisch und wird es noch lange bleiben, denn sie erfordert das gleichzeitige Messen von eilichen Dutzend Variablen.
2. Die Beurteilung der Leistungsfähigkeit eines Lehrers ist nicht trennbar von einem Werturteil, von einer Erziehungsphilosophie. Daher wird selbst der Begriff der Leistungsfähigkeit immer relativ sein — eine Erklärung dafür, warum alle Untersuchungen, die eine Beurteilung der Leistungsfähigkeit *ad absolutum* anstreben, zu Mißerfolgen geführt haben—.
3. Angesichts dieser Schwierigkeit bevorzugen es die Forscher immer mehr, sich bei ihren Untersuchungen auf einem zuverlässigen und mehr fundierten Boden zu bewegen:
 - a) Indem sie sich auf eine objektive Beschreibung der Verhaltensweise der Lehrer konzentrieren, eine Tendenz, die wir anhand der Arbeiten von M. Hughes erläutert haben;
 - b) Indem sie die Relationen zwischen den internen und externen sowie den menschlichen und materiellen Bedingungen einerseits, und den Verhaltensweisen der Erzieher andererseits untersuchen.

Es wäre jedoch falsch, sich auf diese Begrenzungen zu berufen, um auf die Verwendung aller objektiven Meßinstrumente für die Auswahl bzw. die Beurteilung von Lehrern zu verzichten. In der Praxis kommt es darauf an, parallel zu den Revisionen und üblichen Prüfungen eine möglichst hohe

132 D.G. Ryans, Assessment of Teacher Behavior and Instruction (Review of Ed. Res., XXXIII, Oktober 1963, S. 415-441).

Zahl an Messungen zu erreichen, derart, daß man eine einheitliche Datenzusammenstellung gibt und Resultate freilegt, die man nach und nach validieren sollte.

8. Die Bewertung der Schulen

Wir haben soeben gesehen, wie schwierig es ist, den Wert der Lehrer zu messen. Die Problematik, eine Schule in ihrer Gesamtheit zu beurteilen, stellt ein noch gewagteres Unterfangen dar. Man ist noch weit davon entfernt, einigermaßen präzise die relative Bedeutung der materiellen Faktoren: örtliche Voraussetzungen, Schulausrüstung, -organisation, Lehrpläne sowie der menschlichen Faktoren: Schüler, Lehrer, Direktoren und Verwaltungspersonal zu bestimmen. Zum gegenwärtigen Zeitpunkt müssen wir daher viel mehr die getrennte Beurteilung dieser Faktoren ins Auge fassen.

In einem stark zentralisierten Land wie Belgien oder Frankreich, wo der Lehrplan bei gleicher Schulart und für die gleichen Klassen derselbe ist, wird sich die Erhebung normalerweise nicht auf den Lehrplan als solchen erstrecken, sondern lediglich auf seine Anwendung.

Im Gegensatz dazu stellen in anderen Ländern, in denen die einzelnen Gemeinden viel mehr Initiative zeigen, die speziellen Lehrprogramme einen der Hauptbeurteilungspunkte dar. Kurz gesagt, besteht die kritische Prüfung der Lehrprogramme darin, daß man den Inhalt und die damit verbundenen Methoden im Hinblick auf die anerkannten Erziehungsgrundsätze und unter Berücksichtigung der typischen Merkmale des derzeitigen kulturellen Entwicklungsstandes erörtert.

Abgesehen von Aspekten der Bauweise, die direkt durch die Hygiene, die Sicherheit, durch den generellen Verwendungszweck oder die Ästhetik geboten sind, wird man kein absolutes Urteil über die Schulgebäude abgeben können. Denn auch sie sind Ausdruck einer pädagogischen Grundauffassung.

Tatsächlich wird man sich — je nachdem, ob man einen aktiven oder inaktiven Unterricht, ein Arbeiten mit flexiblen Gruppen oder dem starren Klassenverband, eine Bibliothek, die nur als Lagerungs- und Ausgaberaum für Bücher oder im Gegenteil als Ort für individuelle Nachforschungen dient wünscht, — den verschiedenartigsten architektonischen und funktionellen Anforderungen gegenübergestellt sehen. N. Hans hat mit Recht darauf hingewiesen, daß entsprechend dem Baustil der jeweiligen Epoche die Schul-

gebäude Ähnlichkeit mit Gefängnissen, Klöstern, Kasernen oder Krankenhäusern haben¹³³...

Die Organisation einer Schule hingegen ist relativ leichter zu beurteilen. Dennoch untersucht man sie relativ selten mit den Mitteln, die die moderne Technik bietet. Zweifellos aus diesem Grunde bleiben gerade im administrativen Bereich die Arbeitsmethoden vielfach derart primitiv, daß kein einziges Handels- oder Industrieunternehmen sie ertragen könnte.

Es bliebe auch noch viel zu sagen über die Beurteilung der Schulleiter, die man entsprechend den angewandten Methoden zur Beurteilung der Geschäftsführer vornehmen könnte. Aber hierbei berührt man eine empfindliche Stelle. Einmal, weil die Schulleiter im allgemeinen dazu berufen sind, zwei Funktionen zu erfüllen: ihnen obliegt die administrative sowie die pädagogische Leitung; diese Funktionen verlangen unterschiedliche Fähigkeiten und eine unterschiedliche Vorbildung, die sich gegenseitig nicht unbedingt ausschließen, aber vielleicht nur in Ausnahmefällen vereint anzutreffen sind. Wie dem auch sei, eine erfolgreiche pädagogische Leitung setzt vornehmlich eine umfangreiche Tätigkeit in der Handlungsforschung voraus, so daß wenig Zeit für andere Aufgaben bleibt.

Schließlich enthüllt die Beurteilung der amtierenden Schulleiter nach wissenschaftlichen Gesichtspunkten, während ihre Ernennung im allgemeinen nach nicht-wissenschaftlichen Kriterien erfolgt, oftmals peinliche Mängel und Lücken^{133a}.

Die wenigen Aspekte, die wir uns gerade vergegenwärtigt haben, vermitteln dennoch eine Vorstellung von der Breite des Untersuchungsprogramms, die für eine objektive Beurteilung einer Schule notwendig ist. Der begrenzte Rahmen dieses Werkes erlaubt es uns nicht, auf diese Frage näher einzugehen, so wichtig sie auch sein mag.

In verschiedenen Formen angewandt (Fragebogen, Inventarien, rating scales), muß man bei der Verwendung der Untersuchungsverfahren zur Beurteilung der Schulen etwa folgende Regeln des Vorgehens beachten:

1. Das angestrebte Modell wird so exakt wie möglich definiert.
2. Man macht sodann eine Aufstellung derjenigen Anforderungen, denen der jeweils zu unterscheidende Aspekt (z. B. Schulgebäude) entsprechen sollte, um sich an dem gewählten Modell zu orientieren.

133 N. Hans, *The historical approach to comparative education* (Thoughts on Comparative Education, op. cit., S. 50—51).

133a Vgl. hierzu: N. Gross und R. Herriott, *Staff Leadership in Public Schools, a Sociological Inquiry*, New York, J. Wiley and Sons, 1965, 247 S.

3. Man redigiert die Items, die eine Handhabe für die Beurteilung darstellen. Von Fall zu Fall beschränkt man sich darauf, das Vorhandensein oder Fehlen des infrage stehenden Merkmals festzustellen, oder man ordnet Punkte zu, die eine Gewichtung erlauben.

Nach unseren bisherigen Erfahrungen erscheint es nicht erstrebenswert, einen Gesamtpunktwert zu erstellen, der alle zu untersuchenden Aspekte erfaßt.

Als Beispiel folgen zwei amerikanische Untersuchungsmethoden, die uns für die gegenwärtige Tendenz ziemlich repräsentativ zu sein scheinen.

1. *Scale for Elementary Schools* von J. Morrison und V. Ruegsegger, Albany, University of the State of New York, 1943

Diese Skalenreihen sind für die Beurteilung von Volksschulen bestimmt und erfassen 58 Merkmale, und zwar der Lehrmethoden, des Lehrstoffs, der Klassenatmosphäre und des Milieus.

Man spürt deutlich, daß die Urheber dieses Untersuchungsverfahrens sich an der Pädagogik von Dewey orientiert haben.

Der Beobachter verfügt jeweils über drei schriftlich formulierte Merksätze und kreuzt auf einem Kontinuum jeweils denjenigen an, der nach seiner Meinung die beobachtete Verfahrensweise wiedergibt. Ein Syntheseprofil läßt unmittelbar erkennen, wo der Unterricht fortschrittlich oder formell, bzw. eine Zwischenstufe davon ist.

Beispiel: Item = 5: Mittel und Wege, Informationen zu erhalten¹³⁴.

1	2	3
Die Kinder akzeptieren schlechthin das, was ihnen durch einen Text oder den Lehrer gesagt wird, ohne es zu diskutieren, Fragen zu stellen oder eine andere kritische Stellung dazu einzunehmen.	Die Mehrzahl der Kinder prüft die Informationen nur dann, wenn man sie dazu auffordert. Man stellt eine gewisse Tendenz zum Diskutieren der Informationen fest. Einige Kinder finden manchmal die Antworten selbständig, stellen Fragen und sammeln Auskünfte an Ort und Stelle.	Die Kinder verwenden die elementaren Untersuchungstechniken, um Informationen zu finden, und begnügen sich nicht damit, passiv alles, was man ihnen sagt, zu akzeptieren. Sie suchen Rat in Nachschlagewerken und Zeitschriften. Die Informationen werden auf ihre Genauigkeit hin geprüft.

134 Zitiert nach Wrightstone et al., op. cit. S. 425-426.

2. *Guide for Evaluating and Improving Nebraska Elementary Schools* von E. Greer, et al., Lincoln, 1955, 90 Seiten

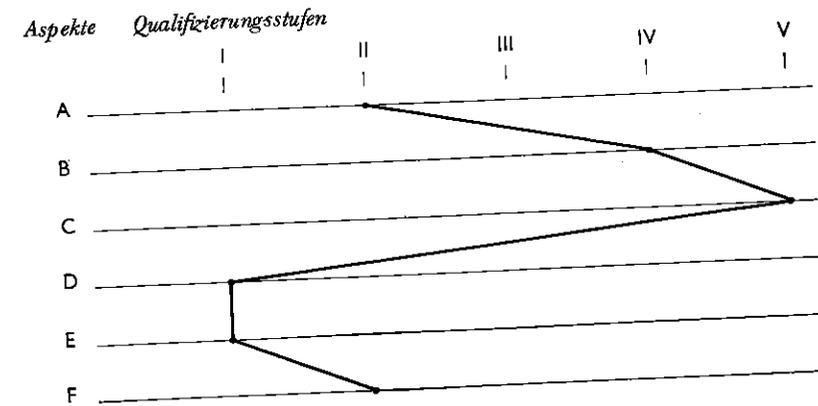
Diese ausführliche Anleitung ist für die Beurteilung der Nebraska-Schulen und ihres Lehrpersonals entworfen und ist tatsächlich nur für diese zu gebrauchen. Dennoch bringen wir eine Darstellung dieses Untersuchungsverfahrens, und zwar in ziemlich eingehender Form, weil es eine nützliche Anregung vermitteln kann.

Es werden folgende Aspekte untersucht:

1. Lehrstoff und -methoden;
2. Pädagogisch-psychologische Führung und Beratung (Guidance);
3. Bibliotheken;
4. Verwaltung und pädagogische Leitung;
5. Lehrer;
6. Kontakt zwischen Schule, Elternhaus und Gemeinde;
7. Schulgebäude und -einrichtung.

Jeder einzelne Aspekt wird in verschiedenen Abschnitten untersucht und nach fünf Qualifizierungsstufen beschrieben, ausgehend vom mäßigen bis zum besten. Ein Profil synthetisiert die Situation.

Beispiel: Organisation



Die folgende Tabelle ist eine Zusammenstellung einiger typischer Items.

Lehrprogramm (Allgemeines) (8 Aspekte)

<i>Stufe I</i>	<i>Stufe II</i>	<i>Stufe III</i>	<i>Stufe IV</i>	<i>Stufe V</i>
A. Man unterrichtet bei allen Schülern den gleichen Stoff unter Anwendung der gleichen Techniken. Die Unterrichtsstoffe werden als isolierte Elemente dargelegt. Das Lehrprogramm ist nicht speziell auf die Gruppe, an die man sich wendet, abgestimmt.	Es zeigt sich eine gewisse Tendenz, die Unterrichtsstoffe und -techniken im Hinblick auf die zukünftigen Anforderungen, die an das Kind gestellt werden, auszuwählen.	Man unterweist alle in den gleichen Unterrichtsgegenständen und nach den gleichen Unterrichtstechniken, aber das Lehrprogramm ist teilweise darauf ausgerichtet, den Kindern zu helfen, ihre aktuellen Probleme zu lösen.	Das Lehrprogramm erlaubt allen, Kenntnisse, Lerntechniken und -gewohnheiten und selbstständiges Denken zu entwickeln; die gewählten Erfahrungsaspekte werden für das gegenwärtige wie für das zukünftige Leben des Kindes in einer Demokratie zu entwickeln. Diese Erfahrungspunkte sind im Hinblick auf die Anforderungen, die Fähigkeiten und Interessen der Gruppensmitglieder ausgewählt.	Das Lehrprogramm enthält Erfahrungsaspekte, die dazu beitragen werden, Lerngewohnheiten, Erkenntnisse, nützliche Einstellungen und Techniken für das gegenwärtige wie für das zukünftige Leben des Kindes in einer Demokratie zu entwickeln. Diese Erfahrungspunkte sind im Hinblick auf die Anforderungen, die Fähigkeiten und Interessen der Gruppensmitglieder ausgewählt.

Geschichte, Geographie, Staatsbürgerkunde (Politische Bildung) (15 Aspekte)

I. Disziplin

<i>Stufe I</i>	<i>Stufe II</i>	<i>Stufe III</i>	<i>Stufe IV</i>	<i>Stufe V</i>
Die Schüler werden aufgrund von schriftlichen oder mündlichen Prüfungen beurteilt.	Wie Stufe I, mehr schriftliche und mündliche Übungen; Standardisierter Leistungstest am Ende des Schuljahres.	Wie Stufe II, ferner: Leistungs- und Intelligenztest zu Beginn des Schuljahres. Die Testaufgaben dienen der Anpassung des Unterrichtsprogramms an jeden Schüler.	Wie Stufe III, ferner: Interessentest und regelmäßige Beobachtung durch den Lehrer. Die Schüler sind an ihrem eigenen Fortschritt interessiert.	Wie Stufe IV, ferner: Beobachtungstabellen der Entwicklung jedes einzelnen Schülers. Die Kinder bemühen sich, die erzielten Fortschritte in theoretischen und praktischen Bereich, sowie die Fortschritte im Bereich des menschlichen Verhaltens selbst wahrzunehmen.
Die Schüler werden autoritär vorgeschriebene Regeln, Spannungen, Atmosphäre der Angst in der Klasse.	Der Lehrer schreibt in strenger Form Disziplinarregeln vor, seine Haltung jedoch ist freundschaftlich. Weniger gespannte Atmosphäre als bei Stufe I.	Der Lehrer ruft die Schüler gelegentlich zur Zusammenarbeit auf, um Arbeitspläne aufzustellen. Die Schüler lieben ihren Lehrer und folgen willig seinen Weisungen.	Lehrer und Schüler arbeiten häufig gut zusammen, um Arbeitspläne aufzustellen. Die Schüler nehmen an der Ausarbeitung von Disziplinarregeln teil. Der Lehrer wirkt oftmals bei der Gruppenarbeit mit.	Wie bei Stufe IV, ferner: Atmosphäre des Vertrauens, der Freundschaft und des Verantwortungsgefühls. Die disziplinarischen Probleme werden unter Leitung des Lehrers gelöst.

Naturwissenschaften (7 Aspekte)

D.

Stufe I	Stufe II	Stufe III	Stufe IV	Stufe V
Wenig oder keine Experimente bzw. Untersuchungen vorhanden. Man untersucht vor allem die in dem Lehrbuch erwähnten Stoffe. Vom Lehrer vorgetragene Stoffe.	Gelegentliche Beobachtung, zusätzlich zur Behandlung des Lehrbuchstoffs. Der Lehrer führt einige Versuche vor der Klasse vor. Die Schüler diskutieren über selbst durchgeführte Versuche.	Verschiedene Aktivitäten außerhalb der Behandlung des Lehrbuchstoffs. Die Schüler führen selbständige Versuche durch, Einrichtung von Sammlungen; Untersuchungen in der Natur selbst; mäßige Benutzung von Bildwerfern.	Wie III, ferner: Lehrer und Schüler arbeiten zusammen. Die Aktivitäten erlauben es den Schülern, bestimmte Tabestände aufzudecken und sich Beobachtungsmethoden für den unmittelbaren und späteren Gebrauch anzueignen.	Wie IV, ferner: Gruppendiskussion, um zu einem wissenschaftlichen Ergebnis zu kommen. Dieses Resultat wird mittels diverser Informationsquellen verifiziert und bestätigt: Bücher, Zeitschriften, Befragung von Fachleuten. Die Aktivitäten tragen den individuellen Unterschieden Rechnung. Aneignung besserer Lebensgewohnheiten und Entwicklung wissenschaftlicher Denkfähigkeit.

Muttersprache (23 Aspekte)

J.

Stufe I	Stufe II	Stufe III	Stufe IV	Stufe V
Beim Lesen machen alle Schüler die gleichen Fortschritte. Keine Gruppenarbeit.	Wie I, aber manchmal erhält ein sehr langsame Schüler ein bißchen weniger Arbeit als die anderen.	Aufstellung von 2 oder 3 Gruppen nach dem Kriterium ihrer Lesefähigkeit. Wenig Beweglichkeit in der Gruppierung.	Gruppierung nach den Fertigkeiten. Beweglichkeit.	Gruppierung nach eingehender Untersuchung der Fertigkeiten und der angetroffenen Schwierigkeiten. Große Beweglichkeit.
Um einen neuen Rechenvorgang zu lehren, liest der Lehrer das vor, was im Lehrbuch darüber steht und läßt dann alle Schüler die gleichen Anwendungsübungen machen, die im Buch vorgeschlagen werden.	Der Lehrer liest aus dem Lehrbuch vor, aber er erläutert dann anhand eigener Beispiele. Alle Kinder machen die gleichen Anwendungsübungen.	Wie II, aber darüber hinaus provoziert Diskussion aufgrund von Fragen, die der Lehrer stellt.	Schüler und Lehrer lesen gemeinsam die in dem Lehrbuch vorgeschlagenen Erklärungen und diskutieren sie. Es werden gemeinsam einige Übungen gemacht. Alle üben die gleichen Anwendungen.	Sooft wie möglich entdecken die Kinder unter der Leitung des Lehrers den Rechenvorgang selbst. Die gegebenen Anwendungen werden individualisiert.
<i>Rechnen (7 Aspekte)</i>				
E.				

Musik (11 Aspekte)

B.

Stufe I	Stufe II	Stufe III	Stufe IV	Stufe V
Gelegentlich werden einige Lieder gesungen.	Man schenkt der Musik gewisse Beachtung, jedoch ist im Stundenplan kein spezieller Musikunterricht eingeplant.	Der Musikunterricht ist im Lehrprogramm eingeschlossen. Es beteiligen sich alle, nicht nur die Begabtesten, an den Übungen. Der Lehrplan sieht Vokal- und Instrumentalmusikunterricht sowie rhythmische Übungen vor. Wenig oder gar nicht werden schöpferische Erfahrungen berücksichtigt.	Wie III, ferner: Individualisierung der Aktivitäten. Einige Erfahrungen erlauben es, das musikalische Verständnis sowie schöpferische Befähigung zu entwickeln.	Wie IV. Der Entwicklung des musikalischen Verständnisses und der schöpferischen Tätigkeit wird größere Bedeutung beigemessen.

Hygiene und Unfallverhütung (7 Aspekte) — Kunstverziehung (7 Aspekte) — Leibeserziehung (11 Aspekte) — Pädagogische Führung und Beratung (Guidance) (21 Aspekte)

B.

Kein Experte (Psychologe, pädagogischer Berater, Gehörgeschädigtenlehrer) hilft den Lehrern. N. Schülerzahl pro Klasse: mehr als 40.	Schule, die gelegentlich von einem Fachspezialisten im Dienst des Staates aufgesucht wird. Schülerzahl pro Klasse: 40.	Dem Schulbezirk steht ein psycho-pädagogischer Dienst für gelegentliche Beratungen zur Verfügung. Schülerzahl pro Klasse: 35.	Gute Experten stehen den Lehrern ohne weiteres zur Verfügung. Schülerzahl pro Klasse: 30.	Die Lehrer werden von allen notwendigen Experten unterstützt. Schülerzahl pro Klasse: 25.
--	--	---	---	---

Bibliothek (7 Aspekte) — Verwaltung und Inspektion (8 Aspekte) — Rechtsstand der Lehrer (17 Aspekte)

Relationen zwischen Schule und Elternhaus und Schule und Gemeinde (9 Aspekte)

Stufe I	Stufe II	Stufe III	Stufe IV	Stufe V
Die Lehrer empfinden die Besuche der Eltern als lästig.	Die Eltern werden zu bestimmten Gelegenheiten in die Schule eingeladen.	Die Eltern werden in kleinen Gruppen in die Schule eingeladen.	Die Eltern werden eingeladen, die Schule, sofern sie es wünschen, zu besuchen, um die Kinder beim Lernen zu beobachten.	Wie IV, ferner: Die Eltern sind immer herzlich willkommen.
Die Lehrer versuchen nicht, die ihnen durch das Gemeinwesen gebotenen Hilfsquellen zu nutzen.	Es werden bestimmte Gesichtspunkte des Gemeinwesens bei Exkursionen untersucht. Aberheit gesammelten Erfahrungselemente werden wenig genutzt.	Bemühungen um eine systematische Erforschung des Gemeinwesens.	Intensive Ausnutzung der durch das Gemeinwesen gebotenen Hilfsquellen. Die Kinder sind sich ihrer Verantwortung gegenüber der Gemeinschaft bewußt.	Die Untersuchung des Gemeinwesens steht im Mittelpunkt des Unterrichtsprogramms. Die Schule führt Karree über interessante Erfahrungen. Die Lehrer wirken mit, um die Untersuchungen über das Gemeinwesen zu koordinieren.

Schulebände und -einrichtung (39 Aspekte)

9. Die Bewertung des sozio-ökonomischen Status

I. Bedeutung

Analog zu ihrer fortschreitenden Entwicklung rückt die Soziologie mehr und mehr den Einfluß des sozio-ökonomischen Status von Elternhaus und Lehrern auf die Erziehung des Kindes ins rechte Licht.

Eine eingehende Diskussion über dieses Problem ist im Rahmen dieses Werkes nicht möglich. Wir beschränken uns darauf, einige Elemente aufzuzeigen, die die Bedeutung der sozio-ökonomischen und kulturellen Faktoren in der pädagogischen Forschung hervorheben.

In einem Aufsatz, der großen Widerhall gefunden hat, haben W. Allison und R. Havighurst¹³⁵ gezeigt, daß die Mehrzahl der großen allgemeinen Leistungstests zur Bevorzugung bestimmter sozio-kultureller Populations-schichten tendieren: „Bei allen in den USA gebräuchlichen Testverfahren befassen sich zahlreiche Items mit dem kulturellen Niveau höherer sozio-ökonomischer Gruppen und nicht etwa mit denjenigen der ungefähr 60% Amerikaner, die in den niederen Gruppen heranwachsen“.

Die folgende Aufstellung gibt die Ergebnisse der Analysen wieder, die sich auf mehrere tausend Probanden erstrecken.

Tests	Prozentsatz der Items, bei denen sich eine signifikante Differenz entsprechend dem sozio-ökonomischen Niveau zeigt:	
	Höheres Niveau	niederes Niveau
<i>I. Kinder von 9 bis 10 Jahren</i>		
Henmon-Nelson	93 %	
Otis Alpha (nicht-verbal)	46 %	
Otis Alpha (verbal)	70 %	
Kuhlmann-Anderson (3. Schuljahr)	56 %	
Kuhlmann-Anderson (4. Schuljahr)	85 %	
<i>II. Kinder von 13 bis 14 Jahren</i>		
Terman-McNemar	100 %	
Otis Beta	91 %	
California Mental Maturity	69 %	
Thurstone: Raumvorstellung	84 %	
Thurstone: Logisches Denken	100 %	

135 In The Scientific Monthly, Nr. 66, 1948, S. 301-316.

Alle diese Tests sind ins Französische^{135a} übertragen, bzw. den französischen Verhältnissen angepaßt. Andere bei uns gebräuchliche Testverfahren unterscheiden sich nicht wesentlich von ihnen. Es ist wahrscheinlich, daß die Schlußfolgerungen, die Allison und Havighurst ziehen, uns gleichermaßen angehen, und zwar in einem Maße, dessen nähere Bestimmung selbstverständlich von Wichtigkeit wäre.

Im gleichen Zusammenhang stellen Davis und Haggard fest, wie sehr eine gewählte Ausdrucksweise die Kinder eines niedrigen sozio-ökonomischen Niveaus benachteiligen könnte. Das Experiment besteht darin, daß bei zwei Parallelgruppen der gleiche Test in zwei verschiedenen Formen abgefaßt durchgeführt wird.

Beispiel: Item der Form I

Jemand, der einem anderen, ohne es zu wollen, einen Stoß versetzt, soll:

- sagen, daß er es nicht gewesen ist;
- sich entschuldigen;
- so tun, als sei nichts gewesen;
- sich davon machen.

Form II

Ein Kind, das aus Versehen ein anderes anstößt, soll:

- die Tat leugnen;
- Abbitte tun;
- die Sache ignorieren;
- die Flucht ergreifen.

Während bei der ersten Itemsform die Erfolgsdifferenz zwischen höheren und niederen sozio-ökonomischen Gruppen bei 12% lag, stieg sie bei Form II auf 32% an. Aber die Autoren stellen fest, daß das Grundproblem das gleiche bleibt. Wenn es das Ziel des Tests ist, die Einstellung der Kinder gegenüber diesem Problem kennenzulernen, ist es daher wichtig, daß der Einflußfaktor des Wortschatzes nach Möglichkeit vollständig ausgeschaltet wird¹³⁶.

A. Descoedres hatte bereits systematisch Differenzen in dieser Größenordnung aufgezeigt.

135a Anm. d. dtsh. Bearbeiters: Meistens auch ins Deutsche.

136 A. Davis, Education for the Conservation of Human Resources (Progressive Education, Nr. 27, Mai 1950, S. 221-224).

Siehe auch: T. Husén, Begåvning och miljö, Stockholm, Gebers, 1951.

Beispiel: Altersstufen, bei denen die gegensätzliche Bedeutung der folgenden Adjektive bekannt ist:

	Höheres Niveau	Niedriges Niveau
hart	8 Jahre	10 Jahre
traurig	8	13
ruhig	8	14
dicht	9	9
steif	9	14
mutig	10	12

Es ließen sich wahrscheinlich in allen Bereichen schulischer Aktivität gleichermaßen wichtige, wenn nicht sogar noch ausgeprägtere Variationen, je nach den behandelten Probanden und nach dem Grad der Heterogenität der Schülerpopulationen finden.

Selbst bei gleichem Intelligenzniveau nehmen die Kinder aus Familien unterschiedlichen sozio-ökonomischen Niveaus den Unterrichtsstoff nicht mit gleicher Erfolgchance auf¹³⁷.

Es sollten zudem die Unterschiede berücksichtigt werden, die zwischen den von den Kindern mit in die Schule gebrachten Wertsystemen und dem erstrebten Anspruchsniveau bestehen. Die Bedeutung dieser Faktoren wird von H. H. Hyman¹³⁸ klar herausgestellt.

Im allgemeinen scheint es, daß die Schüler, die einem niederen sozio-ökonomischen Niveau angehören:

1. sich der Behinderung bewußt sind, die ihre Herkunft und ihre pekuniäre Situation darstellt;
2. dazu neigen, den Status ihrer Herkunft als unüberwindbare Determinante zu betrachten;
3. und folglich nicht alles daransetzen, um die ihnen zur Erlangung eines höheren Status gebotenen Mittel voll auszuschöpfen.

Hyman spricht in diesem Zusammenhang von „Barrieren, die sich der einzelne selbst aufbaut“. Dieses Phänomen wird im erzieherischen Bereich deutlich sichtbar. Tatsächlich macht man die Feststellung, daß es nicht nur darum geht, daß offenbar aus finanziellen Gründen das Hochschulstudium für die ärmeren Schüler schwerer zugänglich ist als für Schüler aus wohl-

137 Siehe in diesem Zusammenhang: W. Sewell, A. Haller und M. Straus, *Social Status and Educational and Occupational Aspiration* (*American Sociological Rev.*, Bd. 22, Februar 1957).

138 H. H. Hyman, *The Values Systems of Different Classes* (*Class, Status and Power*, The Free Press, 1953).

habenden Kreisen, sondern daß darüber hinaus der Zugang zum Hochschulstudium von den ersteren weniger stark angestrebt wird als von den letzteren.

Die folgende Tabelle zeigt, daß sich diese Differenz der Erstrebung des Hochschulstudiums in signifikanter Weise dokumentiert, und zwar unabhängig von der verwendeten Schichtung.

Herkunftsschicht	Prozentsatz der Familien, die für ihre Kinder eine Hochschulausbildung wünschen	
	%	N
<i>Ökonomisches Niveau</i>		
Vermögende Bevölkerungsschicht	68	512
Mittelstand	52	1 531
Besitzlose Bevölkerungsschicht	39	856
<i>Beschäftigung</i>		
Freie Berufe	74	301
Geschäftsleute	62	421
Büroangestellte	65	457
Facharbeiter	53	392
Angelernte Arbeiter	49	416
Hausangestellte u. Dienstpersonal	42	194
Landwirte	47	417
Hilfsarbeiter	35	132
<i>Schulbildung</i>		
Hochschulbildung	72	564
Höhere Schulbildung	55	1 411
Volksschulbildung	36	926
<i>Monatsmiete</i>		
Mehr als ca. 60 \$	70	327
Zwischen ca. 40 \$ und 60 \$	64	666
Zwischen ca. 20 \$ und 40 \$	54	990
Weniger als ca. 20 \$	37	403

Schließlich soll man auch der sozio-ökonomischen Herkunft der Lehrer Beachtung schenken. So weiß man beispielsweise, daß heutzutage die Mehrzahl der Lehrer aus einfachen Populationsschichten hervorgeht. Ihr Wertsystem, das sie in die Schule mitbringen, kann in bestimmten Bereichen beträchtlich

abweichen von den Werten, die die aus höheren oder niederen Milieus stammenden Kinder anerkennen. Die Auswirkungen einer solchen Situation scheinen erheblich zu sein; sie sind jedoch bisher noch wenig erforscht.

II. Bewertung

Die Bewertung des sozio-ökonomischen Niveaus ist nicht ganz einfach. Etliche Klassifizierungssysteme sind so simpel konstruiert, daß sie ihre ganze praktische Anwendbarkeit einbüßen. Andere wiederum sind so schwierig zu handhaben (Einzelerhebung über das Einkommen, die Wohnverhältnisse ...), daß sie für die laufende Forschung kaum von Interesse sind, um so weniger, wenn sie sich auf eine höhere Probandenzahl erstreckt.

Unter den einfachen Klassifizierungen findet man zunächst die Aufteilung in die drei herkömmlichen sozialen Schichten: niedere, mittlere und höhere. Diese Gruppierung ist bei weitem nicht zufriedenstellend; es finden sich zum Beispiel in der mittleren Gruppe Kleinhändler, deren Einkommensniveau unter dem eines Facharbeiters liegen kann, und reiche Bürgerliche, deren Existenzniveau sich nur wenig von dem der höheren Populationsschicht unterscheidet.

Bei der Klassifizierung von Fourastié in primäre (Landwirtschaft), sekundäre (Industrie) und tertiäre Sektoren (Dienstleistungsberufe, Freiberufliche, leitende Angestellte) ist die Problemstellung ähnlich: ein Friseurlehrling rangiert in der dritten Gruppe an gleicher Stelle wie der Rechtsanwalt. Tatsächlich hängt der sozio-ökonomische Status von einer Reihe von Faktoren und ihren Wechselbeziehungen ab: dem Beruf und dem beruflichen Erfolgsniveau, dem Einkommen (Verdienst und andere Einnahmequellen: beruflich, außerberuflich; erworbener oder ererbter materieller Besitz), den familiären Banden (selbst bei einer einfachen Tätigkeit genießt ein Junge „aus gutem Hause“ im allgemeinen ein besonderes Ansehen und kann somit leichter günstige Beziehungen anknüpfen), dem kulturellen Niveau, der Zugehörigkeit zu einer Rassenminderheit, der ländlichen oder städtischen Herkunft, dem Wohnbezirk, dem Bevölkerungstypus, und im weiteren Sinne von der Lebensart und dem -standard.

Im weiteren Verlauf dieses Kapitels bringen wir einige Meßverfahren, die eine Bewertung des sozio-ökonomischen Status erlauben. Diese Bewertungen sind relativ wenig differenziert, aber sie geben dennoch das Bestreben um Objektivität wieder und helfen auf diese Weise einstweilig dem Forscher, der

in diesem Bereich allzu oft gewagten Spekulationen ohne präzise Bedeutung preisgegeben ist.

A. Die beruflichen Klassifizierungen

Sie werden häufig als Bewertungsgrundlage für das sozio-ökonomische Niveau genommen, obgleich — wie wir soeben angedeutet haben — der Beruf dabei nur einer der Faktoren ist.

1. Einfache Klassifizierung¹³⁹

Privater Sektor

- Leiter eines Landwirtschaftsbetriebes von weniger als 12,5 ha;
- Leiter eines Landwirtschaftsbetriebes von 12,5 ha und mehr oder Gartenbauer;
- Einzelhändler oder Klein-Unternehmer, der weniger als 5 Personen beschäftigt;
- Leiter eines Industrie- oder Handelsunternehmens, der zwischen 5 und 49 Personen beschäftigt;
- Leiter eines Industrie- oder Handelsunternehmens, der 50 und mehr Personen beschäftigt;
- Freiberufliche Tätigkeit;
- Anderer selbständiger Beruf (Makler, Sachverständiger ...).

Öffentliche oder private Sektoren

Privat — Öffentlich

- Nicht-qualifizierter Arbeiter;
 - Facharbeiter, Vorarbeiter, Werkmeister;
 - Beamter oder subalternen Angestellter (Büroangestellter, Polizei);
 - Beamter oder qualifizierter Angestellter des mittleren Dienstes: Volksschullehrer, Mittelschullehrer, Gymnasiallehrer der Unterstufe;
 - Gymnasiallehrer der Oberstufe;
 - Höherer Beamter, Führungskräfte des gehobenen Dienstes;
 - Hochschullehrer, Ministerialbeamter.
- Nicht berücksichtigt wurde in obiger Aufstellung die Angabe: „Ohne Beruf“.

139 Univ. Lüttich, Ergänzungsformular zum Einschreibungsschein für das Studium im akad. Jahr 1962–1963. Diese Liste ist keine Skala im eigentlichen Sinn. Sie erlaubt dennoch eine schnelle und ziemlich genaue Klassifizierung; in dieser Eigenschaft kann sie für die laufende Forschung von Nutzen sein. Siehe vergleichsweise die weniger genaue, vom Institut Pédagogique National de France angewandte Klassifizierung: Landwirte, landwirtschaftliche Arbeiter, Handels- und Industrieunternehmer: a) Industrielle; b) Großhändler; c) Handwerker — freie Berufe und Führungskräfte des gehobenen Dienstes — mittlere Dienste — Angestellte — Arbeiter — Dienstpersonal — Rentner, ohne Beruf — andere Berufsgruppen.

2. Skala von Beckman¹⁴⁰

- I. Nicht-qualifizierte manuelle Tätigkeiten: Hofknecht, Hilfsarbeiter usw.
- II. Halb-qualifizierte Tätigkeiten: Fischer, Scherenschleifer ...
- III. A. Qualifizierte manuelle Tätigkeiten: Landwirt, Bäcker ...
B. Qualifizierte administrative Tätigkeiten: Angestellter, Telephonist, Telegraphist ...
- IV. A. Halbfrei-berufliche Tätigkeiten: Optiker, Schauspieler ...
B. Kommerzielle Tätigkeiten: Besitzer einer Autoreparaturwerkstatt, Transportunternehmer ...
C. Leitende Posten niedere Stufe: Vorarbeiter, Werkmeister ...
- V. A. Frei-berufliche Tätigkeiten: Literarischer Zweig: Schriftsteller, Pfarrer, Hochschullehrer ...
B. Freie Tätigkeiten: wissenschaftlicher Zweig: Architekt, Apotheker ...
C. Führungskräfte des gehobenen Dienstes (Exekutive).

Diese Skala läßt eine Reihe von Ungenauigkeiten, wenn nicht sogar von Klassifizierungsfehlern entstehen, vor allem in Punkt IV. und V. Wir zitieren sie trotzdem, weil sie seit ihrer Veröffentlichung einige andere Aufstellungen inspiriert zu haben scheint, doch können auch diese nicht zufriedenstellen. Tatsächlich handelt es sich mehr um einen quantitativen Bewertungsversuch des Prestiges als um das sozio-ökonomische Niveau im eigentlichen Sinne.

B. Die sozio-ökonomischen Indizes

Einige Autoren schlagen Methoden vor, die — ausgehend von einer Reihe von Faktoren, die in ihrer Zahl weitmöglichst reduziert sind — eine schnelle Errechnung des sozio-ökonomischen Index erlauben.

1. Kerr-Remmers American Home Scale (SRA)

Diese Skala ist aufgebaut auf einem Fragebogen mit 50 Items, die sich auf das kulturelle, das ökonomische und das ästhetische Niveau erstrecken. Sie bietet den Vorteil, daß sie auch für den Nicht-Experten zu handhaben ist.

Itemsbeispiele

- Besitzen Sie zu Hause: — einen Staubsauger? ja — nein
— einen Kühlschrank?
— ein Badezimmer oder eine Dusche mit fließendem Wasser?
— ein Telefon?
— ein Auto?

¹⁴⁰ Siehe: A New Scale for Gauging Occupational Rank (Personnel Journal, Nr. 13, 1934, S. 225—233, zitiert nach Remmers u. Gage, Educational Measurement and Evaluation, New York, Harper, 1955, rev. Aufl.

Lassen Ihre Eltern Ihnen Privatstunden neben dem Schulunterricht erteilen? (Tanzstunde, Schauspielunterricht, Rezitationsunterricht, Musikunterricht ...)?

Es ist jedoch bekannt, daß viele Prüflinge bei der Beantwortung solcher Fragebogen unaufrichtig sind. Die Kriterien des Komforts, des Wohlstandes und des kulturellen Niveaus variieren zudem je nach Land und mitunter selbst nach regionalen Zonen.

2. Minnesota Home Status Index (Univ. of Minnesota Press, Minneapolis)

Dieses Untersuchungsverfahren ist präziser als das vorher beschriebene, zugleich aber auch weit schwieriger zu handhaben, denn es stützt sich auf ein Interview, welches 50 Fragen umfaßt (den Kindern gebotene Vergünstigungen, wirtschaftliche Situation, kulturelle Tätigkeiten, Sozialstatus, Beschäftigung, Bildungsstand der Eltern).

3. Die Formel von Warner

Zur Erstellung eines präziseren und zuverlässigeren Bezugssystems unterscheidet Lloyd Warner sechs Sozialschichten¹⁴¹ anstelle der drei herkömmlichen:

Höhere Populationsschicht

1. höheres Niveau (upper upper)
2. niederes Niveau (lower upper)

Mittlere Populationsschicht

1. höheres Niveau (upper middle)
2. niederes Niveau (lower middle): Durchschnittsmensch. Solche, die die obere Schicht der Arbeiterklasse bilden; gute Familien, die aber gesellschaftlich nicht zählen.

Niedere Populationsschicht

1. höheres Niveau (upper lower): die kleinen Leute, solche, die hart arbeiten, aber wenig verdienen.
2. niederes Niveau (lower lower): die Armen, solche, die von öffentlicher oder privater Unterstützung leben. Solche, die von der Hand in den Mund leben.

Nach Warner und seinen Mitarbeitern erlaubt die folgende Formel zu 90% aller Fälle eine exakte Bestimmung des sozialen Status, und zwar für das städtische Milieu.

¹⁴¹ L. Warner, M. Meeker u. K. Eels, Social Class in America, Chicago, SRA, 1949. Der Begriff soziale Klasse ist äußerst komplex. Die Klassifizierung von Warner hat in unseren Augen vor allem einen operationalen Wert.

Man ordnet zunächst Punkte zu, und zwar je nach der jeweiligen Beschäftigung bzw. der Einkommensquelle, den Wohnverhältnissen und der nachbarlichen Umgebung:

<i>Beschäftigung</i>	<i>Punktzahl</i>
Freie Berufe. Inhaber bedeutender Handelsunternehmen	1
Halbfreie Berufe, mittlerer Dienst	2
Angestellte und Arbeiter im Angestelltenverhältnis	3
Facharbeiter	4
Inhaber kleiner Handelsunternehmen	5
Angelernte Arbeiter	6
Nicht-qualifizierte Arbeiter	7
<i>Einkommensquelle</i>	
Ererbter Besitz	1
Erworbener Besitz	2
Gewinn und sonstige Zuwendungen	3
Gehälter	4
Löhne	5
Private Unterstützung	6
Öffentliche Unterstützung	7
<i>Wohnverhältnisse</i>	
Erstklassig: groß, schöner Garten, gut gepflegt	1
sehr komfortabel: etwas kleiner als vorige	2
komfortabel: etwas größer als nötig	3
mittelmäßig	4
wenig komfortabel	5
ärmlich	6
sehr ärmlich, baufällig	7
<i>Nachbarliche Umgebung</i>	
Erstklassige Wohngegend in der Stadt	1
Überdurchschnittlich guter Wohnbezirk	2
Ansehnliches und schönes Wohngebiet, aber nicht von der oberen Gesellschaftsschicht bewohnt	3
Mittelmäßige Wohngegend, vor allem von der Arbeiterschicht bewohnt	4
Wohnbezirk in Industrie- und Eisenbahnnähe, sehr gemischte Population	5
An das Armenviertel grenzender Wohnbezirk	6
Armenviertel	7
<i>Rechenvorschrift</i>	
— Beschäftigungspunktzahl x 4 =	
— Einkommenspunktzahl x 3 =	
— Wohnungspunktzahl x 3 =	
— Wohngegendpunktzahl x 2	
Gesamtergebnis = Punktzahl des sozialen Status	

*Interpretation*¹⁴²

- 12–22: Obere Schicht
- 25–34: Obere Mittelschicht
- 37–50: Untere Mittelschicht
- 54–63: Untere Schicht, höheres Niveau
- 67–84: Untere Schicht, niederes Niveau.

Weder das Bildungsniveau noch die Höhe des Einkommens werden direkt berücksichtigt. Die Verfasser sind der Meinung, daß die Gesamtpunktzahl, so wie sie ist, dem sozialen Status hinreichend Rechnung trägt, wobei die Art der Beschäftigung wichtiger ist als die Besoldung. Maccoly, Gibbs u. a.¹⁴³ haben die anhand der Formel von Warner erhaltene Punktzahl (Gewichtung: 2) mit einer konventionellen Punktzahl entsprechend dem Einkommen (Gewichtung: 1) kombiniert. Sie stellen so eine Skala für neun Populationsschichten auf.

C. Indizes des Prestiges und des sozio-ökonomischen Status
Die Klassifizierung „NORC“

Im Jahre 1961 haben Reiss u. a. außerordentlich differenzierte Skalen als Ergebnis der Arbeiten des National Opinion Research Center (NORC)¹⁴⁴ publiziert. Trotz unvermeidlicher Approximationen stellen sie dennoch die derzeit präzisesten und am meisten differenzierten Untersuchungsverfahren dar, über die man verfügt.

Im Jahre 1947 führte das NORC eine umfangreiche Erhebung über die Relation zwischen Beruf und sozialem Status durch. Man benötigte mehr als 10 Jahre, um die gesammelten Informationen auszuwerten.

Das Werk von Reiss schlägt eine sehr umfassende Berufsaufstellung mit spezifizierten Angaben hinsichtlich des Einkommens, der Bildung, des Prestiges und des sozio-ökonomischen Status vor. Da in diesem Rahmen eine Gesamtwiedergabe der Klassifikation nicht möglich war, haben wir die nachfolgende Tabelle zusammengestellt, die bereits eine ziemlich repräsentative Berufsstichprobe darstellt¹⁴⁵.

- 142 Warner unterscheidet hier bei der oberen Schicht keine zwei Niveaus, weil die Stichprobe, auf die sich seine Umfrage erstreckte, zur Ermittlung einer statistisch signifikanten Differenz nicht genügend Fälle für diese Niveaus enthielt. Die leeren Intervalle (z. B.: 22–25) entsprechen unbestimmten bzw. Übergangszonen.
- 143 Siehe: *Methods of Child-Rearing in two Social Classes* (Martin et Stendler, Readings in Child Development, Harcourt, Brace and Co., 1954).
- 144 A. Reiss Jr., O. Duncan, P. Hatt u. C. North, *Occupation and Social Status*, Free Press of Glenco, Inc. 1961.
- 145 Die Angaben auf dieser Tabelle erscheinen im Originalwerk auf S. 122–123.

Streng genommen, sind die aufgeführten Angaben nur für die USA gültig. Es scheint jedoch, daß sich die westeuropäischen Verhältnisse nicht grundsätzlich von den amerikanischen unterscheiden. Offenbar besteht aber das Bedürfnis, eine Validierung für unsere Länder in Angriff zu nehmen.

Durch den „Norc“* vorgenommene Schätzung des Prestige auf der Grundlage: Einkommens- und Bildungsniveau

Beschäftigungen	Einkommen	Bildung	Prestige (NORC)	Sozio-ökonomischer Index
	a	b	c	d
Architekt	75	92	90	90
Barmann	16	28	7	19
Bergmann				
Briefträger	48	55	34	53
Buchhalter, gehobener Dienst	62	86	82	78
Buchhalter, angestellt	29	72	39	51
Chemiker	64	86	90	79
Chirurg	76	97	97	92
Doktor der Rechtswissenschaft — Richter	76	98	89	93
Elektriker	47	39	53	44
Flugzeugpilot — Schiffskapitän	72	76	83	79
Friseur	16	26	20	17
Geschäftsführer, Besitzer eines Unternehmens				
— Baugewerbe (Unternehmer)	53	45	76	51
— Fabrik mit etwa 100 Beschäftigten	60	56	81	61
— Kleiner Einzelhandel	42	44	45	43
— Bank- und Finanzwesen	78	82	92	85
Hochschullehrer (Rektor, Professor)	64	93	93	84
Installateur, Rohrleger	44	25	29	34
Kellner (Kaffee, Restaurant)	8	32	10	16
Koch (Restaurant)	14	22	16	15
Lastkraftwagenfahrer	21	15	13	15
Maschinenmeister (Lokomotivführer)	81	28	67	58
Mechaniker (Werkzeugmaschinen)	36	32	57	33
Monteur	21	20	24	17
Nachtwächter — Portier	17	25	11	18
Redakteur (Tageszeitung)	67	87	52	82
Schaffner (Eisenbahn)	76	34	38	58

*) National Opinion Research Center.

Beschäftigungen	Einkommen	Bildung	Prestige (NORC)	Sozio-ökonomischer Index
	a	b	c	d
Schriftsteller (Romane)	55	90	76	76
Schuhputzer	9	17	3	8
Sicherungsbeamter	34	47	41	40
Sozialdienst	41	84	59	64
Tankwart	15	29	10	19
Taxichauffeur	9	19	10	10
Verkaufspersonal (Kaffee)	12	30	6	17
Verkäufer (Einzelhandel)	29	50	16	39
Versicherungsagent oder Makler	55	71	41	66
Volksschullehrer — Lehrer der höheren Schule	48	91	73	72
Zahnarzt (Dentist)	80	100	90	96
Zimmermann	21	23	33	19
Zivilingenieur	72	86	88	84

- a Prozentsatz der Personen, die im Jahre 1949 ein Jahreseinkommen von mehr als 3 500 Dollar hatten — die Eichungen wurden im Hinblick auf das Alter vorgenommen.
- b Prozentsatz der Personen, die im Jahre 1950 ihre höhere Schulbildung abgeschlossen haben (High School) — Eichungen wurden im Hinblick auf das Alter vorgenommen.
- c Prozentsatz der Personen, die einen Fragebogen beantwortet und die Beschäftigung mit „ausgezeichnet“ oder „gut“ bewertet haben. Auf der Grundlage des Einkommens- und Bildungsniveaus.

III. Schlußfolgerung

Keines der vorbeschriebenen Untersuchungsverfahren erlaubt eine hundertprozentig zuverlässige Beurteilung. Zudem wird es eine Ideallösung zweifellos niemals geben; denn der sozio-ökonomische Status ist tatsächlich ebenso schwer zu erfassen wie die Persönlichkeit. Dennoch können diese Hilfsmittel, so unvollkommen sie noch sein mögen, schon erhebliche Dienste leisten.

Man erhebt manchmal den Einwand, daß die Identifizierung des sozio-ökonomischen Status durch den Lehrer oder durch den Untersucher — wenn gleich meist unbewußt — eine neue soziale Spaltung heraufbeschwört. Doch die Erfahrung lehrt, daß diese Befürchtung unbegründet ist.

Jedenfalls sind die Folgen weit schwerwiegender, wenn man die Schwierigkeiten unbeachtet läßt, die manchen Schülern allein wegen ihrer sozialen Herkunft erwachsen, oder wenn man Forschungsergebnisse als Beurteilungsvoraussetzung nimmt, die mangels Präzision dem Leser wohl allgemein gültig erscheinen mögen, aber in Wirklichkeit nur auf bestimmte Gruppen anwendbar sind.

Dritter Teil:

Die Behandlung und die Analyse der Daten

1. Die Inhaltsanalyse

I. Definition

Die Inhaltsanalyse (*content analysis*) hat eine objektive, systematische und quantitative Beschreibung einer symbolhaften Verhaltensweise zum Gegenstand: Kommunikation mittels Wort, Schrift, Geste, Mienenspiel usw.

Da die Schule im wesentlichen ein Ort der Kommunikation ist, erscheint es verständlich, daß die modernen Methoden der Inhaltsanalyse dazu angetan sind, in der pädagogischen Forschung eine immer größere Rolle zu spielen.

II. Anwendungsbereiche der Inhaltsanalyse

Zusammenfassend geben wir die von Berelson¹ in Vorschlag gebrachten Typen der Inhaltsanalyse wieder und erläutern sie anhand von pädagogischen Beispielen.

A. Analyse der Charakteristiken

1. Beschreibung der sukzessiven Tendenzen des Kommunikationsinhalts. Beispiel: Mit welchen Problemen hat sich die pädagogische Forschung im Laufe der letzten 50 Jahre auseinandergesetzt?
2. Identifizierung der internationalen Differenzen im Kommunikationsinhalt. Beispiel: Vergleichsstudie an Geschichtslehrbüchern.
3. Vergleich der Einstellungen, die durch unterschiedliche Medien der Verbreitung von Informationen übermittelt werden. Die großen Massenmedien (mass media), Zeitungen, Illustrierte, Radio, Fernsehen beeinflussen — direkt oder indirekt — eine Reihe von Bildungsspekten.
Forschungsbeispiel: Ist eine Verleumdungskampagne gegen eine pädagogische Methode eine allgemeine Reaktion, oder wird sie von einer bestimmten Gruppe gelenkt?
4. Normenerstellung. Beispiel: Es ist festzustellen, ob die Schulbücher oder der Unterricht der Lehrer mit bestimmten Kriterien der Objektivität, der Gewichtung und der Unparteilichkeit übereinstimmen.
5. Anwendung technischer Medien in der Forschung. Beispiele: Kodierung von Interviews mit freier Beantwortung.
Fehleranalyse von Schüleraufsätzen (in der Muttersprache usw.).

¹ B. Berelson, *Content Analysis in Communication Research*, Glencoe, The Free Press, 1952.

B. Formanalyse

1. Identifizierung von Propagandatechniken.
2. Messen der „Lesbarkeit“ von Kommunikationsmitteln. Es ist unnötig, noch mehr auf den Vorzug hinzuweisen, den eine objektive Bewertungstechnik des Schwierigkeitsgrades der in den Schulbüchern verwandten Sprache darstellt; zu berechnen, in welchem Maße die angewandten Wörter eine affektive Resonanz auf das Kind ausüben usw.
3. Identifizierung stilistischer Merkmale.

C. Analyse der Determinanten

Hier analysiert man den Inhalt der symbolischen Begriffe, um die Persönlichkeit ihrer Urheber kennenzulernen.

1. Untersuchung der Absichten und Einstellungen.
Beispiel: Eine Analyse der Aufsätze läßt – zumindest teilweise – die Einstellungen der Schüler gegenüber der Schule erkennen.
2. Untersuchung des psychologischen Status von Einzelpersonen sowie von Gruppen.
Die Analyse ist mitunter sehr einfach (Dollard und Mowrer haben einen Unruhequotienten errechnet, indem sie im Laufe eines Interviews mit dem Kind die Zahl der Wörter zählten, die dieses Gefühl ausdrückten), manchmal außerordentlich komplex.
(Beispiel: Projektivtechniken in der Psychologie: Analyse der Kommentare zu den Bildtafeln des TAT oder des Rohrschachttests).

D. Untersuchung über die Wirkung der Massenmedien

Die Zeitungen, Illustrierten, die Radioprogramme oder Filmveranstaltungen sind auf die Psychologie der Masse zugeschnitten, an die sie sich wenden. Die Studie dieser Dokumente oder Filme erlaubt daher, eine Reihe von Rückschlüssen auf die Einstellungen, die Wertsysteme und Interessen dieser Populationsschicht zu ziehen.

III. Kategorisierungssysteme

Lazarfeld und Barton, die von Cartwright zitiert werden, unterscheiden drei Kategorisierungssysteme, die einen qualitativen Stoff zu klassifizieren erlauben:

A. Dichotomien

Sie bestehen aus der Feststellung des Vorhandenseins oder Fehlens eines besonderen Merkmals.

Beispiel:

- Ist die von einem Verfasser gebrachte Information richtig oder falsch?
 - Verrät die Verhaltensweise eines Schülers Ängstlichkeit oder nicht?
- Wenn sich die Beobachtung auf komplexe Phänomene erstreckt, können Definitionen oder typische Antworten die Auswertung erleichtern.

B. Die Skalierungsreihen

Anstatt sich mit der Feststellung über das Vorhandensein oder Fehlen eines Gefühls, einer Meinung bzw. eines Phänomens zufriedenzugeben, kann man versuchen, sie zu bewerten.

Man stützt sich dabei auf Skalen wie: starkes, mäßiges, schwaches Gefühl; völlig einverstanden, einverstanden, unentschieden, nicht einverstanden, durchaus nicht einverstanden; immer, gewöhnlich, gelegentlich, niemals. Jeder Grad muß vorher definiert werden, und zwar so präzise wie möglich (siehe *rating scales*).

C. Die Variablen²

Sie sind Skalen, die gleiche Intervalle und einen absoluten Nullpunkt aufweisen.

IV. Beispiele für die Inhaltsanalyse

Es ist naheliegend, daß es eine Unzahl von Kriterien der Analyse gibt. Eine Quantifizierung des Textes kann darin bestehen, den Platz zu messen, den er auf eine Seite einnimmt, die Worte zu zählen, aus denen er sich zusammensetzt, die Ideen zu identifizieren, die er beinhaltet usw. Ebenso kann eine Verhaltensweise im Hinblick auf ihre Dauer, ihre Natur, ihre Richtung usw. betrachtet werden.

Wir beschränken uns darauf, drei Beispiele für die Inhaltsanalyse zu bringen, die für die pädagogische Forschung von unmittelbarem Interesse sind oder ihr als Anregung dienen können.

² Dieser Punkt wird nur als Erinnerung erwähnt: Er würde lange mathematische Entwicklungen erforderlich machen, die wir hier nicht vornehmen können.

A. Analyse der Determinanten: Rorschach-Test und TAT

Die Tests zur Erfassung der verbal geäußerten schöpferischen Vorstellungskraft – der bekannteste unter ihnen ist der „Tintenkleckstest“ – waren lange Zeit wenig aufschlußreich, da die Analyse der Antworten nur in einer Aufzählung der dargestellten Gegenstände: Personen, Tiere usw. bestand. Man sah sich somit statischen Daten gegenübergestellt, während das verfolgte Ziel darin bestand, die Dynamik der Persönlichkeit kennenzulernen. Es ist das Verdienst von Rorschach, eine Inhaltsanalyse erstellt zu haben, die dieser Anforderung entspricht.

Er schlug drei Bewertungskriterien vor, die sich gründen auf:

1. Erfassungsmodus (Ganzantwort, Detailantwort);
2. Determinanten (Antwort betrifft Form, Farbe, Bewegung oder Tiefe des Gegenstandes);
3. Inhaltsreihe (Blut, Pflanze, Landschaft usw.).

Die Ergebnisse der quantitativen Auswertung erlauben bereits eine Reihe von Rückschlüssen auf die Intelligenz und vor allem auf die Angepaßtheit des Probanden. Es folgt dann die dynamische und darauf die symbolische Interpretation.

Die Analyse des *Thematic Apperception Tests* von Murray stellt gleichfalls ein gutes methodologisches Beispiel dar.

Die Untersuchung der zu den verschiedenen Bildtafeln des TAT gegebenen Geschichten erfolgt nach folgenden Kriterien:

1. Motivationen und Gefühle des Helden: Herrschsucht, Unterwerfung, Unabhängigkeit, Freiheitsdrang, Affektion, Furcht, Schuldbewußtsein, introvertierte und extravertierte Wesensart usw.
2. Umwelteinfluß auf den Helden: Beherrschung, Beschützung, Ablehnung;
3. Verlauf und Ausgang der Geschichte: logisch oder unlogisch, Sieg oder Niederlage für den Helden, Konfliktlösung oder ungelöster Konflikt usw.;
4. Allgemeine Themen;
5. Interessen und Gefühle;

Diese allgemeinen Kriterien werden von den Psychologen auf verschiedene Art ausgewertet. Das folgende Beispiel³ zeigt ein sehr differenziertes Analysenschema, dessen Wert allein der erfahrene Praktiker ganz ausdeuten kann.

³ Entnommen aus: D. Anzieu: *Les methodes projectives*, P.U.F., 1960, S. 130 und S. 134 bis 135.

Geschichte zu Tafel V

„Hier öffnet die Mutter die Tür zum Kinderzimmer, um das Kind morgens zu wecken. Was war am Vorabend geschehen? Es hatte einen Streit zwischen Mutter und Kind gegeben. Die Mutter steht noch unter dem Einfluß dieser Auseinandersetzung. Aber das Kind schickt sich an, artig ‚guten Morgen‘ zu sagen, und alles ist vergessen.“

Analyse:

- Thema: Mutter und Kind wieder ausgesöhnt;
- Formale Analyse: ziemlich gut;
- Held: Mutter;
- Motivation: Familiäre Aussöhnung;
- Innerer Status: Ärger;
- Einfluß der Umwelt: Kind – emotionale Aggression;
- Haltung des Helden: passiv;
- Konfliktlösung: Unwillkürlicher Erfolg;
- Beobachtungen und Interpretation: Die Einigkeit mit dem Kind gleicht die Unstimmigkeit mit dem Gatten aus (die Untersuchung bezog sich auf eine Frau, die in dieser Situation war).

B. Analyse der Charakteristiken und der Determinanten: Untersuchung der Verhaltensweise von Lehrern⁴

M. Hughes und ihre Mitarbeiter haben eine bemerkenswerte Analyse der von den Lehrern ausgeübten Funktionen entwickelt⁵. Wir schlagen den Aufbauplan dieser Analyse als Beispiel vor.

1. Kontrollfunktionen

- a) Der Lehrer ordnet an: Der Pädagoge gibt an, was das Kind lernen und worauf es besonders achten soll.
- b) Regelt, lenkt: Der Lehrer gibt an, wer die eine oder andere Aufgabe übernimmt und wer auf diese oder jene Frage antworten soll.
- c) Schlägt ein Modell vor: Bestimmt oder erinnert an die geeignete und annehmbare Form der Ausführung der Aufgaben.
- d) Schlichtet einen Streit: Schaltet sich im Falle einer Auseinandersetzung der Schüler untereinander oder zwischen Lehrer und Schülern als Schiedsrichter ein.

2. Anordnende Funktionen (schreibt vor, obwohl es die Situation nicht erfordert)

- a) Lenkt, regelt selbst. Beispiel: Ein Schüler braucht einen Bleistift; der Lehrer läßt es nicht zu, daß er sich einen holt, sondern gibt ihm einen.

⁴ Siehe allgemeine Untersuchung zu dieser Frage, S. 185 ff.

⁵ M. Hughes et al., *Development of the Means for the Assessment of Teaching in Elementary Schools*, Salt Lake City, Univ. of Utah, 1959.

- b) *Moralisiert*: Äußert eine Meinung mit einem moralisierenden Unterton. Beispiel: „Du solltest nicht ...“; „Das ist leicht für Dich“; „Das dürfte Dich nicht viel Zeit kosten“.
- c) *Ist der Meinung, daß ein Bedarf besteht*: Gibt eine nicht erbetene Hilfestellung. Beispiel: Ein Schüler hat vergessen, seinen Namen auf sein Konzept zu schreiben; der Lehrer geht vorüber und schreibt den Namen selbst darauf.
- d) *Informiert*, ohne daß das Kind darum bittet oder daß die Situation es erfordert.
- e) *Schreibt ein Werturteil vor*, ohne daß das Kind die Gelegenheit gehabt hat, seine auf eigene Erfahrung gegründete Meinung auszusprechen.

3. Funktionen der Arbeitserleichterung

- a) *Erklärt, macht die Art und Weise bzw. die Reihenfolge deutlich*, nach der zu lernen ist. Beispiel: „Morgen wollen wir über die Muscheltiere sprechen“.
- b) *Demonstriert*. Zeigt, wie man vorgeht.
- c) *Kontrolliert neutral*:
 - Information: „Wer trägt sich für diese Aufgabe ein?“
 - Routine- bzw. rhetorische Fragen, auf die der Lehrer keine spezifische Antwort erwartet: „Sind wir fertig?“ „Verstanden?“

4. Funktionen der Inhaltsentwicklung

- a) *Erläutert*. Beispiel: Der Lehrer wiederholt mit eigenen Worten, das, was das Kind gesagt hat.
- b) *Regt an*: Schlägt verschiedene Dinge vor, die die Klasse machen könnte, überläßt jedoch die Entscheidung den Schülern.
- c) *Drückt Anerkennung aus*: „Zeichnet eine Karte, das ist eine gute Idee!“
- d) *Agiert als Informationsquelle*: Das Kind wendet sich an ihn, um eine Erklärung oder Hilfestellung zu erhalten.

5. Funktionen persönlicher Antworten

- a) *Antwortet auf die Frage eines Kindes*.
- b) *Geht auf persönliche Erlebnisse ein*: „Hast Du Dir die Enten bei Deinem Großvater angesehen?“ „Du plagst Dich für Deinen Bruder, nicht wahr?“
- c) *Interpretiert die Situation*: „Warte einen Augenblick, wir haben nur eine Schere. Sobald N. fertig ist, bekommst Du sie“.
- d) *Gibt zu, daß er sich geirrt hat*.

6. Positive affektive Funktionen

- a) *Anerkennung*
- b) *Zeigt Fürsorge*: „Soll ich Dir helfen?“
- c) *Ermutigt*: „Das geht schon gut, mach nur so weiter!“

7. Negative affektive Funktionen

- a) *Ermahnt*: „Vergiß nicht, die Tür zu schließen!“
- b) *Erteilt einen Verweis* (Fordert zu einer Verhaltensänderung auf).
- c) *Rügt*: „Du hast nicht fleißig genug gearbeitet“.
- d) *Negative persönliche Antwort*: „Nein, Du darfst Dich nicht neben Jean setzen“.
- e) *Vertröstet*: „Wir werden später sehen“.

C. Formanalyse: Die Tests zur Messung der Lesbarkeit (readability)

1. Einführung

Besteht die Möglichkeit oder nicht, mit Hilfe einer einfachen Formel den Schwierigkeitsgrad eines Textes, sein Abstraktionsniveau zu messen und festzustellen, inwieweit der Text für den Durchschnittsleser von Interesse ist? Die Antwort auf diese Frage ist für die Pädagogik von sehr großer Bedeutung, und zwar im Hinblick auf die Abfassung von Schulbüchern, auf die Instruktionen für die Testanwendung, hinsichtlich der Prüfungsfragen usw. Bis zum gegenwärtigen Zeitpunkt konnten sich die Verfasser von Lehrbüchern bei der Beurteilung des Schwierigkeitsgrads ihrer Texte lediglich auf ihr Feingefühl und auf die langwierigen und unzureichenden Erfahrungen von Versuch und Irrtum stützen. Außerdem können sie selbst, da sie sich der Lesekunst gewidmet haben, die einen Bestandteil ihres Berufes darstellen, sich oft kaum vorstellen, welche Schwierigkeiten scheinbar leicht zu bewältigende Texte für einen ungeübten Leser haben⁶.

Es ist das Verdienst von M. Vogel und C. Washburne, das erste objektive Bewertungsverfahren zur Bestimmung des Schwierigkeitsgrades eines Textes⁷ erstellt zu haben.

Anhand einer Aufstellung von 1000 aufeinanderfolgenden Wörtern zählen sie die Präpositionen, die verschiedenen Vokabeln, sodann diejenigen Wörter, die in der Liste der 10000 Wörter von Thorndike nicht aufgeführt sind; gleichfalls bestimmen sie in 75 aufeinanderfolgenden Sätzen die Zahl der einfachen Sätze. Eine Kombination der erhaltenen Resultate führte zu einem Punktwert für die Schwierigkeit⁸.

6 Das Problem der Schwierigkeit von Texten geht weit über den Aufgabenbereich der Schule hinaus. Bereits im Jahre 1930 ergab eine in den Vereinigten Staaten durchgeführte Umfrage, daß die Hälfte der existierenden Texte für 50 % der Erwachsenen zu schwierig war. Siehe W. A. Gray, Reading (Enc. of Educ. Res., New York, Macmillan, 1960, S. 1094). Zwanzig Jahre später haben Michaelis u. Tyler gezeigt, daß die vom O.N.U. für eine breite Öffentlichkeit bestimmten Veröffentlichungen zu schwierig waren, um normalerweise von amerik. Schülern der Oberstufe des höheren Schulunterrichts gelesen zu werden. Siehe Michaelis u. Tyler, A comparison of Reading, Ability and Readability (Journal of Educational Psychology, 42, S. 491–498, 1951, zitiert von Gray, op. cit.).

7 Siehe: An Objective Method of Determining Grade Placement of Children's Reading Material (Elementary School Journal, Nr. 28, 1929).

8 Zehn Jahre später haben C. Washburne u. M. Vogel das Problem wieder aufgegriffen u. das Zählen von Präpositionen abgeschafft: Siehe Grade Placement of Children's Books (Elementary School Journal, Nr. 38, S. 355–364). Anm. d. dtsh. Bearbeiters: Vgl. auch: LT 2 + (Lesetest für 2. Schj.) u. VL 5–6 u. VL 7–9 (Verständiges Lesen im 5.–9. Schj.) Dtsch. Schultests, Weinheim 1967/68.

Im weiteren Verlauf wurden von etlichen Verfassern leichtere Lösungen in Vorschlag gebracht, insbesondere sind hier W. Gray und B. Leary⁹ sowie I. Lorge¹⁰, E. Dale, J. Chall¹¹ und R. Flesch¹² zu nennen.

Die Arbeiten des letzteren sollen uns dabei als Beispiel dienen. Da die Meßtechnik der Lesbarkeit in Europa bisher wenig bekannt ist, wird das Verfahren von Flesch ziemlich detailliert behandelt. Dabei sind notwendige Abänderungen vorgesehen, um es auf die deutsche Sprache anwenden zu können.

2. Der Test zur Messung der Lesefähigkeit von Flesch

In seinem Werk *How to test readability* unterscheidet Flesch zwei Gesichtspunkte: „Den Leichtigkeitgrad, mit dem ein Text gelesen und verstanden wird“ und das „menschliche Interesse des Lesers an dem Text (besser gesagt an dem Gegenstand selbst)“ (Seite 1).

a) Der Leichtigkeitsindex

Errechnung: Zur Erstellung des Leichtigkeitsindex testet man entweder den ganzen Text – sofern er kurz ist – oder eine Reihe von Stichproben von 100 aufeinanderfolgenden nach dem Zufallsprinzip gewählten Wörtern.

Man bestimmt die Durchschnittszahl der Wörter pro Satz sowie die durchschnittliche Silbenzahl für 100 Wörter.

$$\begin{array}{l} \text{Formel}^{13}: \text{Durchschnittliche Länge der Sätze} \text{ mal } 1,015 = x \\ \text{Zahl der Silben für 100 Wörter} \quad \text{mal } 0,846 = y \\ \hline x + y \end{array}$$

$$\text{Index} = 206,835 - (x+y).$$

Flesch stützt sich nur auf formale Elemente. Folglich – wenn er von der „Leichtigkeit“ eines Textes spricht – sind es vor allem die Einfachheit der

9 W. Gray u. B. Leary, *What makes a book readable*, Chicago, Univ. Press, 1935.

10 I. Lorge, *Predicting Readability* (Teachers College Record, 1944, 45, S. 404–419).

11 E. Dale u. J. Chall, *A Formula for Predicting Readability* (Educ. Research Bulletin, Ohio State Univ., Jan. u. Febr. 1948). Das Verfahren von Dale-Chall wird heute vielfach neben dem von Flesch, von dem noch die Rede sein wird, angewandt. Für J. Chall sind die wesentlichen Schwierigkeitsfaktoren: Die Verschiedenartigkeit und die Schwierigkeit des Wortschatzes, die Satzkonstruktion, die Komprimierung der Ideen, das menschliche Interesse. Siehe J. Chall, *The Business of Readability, a second look* (Educ. Res. Bulletin, Nr. 35, April 1956, S. 197–212).

12 R. Flesch, *The Art of Plain Talk*, New York, Harper & Br., 1946.

idem, *The Art of Readable Writing*, Harper & Br., 1949

idem, *How to test Readability*, Harper & Br., 1951

idem, *How to write, speak and talk more effectively*, 1960.

13 Mit Genehmigung des Autors zitierte Koeffizienten. Eine graphische Skala (*How to test Readability*, S. 5) verkürzt die Errechnungen erheblich.

Syntax, die Kürze der visuellen Vorstellung, an die wir denken¹⁴. Und sei es auch nur um dieser einen Eigenschaft willen, so ist sein Untersuchungsverfahren schon wertvoll.

Aber es trägt nicht nur der Form Rechnung. Flesch schreibt: „Der Test mißt die Länge der Wörter, weil dies die Bewertung des Schwierigkeitsgrades bequem verkürzt. Tatsächlich sind im Englischen die kurzen Wörter meistens am leichtesten zu lesen und zu verstehen“ (Seite 40). Eine solche Verallgemeinerung erscheint gefährlich. Es besteht jedoch kein Zweifel darüber, daß viele kunstvoll ausgeklügelte Wörter mit Hilfe von Vor- und Nachsilben konstruiert sind; die durchschnittliche Wortlänge eines abstrakten Textes liegt daher über der eines konkreten Textes. Darüber hinaus sind große stilistische Leistungen kaum denkbar ohne syntaktische Verlängerung.

Wir sind daher der Meinung, daß diese scheinbar so naiven Berechnungen zu einer wirksamen Sondierung einer weitaus komplexeren Realität führen können, als es zunächst den Anschein hat.

Amerikanische Eichung: Wir haben in der nachfolgenden Aufstellung einige in dem Werk von Flesch zerstreut vorkommende Daten in einer Reihe zusammengestellt. Die Punktwerte zwischen 100 und 120 erscheinen nicht in der Originalskala; sie werden jedoch vom Verfasser zitiert, und wir halten es für angebracht, sie mit einzubeziehen (s. S. 228).

b) Der Index des menschlichen Interesses

Nach Ansicht von Flesch ist der Index des menschlichen Interesses noch wichtiger als der Leichtigkeitsindex, denn er mißt die Motivation, die zur Überwindung der Schwierigkeit führt.

Zur Erstellung dieses Index zählt man zunächst die „persönlichen Wörter“, d. h. die Vor- und Zunamen, die eine Person bezeichnen und die maskulin von feminin unterscheiden, die persönlichen Fürwörter und die Adjektive, die sich auf eine Person beziehen. Man erstellt sodann den prozentualen Anteil der „persönlichen Sätze“: direkte Rede, Aufforderungen oder Bitten, die an den Leser gerichtet sind usw.¹⁵.

14 Flesch definiert seine Position eindeutig: „Der Test berechnet einfach die Lesbarkeit; er sagt nichts darüber aus, ob die darin zum Ausdruck gebrachten Ideen absurd sind oder nicht“ (S. 48). Ebensovienig beabsichtigt er, die ästhetische Qualität zu messen (siehe S. 41).

15 Formel: Anzahl der personbezogenen Wörter auf 100 Wörter $\times 3,635 = x$
Anzahl der personbezogenen Sätze auf 100 Wörter $\times 0,314 = y$, Index = $x + y$.

Amerikanische Eichung

Kinder	Leichtigkeitsindex	Interpretation	Textarten	Prozentsatz der Leser, die diese Texte verstehen (grobe Schätzung)
6 Jahre	...120...			
7 Jahre	...115...			
8 Jahre	...110...			
9 Jahre	...105...			
10 Jahre	...100...	sehr einfach	Komikhefte	93%
11 Jahre	... 90 ...	leicht	Kioskliteratur ^a	91%
12 Jahre	... 80 ...	ziemlich leicht	Zukunftsromane ^b	88%
	... 70 ...			
Höherer Schulunterricht — Unterstufe	... 60 ...	mittelschwer	Readers Digest, Times, Tatsachenberichte	83%
Höherer Schulunterricht — Oberstufe	... 50 ...	ziemlich schwer	Zeitschriften mit ziemlich hohem Niveau, Harper's z.B.	54%
Studenten der Universität	... 30 ...	schwer	akademisch	33%
Personen mit Hochschuldiplom	... 0 ...	sehr schwer	wissenschaftlich freie Berufe	4,5%

a Pulp fiction (amerik. Slang): Anspielung auf das aus Holzabfällen hergestellte Rohpapier, das zum Drucken von Literatur mit sehr niedrigem Niveau verwendet wird.

b Slick fiction (amerik. Slang): Anspielung auf satiniertes Papier, das zum Drucken von Romanliteratur mit höherem Niveau als dem vorhergehenden verwendet wird.

Inwieweit bezeugen die von Flesch ermittelten Faktoren tatsächlich das Interesse bei dem Durchschnittsleser?

Auch der Index der menschlichen Interessen gibt nur einen formalen Aspekt, nur eine unter vielen anderen Motivationen mögliche wieder. Gleichwie eine kleine Anekdote eine Gesellschaft unterhält, gestalten die Dialoge, die dramatischen Höhepunkte einen Text oftmals viel attraktiver. Aber systematische Untersuchungen müssen zeigen, bis zu welchem Punkt diese Verfahren wirksam sind.

Wir beschränken uns daher wohlweislich auf den Hinweis der Indizes, entsprechend den dargelegten Beispielen.

Flesch schlägt die folgende Eichung vor:

... 100 ...	sehr spannend	Zukunftsroman
... 60 ...		
... 40 ...	sehr interessant	Die Zeitschrift: „New Yorker“
... 20 ...	interessant	Die Digests; Times
... 10 ...	wenig interessant	Kommerzielle Schriften
... 0 ...	langweilig	wissenschaftliche Texte.

Diese Gradierung zeigt deutlich, daß der Begriff des Interesses an der Reaktion eines Durchschnittstyps, eines Nicht-Spezialisten oder an einem Laien gemessen ist.

c) Anwendungsversuch im französischen u. deutschen Sprachbereich^{15a}

Anpassung

1. Wir haben eine große Anzahl französischer u. deutscher Originaltexte getestet, mit dem Ziel, zu verifizieren, ob die durch die Indizes angegebene Gradierung der subjektiven und sehr vagen Bewertung, die andere oder wir selbst gegeben hatten, entsprach.

15a Es scheint, daß die Formel noch leichter für die deutsche Sprache anwendbar ist. Die Versuche, die wir unternommen haben, sind in jedem Fall sehr ermutigend. Man wird in der Tat erkennen, daß wir für die drei aufgeführten Beispiele (Seite 231–233) den Leichtigkeitsindex für die deutsche Übersetzung errechnet haben (Anm. d. dtisch. Bearbeiters.)

Für die franz. Anpassung siehe:

G. de Landsheere: Pour une application des tests de lisibilité de Flesch à la langue française, in „Le travail humain“, Paris, Presses Universitaires de France, n° 1–2, 1963.

G. de Landsheere: Recherche sur l'évaluation objective de la lisibilité des manuels scolaires et des tests, in XI^e Colloque international de l'Association Internationale de Pédagogie Expérimentale de Langue Française, 1964, S. 73–97.

Ders.: Lecteurs et lectures: Recherches expérimentales, XIII Colloque international de l'A.I.P.E.L.F., 1966, S. 139–167.

2. Wir haben Texte im Original und in einer vereinfachten Form miteinander verglichen, vereinfacht einmal durch eine stilistische Abwandlung, zum andern durch Herausnahme gewisser unwesentlicher Ideen, und haben geprüft, inwieweit die Indizes variieren.

Veränderungen der Berechnungsweise von Flesch haben sich als notwendig erwiesen:

1. Da beim Zählen der englischen Wörter die Zusammenziehung wie „don't“ als eine Einheit genommen wurden, haben wir uns dazu entschlossen, die Kurzformen (l'heure, j'avais, qu'il, usw.) wegzulassen. Die Jahreszahlen, Preise, Abkürzungen und die zusammengesetzten Wörter zählen nur als ein Wort (1964, 10 F, CQFD, week-end).
2. Beim Zählen der Silben geht Flesch von der Aussprache aus. Aber, das im Französischen stumme, häufig ausgelassene *e* stellt ein schwieriges Problem dar, zumal die regionalen Unterschiede mitunter beträchtlich sind. Aufgrund sehr zahlreicher Erfahrungen sind wir zu der Überzeugung gelangt, daß — selbst wenn man sich streng an Regeln wie die von L. Remacle¹⁶ zu halten entschließt — das Zählen umständlich, zeitraubend und trotz allem häufig anfechtbar ist.

Die Lösung ist — so glauben wir — sehr einfach. Flesch begeht hier nach unserer Meinung einen Fehler, nämlich insofern, als er von der Aussprache ausgeht, um das stille Lesen zu untersuchen. Es ist bekannt, daß hierbei die Geschwindigkeit, mit der ein Text verstanden wird, den Fortschritt bestimmt: Man überspringt oftmals das eine oder andere Wort, welches sich aus dem Zusammenhang ergibt¹⁷.

Unter diesem Gesichtspunkt betrachtet, steht dem Zählen von Silben nichts mehr im Wege (Beispiel: eine große Tochter = 6 Silben); und das haben wir getan¹⁸.

Fügen wir noch hinzu, daß die Art und Weise, wie man die Abkürzungen liest, entscheidend ist (z. B. = 3 Silben). Als eine Silbe zählen die folgenden Gruppen, die phonetisch mit einem Halbkonsonanten beginnen: *ié, ieu, ien, ion, oi, ui, oui, oin, ouin, uin*...

3. Beim Zählen der persönlichen Fürwörter haben wir die Reflexivpronomina nicht berücksichtigt, die sich im Französischen nicht nach dem Geschlecht richten.
4. Indem er seine Untersuchung zur Definition des Satzes auf das doppelte Kriterium, den Sinn und die Formgebung, aufbaut, öffnet Flesch Tür und Tor für zahlreiche Erörterungen, die die praktische Anwendung des Verfahrens zu bremsen scheinen. So betrachtet Flesch: „Therefore I say unto you, take no thought for your life ...“ als einen einzigen Satz (Seite 11).

16 L. Remacle, *Orthophonie française*, Lüttich, Michiels, 1948, S. 101ff.

17 Wir meinen, auch die Hypothese aufstellen zu können, daß für einen geschulten und intelligenten Leser der am leichtesten lesbare Text nicht unbedingt aus kurzen Sätzen zu bestehen braucht. Eine übertriebene Zerstückelung kann tatsächlich den Denkprozeß behindern.

18 Ohne Maschine wird einem das Zählen schnell zum Überdruß. Es hat sich gezeigt, daß die an der Schreibmaschine angebrachte Gradenteilung als Registrator dienen kann. Es genügt, wenn man leise liest, indem man bei jeder Silbe einen Trennungsstrich anschlägt. Kleine Ungenauigkeiten beim Zählen sind fast unvermeidlich, sie beeinflussen die Punktzahlen jedoch nicht wesentlich. Bei umfangreicheren Forschungsvorhaben empfiehlt sich der Einsatz eines Computers.

Demgegenüber betrachtet er als zwei Sätze: „... the hall resounded with these words: / We have been misdirected ...“ (Seite 15). In beiden Fällen handelt es sich um Beispiele in Form einer direkten Rede, beim ersten Beispiel ist ein Komma gesetzt anstelle des sonst üblichen Doppelpunktes.

In Ermangelung einer einfachen und klaren Definition des Satzes¹⁹, haben wir uns eigenmächtig dazu entschlossen, nur solche als ganze Sätze zu betrachten, bei denen ein zusammenhängender Gedankengang mit einem Punkt, einem Fragezeichen oder einem Ausrufezeichen endet. Wir stimmen daher mit F. Brunot überein, der als Beispiel eines einzigen Satzes eine Periode von „La Bruyère“ zitiert, die zwei durch einen Strichpunkt bzw. Doppelpunkt getrennte Sätze als ganzen Satz ansieht²⁰, Fälle also, bei denen Flesch zwei verschiedene Sätze unterschieden hätte.

Beispiele: Es folgen einige original-französische Texte, die nach ihrem Leichtigkeitsindex klassifiziert sind.

Der Wandel der Sprache und des kulturellen Milieus sowie die an den Zählregeln vorgenommenen Abänderungen machen die Eichungen von Flesch fragwürdig. Die erstellten Indizes besitzen daher nur einen relativen Wert und müssen in ihrer Relation betrachtet werden.

Wir hätten zunächst gerne einen Text mit dem von Flesch höchsten Leichtigkeitsindex vorgestellt, nämlich 120²¹. Wir haben in den uns zur Verfügung stehenden Schulbüchern vergeblich nach einem Text mit einem solchen Leichtigkeitsgrad gesucht.

Theoretische Konstruktionsversuche haben ergeben, daß der Leichtigkeitsindex von 120 einem Text entspricht, bei dem alle Wörter nur aus einer Silbe und alle Sätze aus zwei Wörtern bestehen, genauer Index: 120,205. Dieser Index stellt offenbar nur eine konventionelle Grenze dar. Die absolute Grenze wäre zweifellos durch einen Text erreicht, dessen Sätze nur aus einem einzigen einsilbigen Wort bestehen.

Selbst das 0-Niveau kennzeichnet keineswegs die größtmögliche Schwierigkeit, sondern eher eine noch konventionellere Grenze als Index 120.

19 Schreiben F. Brunot u. Ch. Bruneau nicht, wenn sie die zwei wesentlichen Grundlagen einer Sprache, das Wort und den Satz untersuchen: „Die sprachlichen Fakten, die lebendige Tatsachen darstellen, widersetzen sich jedem logischen Klassifizierungsversuch.“ (*Précis de grammaire historique*, Paris, Massou, 1937, S. 243).

F. Brunot schreibt andererseits: „Es gibt keine strenge Trennungslinie zwischen einem Einzelsatz und einem Satzgefüge, dies um so weniger, wenn man mehr Wert auf den Inhalt als auf die Form legt.“ (*La pensée et la langue*, Paris, Massou 1936, S. 28.)

20 Dgl. S. 32.

21 Flesch ist der Ansicht, daß dieser Index den ersten Texten der Lesebücher für die Kinder der Volksschule entsprechen sollte. In seinem Buch jedoch schlägt der Autor kein Beispiel für ein solches Niveau vor.

Beispiel 1

Auszug aus L. Jeunehomme und G. Collette, *Mon livre de français*, 2. Schuljahr, Lüttich, Desoer, 1950, Seite 104.

Ich besitze eine eigene Uhr. / Sie ist hier in meiner Tasche. / Ich kann sie heraus-holen, um die Uhrzeit zu sehen. / „Wieviel Uhr ist es, Du, meine liebe Uhr?“

— Es ist 12 Uhr, mein lieber Rémi. / Ich rate Dir, Dich zu beeilen, denn Du bist ein bißchen spät dran.

— Oh, 12 Uhr, das ist nett von Dir, mich daran zu erinnern. / Ohne Dich würde ich die Zeit verpassen.“

Ich bin stolz, eine Uhr in meiner Tasche zu haben, eine richtige Uhr, die tick-tack macht. / (Nach Hector Malot).

*Leichtigkeitsindex im
Französischen: ± 82*

*Leichtigkeitsindex dieser
deutschen Übersetzung: ± 75*

Beispiel 2

Auszug aus Saint-Exupéry, *Terre des Hommes*, Paris, N.R.F., 1942, S. 9–10 (deutsche Übersetzung von Henrick Becker: *Wind, Sand und Sterne*).

„Wenn ich mit solchen Gedanken vor andere Menschen trete, ersteht mein erster Nachtflug in Argentinien vor mir, das Bild einer dunklen Flugnacht, in der nur die weit verstreuten Lichter in der Ebene gleich fernen Sternen leuchteten. / Jedes von ihnen meldete in diesem Weltmeer von Finsternis das Wunder eines Bewußtseins. / Dort, in diesem Heim, da las ein Mensch, da dachte ein Mensch, da gingen vertraute Mitteilungen von Mund zu Ohr. / Dort, in jenem anderen, da mühte sich ein Forschergeist, in den Weltenraum zu dringen, und zerbrach sich den Kopf mit Berechnungen über den Andromedanebel. / Dort drüben hatten sich zwei Menschen lieb. / In weiten Zwischenräumen leuchteten die Feuer im Lande und forderten ihre Nahrung. / Selbst die kleinsten und bescheidensten riefen, beim Dichter, beim Lehrer, beim Zimmermann. / Aber in dem Raum zwischen diesen lebenden Sternen, wie viele verschlossene Fenster gab es da, erloschene Sterne, schlafende Menschen!...“

*Leichtigkeitsindex im
Französischen: ± 33*

*Leichtigkeitsindex der
deutschen Übersetzung: ± 42*

Beispiel 3

Auszug aus M. Proust, *Du côté de chez Swann*, I, Paris, Pléiade, S. 59 (deutsche Übersetzung von Eva Rechel-Mertens: *In Swanns Welt*).

„Wie liebte ich diese Kirche, und wie deutlich sehe ich sie noch vor mir! / Das alte, schwarze, pockennarbige Portal, durch das wir eintraten, war an den Seiten ausgebuchtet und abgewetzt (genau wie ein Weihwasserbecken, an das sie uns führte), so, als ob das leichte Tücherstreifen der Bäuerinnen beim Eintreten und die Berührung ihrer scheuen Finger, wenn sie Weihwasser entnahmen, in ihrer Wiederholung durch die Jahrhunderte hindurch eine zerstörerische Kraft bekommen, die Steine zur Seite gedrängt und sie mit Furchen durchzogen hätten, wie sie die Räder der Bauern-

wagen in den Randstein schleifen, gegen den sie alle Tage im Vorbeifahren stoßen. / Die Grabsteine in der Kirche, unter denen der edle Staub der Äbte von Combray, die dort begraben waren, im Chor eine Art von geistlicher Untermauerung schuf, waren selbst schon keine leblose, harte Materie mehr, denn die Zeit hatte sie mürbe gemacht und eine honigartige Substanz aus ihren Quadern gezogen, die sie mit blondem Schimmer umfloß und in ihr unmerklich leichtes Strömen hier eine rankengeschmückte gotische Letter und dort den weißlichen Veilchenton des Marmors hineingezogen und halb ausgelöscht hatte; darüber hinaus waren sie an anderen Stellen gleichsam geschrumpft, hatten die abgekürzte lateinische Inschrift noch mehr zusammengezogen und die Anordnung der nur andeutenden Ziffern noch launenhafter gestaltet, während andere durch einen unangemessenen Zwischenraum getrennt wurden. / Die Fenster der Kirche waren nie farbenprächtiger als an Tagen, da die Sonne nur wenig schien, so daß man, wenn es draußen bedeckt war, sicher sein konnte, in der Kirche werde es wunderschön sein;“ .../

*Leichtigkeitsindex im
Französischen: negativ – 9*

*Leichtigkeitsindex der deutschen
Übersetzung: negativ – 17,5*

Schlußfolgerungen

1. Die durchgeführten Sondierungen scheinen zu bestätigen, daß auch für den Franzosen und den Deutschen der Leichtigkeitsindex eine objektive Graduierung der Texte erlaubt in dem Sinne, wie wir ihn weiter oben ausgeführt haben.
2. Die amerikanische Eichung hingegen scheint weder auf die französische noch auf die deutsche Sprache anwendbar zu sein.
 - a) Eine neue Schwierigkeitsskala muß erarbeitet werden. Dabei wird man berücksichtigen müssen, daß die Silbenzahl von hundert Wörtern im Französischen normalerweise höher und im Deutschen viel höher ist als im Englischen. Scheinbar hat Flesch den prozentualen Anteil der Silben stärker gewichtet als die durchschnittliche Satzlänge, was im Französischen und im Deutschen eine noch stärkere Verzerrung verursacht.
 - b) Die von Flesch gemachten Angaben hinsichtlich der einzelnen Schulstufen lassen sich nicht auf französisch- oder deutschsprachige europäische Länder anwenden.
 - c) Wir sind der Meinung, daß Untersuchungen bei Erwachsenen über das Verständnis von Presstexten, Veröffentlichungen, amtlichen Verfügungen usw. im Rahmen dieser Forschungsarbeit nicht nur notwendig sind, sondern daß sie auch den Verantwortlichen in Kultur und Wirtschaft die größten Dienste erweisen würden.
3. Es stehen uns noch keine ausreichenden Elemente zur Verfügung, um die Signifikanz des menschlichen Interessenindex deutlich zu machen.

Man rechnet meist mit Erfolg auf die egozentrischen Einstellungen der Einzelpersonen. Darüber hinaus wäre es jedoch erforderlich festzustellen, inwieweit die Erhöhung des Index der menschlichen Interessen einer größeren Motivation bei dem Probanden entspricht^{21a}.

- a) Erfahrungen mit Kontrollgruppen werden zeigen, wie Kinder und Erwachsene reagieren würden.
- b) Man kann die Hypothese aufstellen, daß die Auswirkung der infrage stehenden Motivationstechnik je nach Kultur differiert.

2. Die elektronische Datenverarbeitung

In der erziehungswissenschaftlichen Forschung findet in zunehmendem Maße die elektronische Datenverarbeitung ihre Anwendung. Sie erleichtert nicht nur die Arbeit und gewährleistet einen hohen Präzisionsgrad, sondern sie eröffnet dem Forscher weit bessere Wirkungsfelder und -möglichkeiten. Der Studierende soll sich so frühzeitig wie möglich mit den in diesem Kapitel beschriebenen Techniken und apparativen Einrichtungen vertraut machen: man kann sie heutzutage als grundlegende Instrumente²² betrachten.

I. Die Lochkarten und ihre Behandlung

A. Beschreibung

Es handelt sich hierbei um genormte Karten, die in Form von Lochungen Auskünfte tragen, die für die Mechanographie unmittelbar brauchbar sind.

Die Hauptmerkmale dieser schematisierten Karte (siehe Seite 236) sind folgende:

1. Sie besteht aus 80 Spalten, die von 1 bis 80 numeriert sind;
2. Jede Spalte umfaßt 12 Lochungspositionen, d. h. insgesamt 960 Positionen pro Karte;

^{21a} Auf experimentellem Wege hat der Verfasser die Grenze der Validität des Index der menschlichen Interessen nachgewiesen.

Vgl. G. de Landsheere: Recherche sur l'évaluation objective de la lisibilité des manuels scolaires et des tests. XI^e Colloque international de pédagogie expérimentale 1964.

²² Wir danken der Firma IBM, die uns eine umfangreiche Dokumentation zur Verfügung gestellt hat und es uns erlaubt hat, ihre Veröffentlichungen zu benutzen. Wir haben die IBM als Beispiel gewählt, weil uns ihr Material vertrauter ist als das anderer Firmen. Die Maschinenfabrik Bull hat uns auch mit Informationsmaterial unterstützt.

3. Für jede Spalte sind die Lochungspositionen von oben nach unten mit folgenden Nummern versehen: 12, 11, 0, 1, 2, 3, 4, 5, 6, 7, 8 und 9;
4. Die Position von 0 bis 9 entsprechen den auf der Karte aufgedruckten Ziffern.

Die Position 12 und 11 sind nicht vorgedruckt.

Jede dieser 80 Spalten kann der Registrierung einer Ziffer, eines Buchstaben oder eines anderen Symbols dienen.

1. *Darstellung einer Ziffer.* Es genügt eine einzige Lochung in einer der Positionen von 0 bis 9.
2. *Darstellung eines Buchstabens.* Da eine Spalte nur 12 Positionen umfaßt, ist es nicht möglich, alle 26 alphabetischen Buchstaben durch eine einzige Lochung darzustellen. In der Praxis nimmt man zwei Lochungen vor und hält sich dabei stets an die gleiche Kodierung:

	1	2	3	4	5	6	7	8	9
12	A	B	C	D	E	F	G	H	
11	J	K	L	M	N	O	P	Q	R
0	S	T	U	V	W	X	Y	Z	

Zur Registrierung eines A in einer beliebigen Spalte locht man daher die 12 und die 1. In der Praxis erfolgt dieser doppelte Vorgang bei den alphabetischen Tastaturlochern in einem Arbeitsgang.

3. *Darstellung anderer Zeichen.* Beispiel: Das Zeichen % wird durch drei Lochungen in der gleichen Spalte bei 0, 4 und 8 übertragen.

B. Lochungsmaschinen

Man unterscheidet die einfachen Lochungsmaschinen — mit numerischer oder alphabetisch-numerischer Tastatur (zum Lochen von Buchstaben und Zahlen) — und die Kartenlocher mit Lochschriftübersetzer, wobei auf dem oberen Rand der Karte in Klarschrift der gesammelte Text wiederholt wird. In der pädagogischen Forschung gibt man im allgemeinen der letzteren den Vorzug.

C. Prüflocher

Um die Genauigkeit der Lochungen zu verifizieren, nimmt ein Locher die Originaldokumente erneut zur Hand, und nach dem Einführen der Kon-

— Geschlecht: Spalte 7: Männlich = Position 1;
Weiblich = Position 2. (7) ()

— Geburtsdatum: (8) ()
Monat: 2 Spalten (9) ()
Jahr: 2 Spalten (die beiden letzten Ziffern der Jahreszahl) (10) ()
(11) ()

Fragen

1. Warum sind Sie in die Pädagogische Hochschule eingetreten — (12) ()
Weil Sie es wollten? ja (1) — nein ()
Auf Anraten Ihrer Eltern? ja (2) — nein ()
Auf Anraten eines anderen Familienmitglieds? ja (3) — nein ()
Auf Anraten eines Freundes? ja (4) — nein ()
Auf Anraten einer Berufsberatungsstelle? ja (5) — nein ()
Aus einem anderen Grund? ja (6) — nein ()
(bitte Grund angeben) a

2. usw.

a) Wenn diese offene Antwort Gegenstand einer getrennten Untersuchung sein soll, erlaubt das Aussortieren der Karten, die mit (6) in der Spalte 12 markiert sind, eine unmittelbare Identifizierung der Fragebogen, auf denen sich eine solche Antwort befindet.

Von 80 verfügbaren Spalten wurden nur 12 verwendet, um alle obigen Informationen festzuhalten. Man kann sich leicht eine Vorstellung davon machen, wieviele Informationen eine einzige Karte zu registrieren in der Lage ist. Notfalls lassen sich auch mehrere Karten für den gleichen Gegenstand verwenden.

2. Kodierung einer Schulakte

Die Schüler der Oberstufe eines Gymnasiums wurden Individual- und Gruppentests unterzogen. Die Ergebnisse müssen in Klarschrift auf den Lochkarten erscheinen.

Nummer
der
Spalte Position
in der
Spalte

Name: maximal 15 Buchstaben (1–15)
Spalten: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Anfangsbuchstaben der ersten beiden Vornamen (16–17) ()
Geschlecht: männlich — weiblich (18) ()
Geburtsdatum: Monat (19–20) ()
Jahr (die 2 letzten Ziffern der Jahreszahl) (21–22) ()
(23) ()
Schuljahrgang: (24–26) ()
Unterrichtszweig: IG — LN — LM — NSNA — NSW (27–28) ()
Testdatum: Monat (29–30) ()
Jahr (die 2 letzten Ziffern der Jahreszahl) ()
Testergebnisse
— Primary Mental Abilities (PMA) von Thurstone (31) ()
Im Prozentrang: Wortverständnis (V) (32) ()
Räumlicher Faktor (R) (33) ()
Numerischer Faktor (N) (34) ()
Sprachflüssigkeit (W) (35) ()
Gesamtpunktwert (T) (36–38) ()
Intelligenzquotient (IQ) ()
— Skala von Wechsler (39–41) ()
Intelligenzquotient, Verbalteil (VIQ) ()
Intelligenzquotient, Handlungsteil (HIQ) (42–44) ()
Intelligenzquotient: Gesamt (GIQ) (45–47) ()
— Fragebogen zur Messung der Anpassung von Bell (Rohpunktwerte) (48–49) ()
Familiäre Anpassung (FA) ()
Anpassung nach gesundheitlichen Gesichtspunkten (GA) (50–51) ()
Soziale Anpassung (SA) (52–53) ()
Emotionale Anpassung (EA) (54–55) ()

– Prüfung der beruflichen Interessen von Derivière
Neun Niveaus. Das höchste Niveau ist 1, das niedrigste ist 9. Man kann so das Interessenprofil auf die Lochkarte übertragen.

	Nummer der Spalte	Position in der Spalte
Sprachen (A)	(56)	(())
Naturwissenschaften (B)	(57)	(())
Aktive Berufe (C)	(58)	(())
Mathem. Wissenschaft (D)	(59)	(())
Soziale Berufe (E)	(60)	(())
Technische Berufe (F)	(61)	(())
Graphik (G)	(62)	(())
Verwaltung (H)	(63)	(())
Handel (I)	(64)	(())

E. Kartendoppler

Zwei Funktionen, die der Kartendoppler erfüllt, sind für die Pädagogische Forschung von unmittelbarem Interesse:

1. Reproduktion von einer bereits gelochten Karte

- Der Kartendoppler liefert beliebig viele Kopien von einer einmal gelochten Karte (Hauptkarte).
- Zur Verkürzung der späteren Lochungsarbeit kann man für einen Schütler jeweils einen Kartenvorrat einrichten mit einer Hauptkarte als Ausgangspunkt, die die wichtigsten allgemeinen Angaben – wie Name, Anfangsbuchstaben der Vornamen, Geburtsdatum usw. – enthält.

2. Das „Mark-Sensing“

Elektronische Lochung oder elektronische Lesen und Transformation in Lochung von Graphit-Bleistiftstrichen, die an festgelegten Stellen einer speziellen Strichlochkarte eingezeichnet wurden.

- In manchen Fällen verzichtet man aus Gründen der Zeitersparnis auf die Kodifizierung der Antworten auf dem Fragebogen, und man kodifiziert unmittelbar auf einer Karte, indem man mit einem Bleistiftstrich die zu lochenden Stellen markiert.
- Dieser Vorgang ermöglicht es auch denjenigen Schulen bzw. Untersuchern, für die solche Einrichtungen nicht direkt zugänglich sind, mit geringem Kostenaufwand, Lochkarten zu erhalten.

J 1712912505091 408322161689

TESTERGEBNISSE

NAME	Geburtsdatum		Schuljahr		In-Zweit		Testdatum		PMA		Wechsler		Bell		Derivière	
	M	J	M	J	M	J	M	J	M	J	M	J	M	J	M	J
0000	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
1234	57	8	10	12	13	14	15	16	17	18	19	20	21	22	23	24
1111	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
2222	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
1234	57	8	10	12	13	14	15	16	17	18	19	20	21	22	23	24
3333	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
4444	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44
1234	57	8	10	12	13	14	15	16	17	18	19	20	21	22	23	24
5555	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
6666	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66
1234	57	8	10	12	13	14	15	16	17	18	19	20	21	22	23	24
7777	77	77	77	77	77	77	77	77	77	77	77	77	77	77	77	77
8888	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88
1234	57	8	10	12	13	14	15	16	17	18	19	20	21	22	23	24
9999	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99
0000	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

IBM Belgium S. A. 8035
Copyright IBM

Diese Karte enthält die vorgedruckten Anmerkungen hinsichtlich der soeben beschriebenen Schulakte sowie der Ergebnisse eines Schülers.

c) Karten, die also mit entsprechendem Vordruck ausgestattet sind, erlauben es dem Befrager oder Beobachter, die während eines strukturierten oder halbstrukturierten Interviews gegebenen Antworten direkt (verschlüsselt) festzuhalten bzw. die beobachteten Verhaltensweisen z. B. mit Hilfe einer Kontrollliste zu notieren.

F. Sortier- und Zählmaschinen

Die bisher beschriebenen Vorgänge dienen der Registrierung von Informationen. Die folgende Stufe betrifft ihre quantitative Erschließung.

Man will eine statistische Erhebung über den anonymen Fragebogen anstellen, dessen Kodierung wir bereits brachten (S. 237). Man möchte insbesondere wissen, wieviel Schüler jedes Unterrichtszweigs geantwortet haben. Dies läuft darauf hinaus, die Schülerzahl festzustellen, die jeweils unter den Rubriken 1, 2, 3, 4, 5, 6 der Spalte 6 rangieren.

Es genügt, die Karten in den Zufuhrkasten der Sortieranlage zu bringen und den Ablesespieß zu dem Punkt zu führen, der der Spalte 6 entspricht. Eine Minute genügt, um mehrere hundert Karten zu sortieren²³, d. h. damit alle mit I markierten Karten in Spalte 6 in das Fach I hineinlaufen usw. Die ungelochten Karten gehen dabei in ein Fach „Ausschuß“. Eine automatische Vorrichtung bringt die Maschine zum Stillstand, sobald ein Fach voll ist. Schließlich zeigt ein Kartenzähler die Anzahl der behandelten Karten an.

Der Sortierer, den wir soeben betrachtet haben, erstreckte sich nur auf eine Spalte. Dies trifft aber nicht immer zu.

Will man einen Kartenbestand alphabetisch nach Namen geordnet klassifizieren²⁴, muß jede Spalte zweimal sortiert werden, da ein Buchstabe in einer Spalte durch zwei kombinierte Lochungen dargestellt wird, wobei die eine 12, 11 bzw. 0 und die andere eine Zahl von 1 bis 9 ist.

Die Karten werden zunächst normalerweise von 1 bis 9 sortiert. In einem zweiten Arbeitsgang werden sie zwischen 12, 11 und 0 unterteilt. Das Fach 12 umfaßt nun die Buchstaben A bis I in alphabetischer Reihenfolge; das Fach 11 die Buchstaben J bis R und das Fach 0 die Buchstaben S bis Z.

Zur Fortsetzung der Klassifizierung wird der Kartenstoß A sodann nach dem zweiten Buchstaben der Namen aussortiert und so fort.

²³ Elektronische Sortieranlage Typ IBM 82: 650 Karten pro Minute; IBM 84: 2 000 Karten pro Minute.

²⁴ Man beschreibt hier die Langmethode. In der Praxis erlauben verschiedene Manipulationen eine Verkürzung des Arbeitsvorgangs.

G. Sortiermaschinen für statistische Zwecke

Es gibt Perfektionierungen bei der Sortiermaschine auf der Grundlage, wie wir sie soeben beschrieben haben. So z. B. die IBM-Type 108 (*Card Proving Maschine*):

1. Sortiert die Karten in numerischer oder alphabetischer Ordnung;
2. Wählt die Karten nach einem bestimmten Kriterium aus;
3. Prüft, ob die gewählte Kodierung berücksichtigt wurde;
4. Wählt die Karten aus, die aufs ganze gesehen einer Kombination von Faktoren entsprechen (Beispiel: Items eines Tests, der für ein bestimmtes Alter versagt hat);
5. Addiert die partiellen Punktwerte, die den Untersucher interessieren usw.

H. Tabulatoren

Es handelt sich hierbei im Grunde genommen um automatische Schreibmaschinen, die — ausgehend von Lochkarten — die gesammelten Informationen auf einem Blatt in Klartext wiedergeben.

Sie fertigen entweder Listen an (z. B.: alphabetisch geordnete Klassenlisten der Schüler mit den gesammelten Auskünften, die jedem Namen gegenüber vermerkt sind) oder Tabellen (z. B. Resultate der Schüler einer Schule bei einem bestimmten Test).

II. Elektronische Auswertungsmaschinen für Tests (Test Scoring Machines)

Die jährliche Auswertung einer großen Zahl von Tests stellt eine mühsame Arbeit dar. Die elektronischen Auswertungsmaschinen reduzieren die Kosten des Testens, gewährleisten eine sehr exakte Auswertung und erlauben es vor allem den Untersuchern, den größten Teil ihrer Zeit der eigentlichen psychologischen Tätigkeit zu widmen.

A. Das Antwortblatt

Die Vordrucke des Antwortblattes können verschieden aussehen; sie richten sich nach den Erfordernissen des Tests, den man anwenden möchte (5 Wahlen, 3 Wahlen, richtig oder falsch, *rating scales* usw.).

Das Modell eines Antwortblattes, wie es auf Seite 245 abgebildet ist, ist für einen Test mit 96 Items abgefaßt, davon sind 10 Items der Einübung gewidmet.

Beispiel 1

Item: Welches Wort im zweiten Satz erfüllt die gleiche Funktion wie das im ersten Satz schräg gedruckte Wort?

1. — Maria teilt den *Apfel*.
 — Mein Bruder Johann schlägt seinen Hund mit einem *Stock*
 A B C D E

Antwort: Um anzuzeigen, daß er Lösung D gewählt hat, genügt es, wenn der Schüler in das Kästchen d des Antwortblattes einen Bleistiftstrich zeichnet.

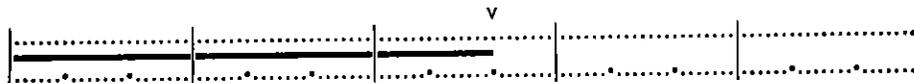
1. a b c d e

Nach Remmers, Gage und Rummel²⁵ ist dieses Antwortsystem anwendbar bei Kindern vom 9.—10. Lebensjahr ab.

Beispiel 2: Auszug aus einer rating scale

Item: Selbstbewußtsein
 Ist er sich seiner selbst sicher oder ist er unschlüssig, fehlt es ihm an Selbstvertrauen, ist er leicht aus der Fassung zu bringen?

Schüchtern zögernd	Scheint zu vorsichtig	Mäßiges Selbst- vertrauen	Selbstbewußtsein	Außerordentlich starkes Selbstbewußtsein
-----------------------	--------------------------	------------------------------	------------------	--



B. Die elektronische Auswertungsmaschine IBM — Typ 805

1. Beschreibung: Siehe Abbildung S. 246

2. Arbeitsvorgang

a) Funktionsprinzip

Man führt das Antwortblatt in die Maschine ein; ein Zeiger, der sich auf einem in 100 Grad eingeteilten Zifferblatt hin- und herbewegt, zeigt das gewünschte Resultat an (z. B. die Anzahl der richtigen Antworten). Eine genauere Prüfung des Mechanismus vermittelt ein besseres Verständnis der Nutzenanwendung, die daraus gezogen werden kann.

²⁵ Op. cit. S. 146.

IBM

Test zur Messung der Einstellung
 für EDPM Programmierer

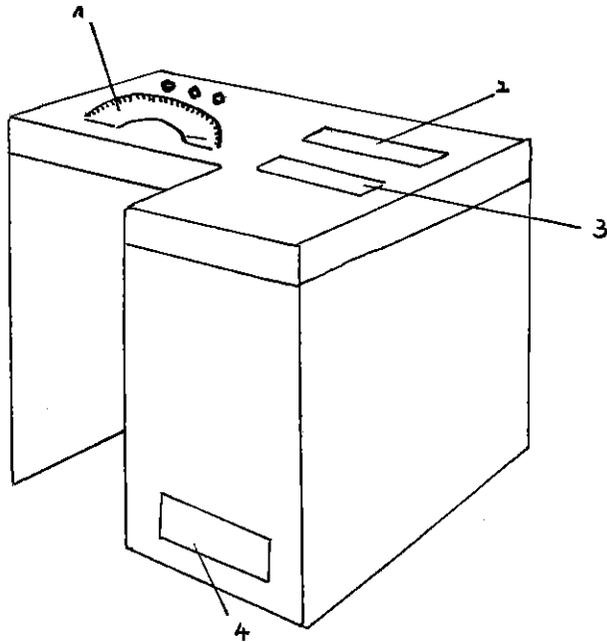
W
 R

TEIL I
 TEIL II
 TEIL III
 GESAMT

NAME
 BÜRO ODER FIRMA

GESCHLECHT
 ALTER
 DATE

PART I					PART II					PART III							
EXAMPLES					EXAMPLES					EXAMPLES							
W	a	b	c	d	e	W	1	2	3	4	5	X	a	b	c	d	e
X	a	b	c	d	e	X	1	2	3	4	5	Y	a	b	c	d	e
Y	a	b	c	d	e	Y	1	2	3	4	5	Z	a	b	c	d	e
Z	a	b	c	d	e	Z	1	2	3	4	5	1	a	b	c	d	e
21	a	b	c	d	e	15	1	2	3	4	5	2	a	b	c	d	e
22	a	b	c	d	e	16	1	2	3	4	5	3	a	b	c	d	e
23	a	b	c	d	e	17	1	2	3	4	5	4	a	b	c	d	e
24	a	b	c	d	e	18	1	2	3	4	5	5	a	b	c	d	e
25	a	b	c	d	e	19	1	2	3	4	5	6	a	b	c	d	e
26	a	b	c	d	e	20	1	2	3	4	5	7	a	b	c	d	e
1	a	b	c	d	e	21	1	2	3	4	5	8	a	b	c	d	e
2	a	b	c	d	e	22	1	2	3	4	5	9	a	b	c	d	e
3	a	b	c	d	e	23	1	2	3	4	5	10	a	b	c	d	e
4	a	b	c	d	e	24	1	2	3	4	5	11	a	b	c	d	e
5	a	b	c	d	e	25	1	2	3	4	5	12	a	b	c	d	e
6	a	b	c	d	e	26	1	2	3	4	5	13	a	b	c	d	e
7	a	b	c	d	e	27	1	2	3	4	5	14	a	b	c	d	e
8	a	b	c	d	e	28	1	2	3	4	5	15	a	b	c	d	e
9	a	b	c	d	e	29	1	2	3	4	5	16	a	b	c	d	e
10	a	b	c	d	e	30	1	2	3	4	5	17	a	b	c	d	e
11	a	b	c	d	e	31	1	2	3	4	5	18	a	b	c	d	e
12	a	b	c	d	e	32	1	2	3	4	5	19	a	b	c	d	e
13	a	b	c	d	e	33	1	2	3	4	5	20	a	b	c	d	e
14	a	b	c	d	e	34	1	2	3	4	5						
15	a	b	c	d	e	35	1	2	3	4	5						
16	a	b	c	d	e	36	1	2	3	4	5						
17	a	b	c	d	e	37	1	2	3	4	5						
18	a	b	c	d	e	38	1	2	3	4	5						
19	a	b	c	d	e	39	1	2	3	4	5						
20	a	b	c	d	e	40	1	2	3	4	5						



Vereinfachte Darstellung einer elektronischen Korrigiermaschine für Tests

1. Strommesser. Die Punktwerte werden durch einen Zeiger auf einem Zifferblatt zu 100 Grad angezeigt
2. Magazin zur Speicherung der auszuwertenden Blätter.
3. Einführungsöffnung zur Auswertung der Blätter.
4. Auffangvorrichtung: Speichert die ausgewerteten Blätter.

Ein Bleistiftstrich dient dabei als Wegweiser. Davon, ob der Bleistiftstrich mit starkem oder leichtem Druck gezogen ist und ob die Intensität des Stromes, der durch den Bleistiftstrich geleitet wird, gleichmäßig konstant gehalten wird, hängt die Genauigkeit des Ergebnisses ab.

Das Antwortblatt umfaßt maximal 750 Positionen, d. h. z. B. 150 Items zu 5 Wahlen. Die Auswertung verläuft schnell, weil alle Items gleichzeitig behandelt werden (ein Test mit 150 Fragen ist ebenso schnell ausgewertet wie ein Test mit 50, 20 oder 5 Fragen). Praktisch besteht die Funktion der Maschine darin, die guten wie die schlechten Antworten zu diagnostizieren, zu trennen und zu addieren, und zwar dies mit der Geschwindigkeit der Elektrizität.

Das in die Maschine eingeführte Blatt, das 750 Positionen, entsprechend den 750 Antwortpositionen, umfaßt, wird gegen eine Kontaktscheibe gedrückt. Jede Posi-

tion setzt sich aus 5 kleinen parallel-laufenden Schleifkontakten zusammen, die wechselweise mit dem positiven bzw. negativen Stromkreis verbunden sind. Jedes Mal, wenn ein Bleistiftstrich gegen die Kontakte gedrückt wird, schließt sich der Stromkreis und der Strom fließt, dieser Vorgang setzt die Anzeigennadel in Bewegung.

Die Maschine unterscheidet richtig und falsch durch Einschalten eines gelochten Hauptantwortblattes (nach dem Test erstellter Korrekturschlüssel), welches die Kontaktpositionen in zwei Gruppen teilt: richtig und falsch. Durch den Einsatz des Korrekturschlüssels, der verschiedenen Regulierungen — z. B. der Stromstärke²⁶ — kann das Korrekturgerät eine große Zahl von verschiedenen Indikationen liefern; die wichtigsten sind folgende:

- Anzahl und prozentualer Anteil der richtigen Antworten;
- Anzahl und prozentualer Anteil der falschen Antworten;
- Differenz zwischen der Zahl der richtigen und der falschen Antworten;
- Anzahl der durch einen Bruchteil oder ein Vielfaches der falschen Antworten verminderten richtigen Antworten;
- Anzahl der richtigen und der falschen Antworten.

b) *Gewichtungsvorrichtung.* Die Maschine kann mit einer Gewichtungsvorrichtung ausgestattet werden (*aggregate weighting unit*), mit der es möglich ist, ausgehend von 30 Variablen mit einem Gewichtungskoeffizienten von 1 bis 20 gewichtete Mittelwerte zu errechnen. Diese Vorrichtung erlaubt insbesondere eine Synthese der gelieferten Antworten in Form einer Auswertungsskala.

c) *Graphische Zählvorrichtung der Items.* Mit dieser dem Auswertungsgerät anzupassenden Vorrichtung erhält man eine graphische Darstellung, die die Zahl der Probanden wiedergibt, die auf jede einzelne Frage richtig oder falsch geantwortet haben. Diese graphischen Darstellungen liefern die für die Test- und Fragebogenanalyse notwendigen Angaben.

d) *Elektronisches Auswertungsgerät zum Photolesen (Optical Mark Scoring Reader IBM 1230).* Diese Maschine wurde im Jahre 1963 auf den Markt gebracht; sie korrigiert 1 200 Tests pro Stunde, wobei jeder Test 200 Fragen mit 5-Wahlantworten umfaßt.

²⁶ Beispiel: Soll ein Test mit 77 Items im Prozentsatz korrigiert werden, wird die Intensität derart reguliert, daß ein vollständig richtiges Antwortblatt den Zeiger des Strommessers auf die Position 100 ausschlagen läßt.

Dieses neue Korrekturgerät bringt zwei bemerkenswerte Fortschritte:

1. Für die Antwort ist die Verwendung eines Spezialstiftes nicht mehr notwendig. Es wird ein gewöhnlicher Stift Nr. 2 empfohlen: hiermit lassen sich schwarze Markierungen vornehmen, und der Schüler kann radieren soviel er will²⁷.
2. Indem man mit einem Kabel die Lochmaschine mit dem Auswertungsgerät verbindet, werden alle Gesamt- und Teilpunktwerte sowie die Antwort auf jedes Item automatisch auf Lochkarten übertragen (1200 Karten pro Stunde).

Diese Maschine ist aufgrund ihres einfachen und schnellen Systems nicht nur ein erstklassiges technisches Hilfsmittel für die Auswertung standardisierter Tests, sondern auch für die Korrektur von Routine-Prüfungen und -befragungen.

III. Elektronische Rechenanlagen

Im Dezember 1951 wurde erstmalig ein Computer in der psychologischen Forschung verwendet²⁸, und das erste Programm dieses Computers wurde von R. B. Cattell erarbeitet, der es speziell für psychologische und pädagogische Zwecke abfaßte²⁹.

Die elektronischen Rechenanlagen erlauben es, Untersuchungen von stets wachsendem Umfang in Angriff zu nehmen. Ende 1960 konnte der Computer LARC von Remington Rand 1 Million Multiplikationen in 8 Sekunden vornehmen, während eine Arbeitskraft für die gleiche Arbeit mit einer Büromaschine 5 Jahre gebraucht hätte.

Bekanntlich haben die Vereinigten Staaten im Jahre 1960 eine erste Bestandsaufnahme der Leistungsfähigkeit bei allen 15jährigen Schülern ihres Landes vorgenommen (*Project Talent*). 25 Tests, mehrere Fragebogen und

27 Wir bemerken hierzu, daß das System des Photolesens schon seit einiger Zeit von der Maschinenfabrik BULL erschlossen ist. Bei dem BULL-System werden die Antworten auf eine kleine Standardkarte eingeschrieben. Ein Photolesegerät locht 3 600 Karten pro Stunde. Man kann an den Photoleser ein Addier-Subtrahiergerät anschließen, einen Relaischrank „M 2 G“ (Abschätzung bzw. Vorgruppierung der zu lochenden Elemente in Hinblick auf andere Indikationen) oder einen Kalkulator Gamma 3, mit dessen Hilfe die Berechnungen bezüglich der Markierungen, die man lochen will, gleichzeitig mit der eigentlichen Lochung vorgenommen werden.

28 C. F. Wrigley u. J. O. Neuhaus, A Re-Factorization of the Burt-Pearson Matrix with the ORDVAC Computer (*British Journal of Psychology*, Juni 1952).

29 R. B. Cattell, *Factor Analysis: an Introduction and Manual for the Psychologist and the social Scientist*, New York, Harper & Br., 1952, 462 Seiten.

andere Untersuchungen wurden an einer stratifizierten Stichprobe durchgeführt, welche etwa 500 000 Schüler aus 1 357 verschiedenen Schulen umfaßte. Das Forschungszentrum der Universität von Iowa wurde mit der Auswertung der Prüfungen betraut: 2,25 Millionen Formulare; ungefähr 1 Milliarde Items. Die Auswertung erfolgte mit Hilfe des zuerst beschriebenen elektronischen Auswertungsgeräts. Die Analyse der Resultate wurde durch einen Computer vorgenommen, der — sobald er einmal im Besitz der vollständigen Daten war — innerhalb von 100 Stunden Schlußfolgerungen erstellte, die andernfalls Millionen normaler Arbeitsstunden erfordert hätten³⁰.

Anfangs traf man eine ziemlich klare Unterscheidung zwischen den elektronischen Rechenmaschinen und den Maschinen zur Behandlung von Lochkarten. Heute trennt man diese beiden Typen nicht mehr so streng; dies haben die vorangegangenen Anmerkungen bereits aufgezeigt. Praktisch können alle in der pädagogischen Forschung notwendigen Berechnungen (Berechnung des Chi-Quadrates, der Faktorenanalyse usw.) von den elektronischen Maschinen ausgeführt werden.

Es liegt jedoch auf der Hand, daß — abgesehen von Projekten, die einen größeren Umfang aufweisen — die Benutzung großer Rechenanlagen unrentabel ist. Fattu ist der Meinung, daß die Benutzung einer elektronischen Rechenanlage erst in dem Augenblick angezeigt ist, wo fünf Bürorechenmaschinen pausenlos beschäftigt sind.

In den Universitäten, die über ein Rechenzentrum verfügen, ist es selbstverständlich nicht notwendig, ein solches Arbeitsvolumen zu erreichen, um von den Vorzügen der modernen Technik zu profitieren, vor allem, wenn die Maschinenprogramme bereits vorliegen.

Die Pädagogen sollen in erster Linie zur Kenntnis nehmen, welche Dienste ihnen die interfakultären Rechenzentren leisten können. Die Spezialisten dieser Zentren werden ihnen — zumindest im Anfang — helfen, ihre Probleme logisch bzw. in mathematischen Termini zu definieren, die arithmetische Übertragung durchzuführen sowie die Programmierung der Rechenanlage, die Registrierung der Angaben auf Karten oder Magnetbändern zu bewerkstelligen und schließlich Anleitung zu selbständigen Berechnungen geben.

30 Das zentrale Ferrit-Speichergerät des Computers IBM 7070 umfaßt 100 000 Informationen, die in einer 6 Millionstel Sekunde abgerufen werden können; ihre Kartenleser arbeiten mit einer Geschwindigkeit von 500 Karten pro Minute; ihre Locher stellen 250 Karten pro Minute her. Ihr stehen 5 Bandspeicher zur Verfügung, von denen jeder 28 000 000 bis 56 000 000 Zeichen aufnimmt; und sie kann die elektronische Schreibmaschine eines Computers 1401 benutzen, die 600 Zeichen in der Minute schreibt.

Es gelte als eine Regel, die mancher Anfänger in seinem eigenen Interesse berücksichtigen sollte: Diese Zusammenarbeit muß zustande kommen, *bevor* die eigentlichen Forschungsarbeiten beginnen.

Die Mehrzahl der Universitäten sehen neuerdings Programmierkurse vor, bei denen den Forschern die notwendige Einführung vermittelt wird.

3. Die Statistik

I. Die Versuchspläne

Wir haben gesehen, daß das Experiment eine provozierte Beobachtung ist. Es hat zum Ziel, besondere Bedingungen im Hinblick darauf zu schaffen, leichter und systematischer eine oder mehrere Variablen bzw. ihre Interaktionen zu untersuchen.

Lange Zeit hat die Statistik — diese junge Wissenschaft — dem Forscher kaum Bewegungsfreiheit gelassen; er sah sich gezwungen, jeweils nur einen Faktor variieren zu lassen. Er war daher auf eine Reihe kleiner gesonderter Experimente angewiesen. Die Addition, die Synthese, die Integration der Resultate zielten darauf hin, ein Bild von der Wirklichkeit zu vermitteln.

Die Isolierung einer einzigen Variablen war in vielen Fällen illusorisch. Herausgelöst aus ihrem Zusammenhang, verlor sie an Form und Bedeutung. Darum ist die statistische Methode — obgleich streng im mathematisch-naturwissenschaftlichen Bereich — in der empirischen Pädagogik und darüber hinaus im Bereich der Humanwissenschaften im allgemeinen enttäuschend gewesen.

Heute ermöglicht die Statistik die gleichzeitige Bearbeitung von mehreren Faktoren und ihren Kombinationen.

Versuchspläne wurden im Verlauf verschiedener Untersuchungen entwickelt. Wir fassen hier die wichtigsten zusammen, um ihre Möglichkeiten und Grenzen aufzuzeigen. Der Forscher muß sie kennen, um die für seine Ziele geeignetsten auszuwählen, und um die für seine speziellen Arbeitsbedingungen erforderlichen Abänderungen vornehmen zu können.

Eine Warnung ist jedoch geboten: Die Pädagogik setzt sich mit einer komplexen und fließenden menschlichen Materie auseinander.

Das Experiment nimmt häufig eine gewisse Zeit in Anspruch. Man muß die eventuelle Intervention bestimmter Faktoren, die dazu angetan sind, den eigentlichen Schulerfolg zu beeinflussen, maximal reduzieren.

Bei einer Untersuchung, die sich z. B. auf die Wirksamkeit von zwei Lehrmethoden bezieht, bleibt die wesentliche Frage diese: „Kann die Differenz in den Schulergebnissen dem Einfluß der Lehrmethoden zugeschrieben werden oder gibt sie einfach Zufallsvariationen wieder aufgrund von externen oder internen Faktoren?“

Die wichtigsten *internen* Fehlerquellen sind folgende:

- Die Variationen, die bei einem Schüler im Verlauf eines Experiments intervenieren: Motivation, Ermüdung, persönliche Probleme...
- Die Differenzen, die zwischen den am Experiment beteiligten Schülern bestehen:
 1. Begabungsdifferenzen;
 2. Differenzen des anfänglichen Kenntnisstandes;
 3. Differenzen der Persönlichkeit: Ausdauer, Motivation, Aufnahme-niveau, Interesse usw.;
 4. Differenzen in der physischen Konstitution;
 5. Differenzen der Lernmethoden;
 6. Differenzen des sozio-ökonomischen und kulturellen Niveaus der Familien usw.

Die wichtigsten *externen* Fehlerquellen sind folgende:

- Die Differenz unter den Lehrern: Persönlichkeit, pädagogische Qualifikation, bewußte oder unbewußte Bevorzugung einer der Untersuchungsmethoden usw.
- Die Differenz innerhalb der Arbeitsbedingungen: Räumlichkeiten, Einrichtung, Heiz- und Beleuchtungsanlage, Klimaanlage, didaktisches Material, Schulorganisation usw.

Der Untersucher bekämpft diese Fehler auf zweierlei Arten:

1. Er versucht, ihnen vorzubeugen oder durch ein Kompensationsspiel ihre Effekte zu neutralisieren:
 - a) indem zunächst möglichst gleichartige Gruppen aufgestellt werden, um die internen Fehler auszuschalten;
 - b) indem diese Gruppen sodann in Untergruppen unterteilt werden, und zwar nach dem Zufallsprinzip, derart, daß die externen Behandlungen und die Fehler, die sie hervorrufen können, keinen einseitigen Einfluß ausüben.
2. Er verfügt über statistische Techniken, die es ihm erlauben, zu bestimmen, inwieweit es sich bei den festgestellten Differenzen um Zufallserscheinungen handelt, d. h. um eine zufällige Kombination der Fehler (siehe weiter unten: Null-Hypothese).

Wir wollen zunächst die Versuchspläne behandeln, die der Vorbeugung von Fehlern dienen³¹. Anschließend werden die statistischen Kontrolltechniken dargelegt.

A. Versuchspläne mit Variation eines Einzelfaktors

1. Ausschaltung interner Fehler (eine einzige Gruppe)

a) Methode der Kontrollgruppe

Vortest
Experiment
Nachtest

Es soll der Einfluß einer bestimmten Anzahl von Abschreibübungen auf die Orthographie einer Klasse gemessen werden. Bei diesem Schema umfaßt das Experiment drei Entwicklungsstufen:

1. Vorausgegangene Durchführung eines Tests zur Messung der orthographischen Kenntnisse;
2. Untersuchung;
3. Durchführung eines orthographischen Tests — Parallelform zum Vortest, — und Feststellung der Leistungssteigerung.

Diese noch recht gebräuchliche Verfahrensweise führt zu wenig signifikanten Ergebnissen. Durch das Fehlen einer der Experimentalgruppe weitgehend äquivalenten Kontrollgruppe haben die internen Fehler freies Spiel. Der Vortest selbst kann eine Leistungssteigerung bewirken, deren Einfluß nicht abzuschätzen ist.

Dieses einfache Schema sollte daher vermieden werden. Es kann allenfalls bei Kurzuntersuchungen, die sich auf wenige Einzelpersonen erstrecken, gewisse Dienste leisten. Aber selbst in diesem Fall bestehen viele Vorbehalte, und die Resultate sind naturgemäß nicht zu verallgemeinern.

b) *Methode der Parallelgruppen.* Bei der Aufstellung von Parallelgruppen kann man das vorausgegangene Experiment weitgehend präzisieren. Der folgende Plan erlaubt außerdem, die durch den Vortest erzielte Leistungssteigerung festzustellen.

³¹ Zur eingehenden Untersuchung siehe: B. Winer, *Statistical Principles in Experimental Design*, London, McGraw-Hill, 1962.

Gruppe A:	Vortest	Experiment	Kontrolltest
Gruppe B:	Vortest	—	Kontrolltest
Gruppe C:	—	Experiment	Kontrolltest
Gruppe D:	—	—	Kontrolltest

Häufig organisiert man zur besseren Erkenntnis der Effekte der experimentellen Variablen mehrere konkurrierende Untersuchungen, bei denen diese Variable in verschiedenen Graden interveniert. In einem solchen Fall muß jede Teiluntersuchung den Kontrollen A, B, C und D unterworfen werden.

Beispiel:

Man möchte auf einem bestimmten Gebiet den Lernerfolg mit Hilfe von Lehrmaschinen untersuchen. Es werden drei Arbeitsmethoden miteinander verglichen:

- a) Herkömmliche Unterrichtsmethode;
 - b) Die Hälfte der Zeit wird der herkömmlichen Unterrichtsmethode gewidmet; in der anderen Hälfte verwendet man Lehrmaschinen;
 - c) Ausschließliche Benutzung von Lehrmaschinen.
- Um die durch den oben aufgezeigten Plan gebotenen Garantien zu ermöglichen, sind 8 Parallelgruppen notwendig:

Experiment a	Gruppe 1	Vortest	Herkömmlicher Unterricht	Kontrolltest
	Gruppe 2	Vortest	—	Kontrolltest
	Gruppe 3	—	Herkömmlicher Unterricht	Kontrolltest
	Gruppe 4	—	—	Kontrolltest
Experiment b	Gruppe 5	Vortest	1/2 herk. Unterricht 1/2 masch.	Kontrolltest
	Gruppe 6	—	1/2 herk. Unterricht 1/2 masch.	Kontrolltest
Experiment c	Gruppe 7	Vortest	Lehrmaschinen	Kontrolltest
	Gruppe 8	—	Lehrmaschinen	Kontrolltest

Aufstellung von Parallelgruppen — Es gibt selbstverständlich keine völlige Übereinstimmung unter den Schülern, so daß der Begriff „Parallelgruppen“ sich nur auf eine durchschnittliche Äquivalenz bezieht. Da der Wert des Experiments insbesondere von der Gruppenübereinstimmung abhängt, läßt man bei der Aufstellung der Gruppen größte Sorgfalt walten. Wir legen im folgenden die drei klassischen Methoden dar:

1. *Nach dem Zufallsprinzip aufgebaute Stichproben.* Wenn man bei einer homogenen Population zwei oder mehrere Gruppen nach dem Zufallsprinzip

zip wählt (durch Lösen oder Verwendung von Zufallstabern), tendieren diese zu einer äquivalenten Zusammensetzung.

Anmerkungen:

- a) Die Gruppen setzen sich vielfach aus vollständigen Klassen zusammen. Insofern, als sie nicht nach besonderen Kriterien homogenisiert worden sind (Gruppierung nach Begabungsniveau beispielsweise), kann man sie als nach dem Zufallsprinzip aufgebaute Stichproben bezeichnen.
- b) Zwei Klassen desselben Schuljahrganges brauchen hingegen nicht unbedingt den gleichen Fortschritt im Lehrprogramm gemacht zu haben. Sofern der Untersuchung ein Vortest vorausgeht, erlaubt die Kovarianzanalyse einen Vergleich der Resultate des Experiments, so als wenn es sich anfänglich um streng parallele Gruppen gehandelt hätte.
- c) Zur besseren Vergleichbarkeit der Gruppen sollte in den Gruppen C, D, 3, 4, 6 und 8 anstelle des Vortests die Teilnahme an Tests geübt worden sein.

2. *Gruppen mit gleichem Mittelwert und gleicher Standardabweichung.* Die Aufstellung von Gruppen gleichen Durchschnittsalters, mit gleichen Intelligenzquotienten oder gleichen pädagogischen Quotienten, d. h. mit gleichen mittleren schulischen Leistungen usw. genügt noch nicht. Es muß auch die Standardabweichung berücksichtigt werden, ohne die die Mittelwerte ihre Bedeutung verlieren können (Ein Mittelwert von 10 Jahren kann auf völlig verschiedenen Alterszusammensetzungen beruhen, z. B. eine Gruppe: *nur* 8- und 12jährige, eine andere *nur* 10jährige).

3. *Gruppenaufstellung nach äquivalenten Paaren.* Die Aufstellung von äquivalenten Paaren dient dem Zweck, die Ungenauigkeit des Mittelwertes zu vermeiden.

Im Idealfall bestünde jedes Paar aus zwei absolut gleichartigen Schülern. Ihr Vergleich würde zu Ergebnissen von großer Genauigkeit führen.

Allein die Beobachtung von eineitigen Zwillingen kann sich diesen vollkommenen Bedingungen nähern. Leider ist sie schwerlich zu verwirklichen.

In der täglichen Praxis versucht man, weitgehendst gleichartige Schülerpaare zu bilden (Alter, pädagogisches Niveau, gleicher Leistungsstand, Intelligenzquotient usw.).

Unter günstigen Voraussetzungen befriedigt diese Methode am meisten. Aber sie erfordert eine lange Vorbereitungszeit, und es ist, wenn man in einer Schule arbeitet, ein Ausschluß vieler Kinder vom Experiment unumgänglich.

- c) *Training-Methode.* Wir wollen nur kurz an diese Methode erinnern, denn sie ist in seltenen Fällen anwendbar. Sie beruht darauf, daß von einem

gewissen Übungsstadium an die Leistung nicht mehr weiter ansteigt. Das gilt vor allem bei den Experimenten zur Messung des Lernvorganges.

Man beginnt mit einem ersten Training, bis zu dem Augenblick, wo die Probanden keine Fortschritte mehr machen. Sodann führt man eine zweite Lehrmethode ein und prüft, ob die erzielten Fortschritte die sich dabei zeigen, dieser zweiten Methode zuzuschreiben sind.

d) *Rotationsmethode.* Ein Lehrer unterrichtet in zwei Klassen des gleichen Schuljahres und möchte die Wirksamkeit von zwei Lehrmethoden vergleichen. Sein Experiment gründet sich auf folgendes Schema:

Klasse I	Klasse II
<i>Unterrichtsgegenstand: Flächenberechnung des Parallelogramms</i>	
– Vortest	– Vortest
– KLASSISCHE METHODE	– NEUE METHODE
– Kontrolltest	– Kontrolltest
M_1	M_2
.....	
<i>Unterrichtsgegenstand: Flächenberechnung des Trapez</i>	
– Vortest	– Vortest
– NEUE METHODE	– KLASSISCHE METHODE
– Kontrolltest	– Kontrolltest
M_3	M_4
Auswertung: man vergleicht M_1 und M_4 mit M_2 und M_3	

Die Rotation erlaubt es, den Einfluß bestimmter Faktoren auszuschalten. Wenn das durchschnittliche Begabungsniveau der Klasse I höher ist als das der Klasse II, wirkt sich die Differenz bei beiden Methoden gleichermaßen aus.

2. Ausschaltung externer Fehler

Zur Ausschaltung externer Fehler wählt man willkürlich unter den Parallelgruppen solche, die verschiedenen experimentellen Behandlungen unterzogen werden sollen (*randomized block design*).

Wir haben gesehen, daß das Problem der externen Fehler vor allem dann auftritt, wenn bei einer Untersuchung verschiedene Lehrer in verschiedenen Schulen eingeschaltet werden.

Beispiel: Man möchte die Wirksamkeit von zwei Lehrmethoden vergleichen, die sich daraus ergebenden Rückschlüsse sollen auf alle Klassen der Sexta (Gymnasium) innerhalb einer Provinz anwendbar sein. Die verschiedenen Gruppen bestehen aus ganzen Klassen, deren Zusammensetzung nicht modifiziert werden kann. Man beschließt beispielsweise, daß man in jeder Schule:

1. zwei Klassen willkürlich wählt;
2. ebenfalls willkürlich die Lehrer benennt, die nach der ersten bzw. zweiten Methode unterrichten.

Für acht Schulen erhält man das folgende Schema:

		<i>Methoden</i>	
		A	B
<i>SCH</i> <i>U</i> <i>L</i> <i>E</i> <i>N</i>	1		
	2		
	3		
	4		
	5		
	6		
	7		
	8		

Anmerkungen:

1. Dieses Schema läßt sich je nach den Erfordernissen des Experiments leicht variieren (größere Anzahl an Schulen, Lehrern, Trennung nach Geschlecht usw.).
2. Dieses Schema entbindet den Untersucher keineswegs von der Kontrolle der internen Fehler.

B. Versuchspläne, die die gleichzeitige Untersuchung von mehreren Variablen erlauben (Varianzanalyse)

Die vorausgegangenen Pläne sehen die Variation eines einzigen Faktors vor, wobei die anderen konstant bleiben oder ihr störender Einfluß neutralisiert wird.

Die Arbeiten der modernen Statistiker, und insbesondere die von Sir Ronald Fisher³², erlauben die gleichzeitige Untersuchung der Variation von mehreren Faktoren und ihren Interaktionen.

³² R. Fisher, *The design of experiments*, Edinburgh, Oliver & Boyd, 1949, 5. Aufl.
R. Fisher, *Statistical methods for research workers*, Edinburgh, Oliver & Boyd, 1950, 11. Aufl.

Dieser Fortschritt ist wesentlich für die pädagogische Forschung, denn die Mehrzahl der Erziehungssituationen resultiert aus der Abwandlung komplex auf das Kind einwirkender Kräfte.

Beispiel: Man möchte den Effekt von Lob und Tadel auf den Lernerfolg der Schüler untersuchen.

Je nachdem, ob die Kinder intro- oder extravertiert sind, reagieren sie unterschiedlich auf Lob und Tadel. Ein Persönlichkeitsfaktor stößt somit mit dem experimentellen Faktor zusammen, und es besteht die Gefahr, daß die Rückschlüsse verfälscht werden, wenn sich die Untersuchung auf zwei nach ausschließlich pädagogischen Kriterien gewählten Parallelgruppen erstreckt. In der Tat ist es notwendig, daß vier Gruppen aufgestellt werden, um die folgenden Kombinationen zu gewährleisten:

- Lob -- Extravertierte;
- Lob -- Introvertierte;
- Tadel -- Extravertierte;
- Tadel -- Introvertierte.

Das Vorhandensein von zwei experimentellen Faktoren und von zwei möglichen Interaktionen erfordert daher 2² Beobachtungen.

In Wirklichkeit sind es nicht nur diese beiden Charaktereigenschaften, denen man Rechnung tragen sollte, sondern noch eine Reihe anderer Faktoren, die auf jeden Lernvorgang einwirken. So erfordern 5 Faktoren bereits 2⁵, d. h. 32 Beobachtungen und so fort.

Da sich derart zahlreiche Beobachtungen schwerlich durchführen lassen, ist man vielfach gezwungen, sich auf zwei oder drei Faktoren zu beschränken und alle anderen hypothetisch zusammenzufassen.

Im Folgenden schlagen wir zunächst Pläne vor, bei denen alle Faktoren getrennt untersucht werden; dann folgen gemischte Schemata, und zwar mit isolierten und miteinander verbundenen Faktoren³³.

³³ Bedeutende Quellen: F. Süllwold: Zur Frage der Präzision und Ergiebigkeit von pädagogischen Experimenten, in: H. Heckel et al., *Pädagogische Forschung und pädagogische Praxis*, Heidelberg, Quelle u. Meyer, 1958, S. 185-221. E. Lindquist, *Statistical analysis in educational research*, Boston, Houghton Mifflin, 1940. E. Lindquist, *Design and analysis of experiments in psychology and education*, New York, Houghton Mifflin, 1953. Siehe auch den Artikel von E. Shacklock Evans, *The design of teaching experiments in education*, in „*Educational Research*“, Bd. V, Nr. 1, Nov. 1962, S. 37 bis 52.

1. Pläne zur Untersuchung der gleichzeitigen Variation von mehreren isolierten Faktoren

Problem 1. Inwieweit werden die Ergebnisse eines Diktats — abgesehen von den bereits vorhandenen orthographischen Kenntnissen — beeinflusst:

- A. durch den Sinnzusammenhang;
- B. durch die Motivation zu Beginn des Diktats;
- C. durch die Geschwindigkeit des Diktierens³⁴?

Experimentelle Faktoren

A. *Sinnzusammenhang.* Man bedient sich der vier folgenden Arten der Darbietung:

- A₁: Der Gesamttext wird vor Beginn des eigentlichen Diktats vorgelesen; dann diktiert man in logischen Abschnitten.
- A₂: Man unterläßt das Vorlesen des Gesamttextes zu Beginn des eigentlichen Diktats. Während des Diktats liest man jeden Satz vollständig, dann diktiert man in logischen Abschnitten.
- A₃: Kein vorheriges Lesen; man diktiert in logischen Abschnitten.
- A₄: Kein vorheriges Lesen; man diktiert in Abschnitten, die keine logische Einheit mehr darstellen, sondern die ungefähr von der gleichen Länge sind wie die stückweisen Diktate aus A₁, A₂ und A₃.

	A ₁		A ₂		A ₃		A ₄	
	B ₁	B ₂						
C ₁	5	5	5	5	5	5	5	5
C ₂	5	5	5	5	5	5	5	5
C ₃	5	5	5	5	5	5	5	5

B. Motivation

- B₁: Der Lehrer kündigt an, daß das Diktat schwer ist.
- B₂: Der Lehrer kündigt an, daß das Diktat leicht ist.

³⁴ In Anlehnung an Süßwold, op. cit., S. 204—208.

C. Geschwindigkeit

- C₁: Langsam
- C₂: Normal
- C₃: Schnell

Zahl der möglichen Kombinationen: 4 mal 2 mal 3 = 24.

Bei einer Beteiligung von 120 Schülern am Experiment wird man 24 Gruppen zu je 5 Schülern aufstellen. Dabei ergibt sich folgendes Schema:

Auswertung des Experiments

1. Untersuchung des Effekts der Faktoren A, B, C — getrennt —

— Faktor A —

A₁: Mittelwert der Resultate von 30 Schülern, die der Kondition A₁ unterzogen wurden, d. h.:

$$A_1B_1C_1 + A_1B_1C_2 + A_1B_1C_3 + A_1B_2C_1 + A_1B_2C_2 + A_1B_2C_3.$$

A₂: gleiches Verfahren.

A₃: gleiches Verfahren.

— Faktor B —

B₁: 12 Zellen enthalten diesen Faktor.

B₂: 12 Zellen enthalten diesen Faktor.

— Faktor C —

C₁: 8 Zellen

C₂: 8 Zellen

C₃: 8 Zellen.

2. Untersuchung der Interaktion der Faktoren

— Primäre Interaktionen

Möglicherweise treten die Effekte der A-Bedingung eher bei einer bestimmten Geschwindigkeit zutage. Zum Beispiel: Je schneller man diktiert, um so wichtiger ist es vermutlich, daß man logische Einschnitte heraushört usw. Man bestimmt die Interaktionen zwischen den Bedingungen A—B; A—C; B—C.

— Sekundäre Interaktionen

Es ist auch möglich, daß die Interaktion A—C von B abhängt. Beispiel: Die Differenz zwischen A₁ und A₄ ist bei großer Geschwindigkeit (C₃) besser gekennzeichnet als bei mäßiger Geschwindigkeit (C₁); außerdem tritt

diese Differenz, sofern man das Diktat als schwierig angekündigt hat (B_1), deutlich hervor. Hier wird man daher die Interaktionen $A_1-B_1-C_3$; $A_4-B_1-C_3$ usw. untersuchen. Die Varianzanalyse erlaubt eine Auswertung dieser Interaktionen.

Problem 2. Zur Untersuchung der Wechselbeziehung zwischen der Wahrnehmung geometrischer Formen und dem Kopfrechnen zu Beginn des Mathematikunterrichts in der Quinta legt G. Mialaret³⁵ das folgende Experiment vor:

Er hat zwei Klassen des gleichen Schuljahres gewählt (Jungen und Mädchen) und legt drei Übungsarten zugrunde:

- Übungen im schnellen Kopfrechnen;
- Übungen der Analyse von geometrischen Formen;
- Vorführung von geometrischen Filmen.

Jede Klasse wurde in zwei Gruppen aufgeteilt (A und B) und jede Gruppe wiederum in zwei Untergruppen (A_1-A_2 ; B_1-B_2):

Jungenklasse	{	A. Analyse geometrischer Formen	{	A_1 mit Filmvorführungen
			{	A_2 ohne Filmvorführungen
	{	B. Übungen im Kopfrechnen	{	B_1 mit Filmvorführungen
			{	B_2 ohne Filmvorführungen

Mädchenklasse: ebenso

Das Schema für das Experiment sieht demnach so aus:

	Jungen		Mädchen	
	Geometrie A	Kopfrechnen B	Geometrie a	Kopfrechnen b
mit Filmvorführungen . . .	A_1	B_1	a_1	b_1
ohne Filmvorführungen . .	A_2	B_2	a_2	b_2

³⁵ Eine vollständige Beschreibung dieses Experiments einschließlich der Varianzanalyse findet sich in: G. Mialaret, *Nouvelle pédagogie scientifique*, op. cit.

Ein schönes Forschungsbeispiel, wobei ein ähnliches Schema verwendet wird, liefert: H. Fischer, *La rapidité de calcul est-elle un critère suffisant pour l'orientation scolaire ou professionnelle?* (*Revue Belge de Psychologie et de Pédagogie*, XIX, 1957, Nr. 77).

2. Pläne zur Untersuchung von mehreren, teilweise verschiedenen Faktoren, von denen sich einige nicht kombinieren lassen

a) lateinisches Quadrat³⁶

Man möchte eine Vergleichsstudie anstellen zwischen dem Erfolg im Auswendiglernen mit und ohne Lehrmaschinen. Das Schema des lateinischen Quadrats erlaubt komplexe Vergleiche mit einer relativ niedrigen Anzahl von Schülern.

Angenommen, man möchte den Effekt der folgenden vier Lehrmethoden messen:

- A: Auswendig lernen, während einer Zeit von 10 Minuten — ohne Lehrmaschinen.
- B: Auswendig lernen, während einer Zeit von 20 Minuten — ohne Lehrmaschinen.
- C: Auswendig lernen, während einer Zeit von 10 Minuten — mit Lehrmaschinen.
- D: Auswendig lernen, während einer Zeit von 20 Minuten — mit Lehrmaschinen.

Es stehen vier Parallelklassen (1, 2, 3 und 4) zur Verfügung sowie vier Texte gleicher Länge und gleichen Schwierigkeitsgrades (I, II, III und IV).

Die Untersuchung nach diesem Schema erfolgt in mehreren Stufen:

1. Jeder Text wird unter vier Gesichtspunkten untersucht.
2. Jede Klasse führt das Experiment nach den vier Verfahrensweisen durch.
3. Die Reihenfolge der Verfahrensweisen wird für jede Klasse verschieden sein.

³⁶ Das Schema des lateinischen Quadrats wurde zunächst in der Landwirtschaft angewandt. Beispiel: Experiment mit vier Düngemitteln, die nur in ihrem Phosphorgehalt variieren. Da der zu untersuchende Boden unterschiedlich fruchtbar sein kann, teilt man das Gelände in 16 gleiche Vierecke und wendet die Düngemittel so an, daß jedes einmal in jeder Reihe und in jeder Spalte nach folgendem Schema erscheint:

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Siehe die bereits zitierten Werke von R. Fisher u. R. Travers!

Das Schema, das dem Experiment zugrunde liegt, wird folgendermaßen aussehen:

	I	II	III	IV
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

- jede Klasse beginnt mit einer anderen Verfahrensweise, was den Einflußfaktor anderer Verfahrensweisen eliminiert.
 - Die Interaktionen (Übertragbarkeit von Lernerfolgen) wirken je nach den vier möglichen Reihenfolgen der Verfahrensweisen mit.
4. Die unannehmbaren Kombinationen werden vermieden. Beispiel: die Gruppe 1, die den Text I während 10 Minuten ohne Lehrmaschine (1A I) gelernt hat, wird nicht den gleichen Text I während 20 Minuten ohne Lehrmaschinen lernen (1B I).
 5. Mehrere mögliche Interaktionen werden nicht getrennt untersucht, sie sind daher mit anderen verbunden: das lateinische Quadrat führt daher nur zu Teilauswertungen.

Auswertung des Experiments

1. Vergleich der Effekte der Verfahrensweise A, bezogen auf jede Klasse:
1A I, 2A IV, 3A III, 4A II.
Gleicher Vorgang für B, C und D.
2. Vergleich der Lernerfolge bei jeder Klasse.
3. Gesamteffekt des Lernprozesses mit und ohne Lehrmaschine usw.

b) griechisch-lateinisches Quadrat

Problemstellung: Bei dem vorausgegangenen Experiment will man außerdem noch in Erfahrung bringen, ob die Lernerfolge je nach Tageszeit differieren. Man legt vier Zeitpunkte fest:

8 Uhr = α
 10 Uhr = β
 14 Uhr = γ
 16 Uhr = δ

Es ist daher wichtig, daß jede Klasse zu jedem der vier Zeitpunkte einmal arbeitet, und daß jede Behandlungsmethode bei jeder Klasse zu verschiedenen Zeiten untersucht wird.
 Man hält sich dabei an folgendes Schema:

	I	II	III	IV
1	A_α	B_β	C_γ	D_δ
2	B_γ	C_δ	D_α	A_β
3	C_δ	D_γ	A_β	B_α
4	D_β	A_α	B_δ	C_γ

Auswertung des Versuchs: Siehe lateinisches Quadrat!

II. Die Stichproben

A. Definition

Eine Stichprobe ist eine begrenzte Anzahl von Einzelpersonen, Gegenständen oder Begebenheiten, deren Beobachtung Rückschlüsse erlaubt, die auf die Gesamtpopulation, innerhalb der die Wahl erfolgte, anwendbar sind.
 Der Wert der Rückschlüsse ist abhängig vom Repräsentationsgrad der Stichprobe: In welchem Maße – mit welchem Wahrscheinlichkeitsniveau – liefert sie ein zuverlässiges Bild von der in Betracht gezogenen Gesamtheit?
 Die Wahl der Stichproben erfordert mathematische Spezialkenntnisse. Auch ruft man bei wichtigen Untersuchungen Experten zur Hilfe, die im Hinblick auf die vom Forscher verfolgten Ziele den zu untersuchenden Populationsanteil bestimmen und die Fehler, die auf Zufallsvariationen beruhen, untersuchen.

Die folgenden Bemerkungen haben zum Ziel:

1. Vor häufig bei der Wahl von Stichproben begangenen Fehlern zu warnen.
2. Einige Grundprinzipien in Erinnerung zu bringen.

B. Fehlerquellen bei der Wahl der Stichproben

Der Laie gründet sein Urteil und seine Meinung oftmals auf allzu begrenzte Beobachtungen — die somit nicht hinreichend repräsentativ für die in Betracht stehende Gesamtheit sind — und zieht Rückschlüsse, die wertlos sind: „Alle Nordländer sind blond und haben blaue Augen usw.“

In der experimentellen Pädagogik stößt man mitunter auf ähnliche Fehlerquellen, indem nämlich die Forscher aufgrund von Beobachtungen Verallgemeinerungen aufstellen, die in Wirklichkeit nur für die jeweils untersuchten Einzelpersonen Gültigkeit haben.

Wir geben nun einige der häufig auftretenden Fehlerquellen wieder³⁷:

- a) Die Lösung des bequemen Weges: Zur Einholung einer Meinung über ein allgemein interessierendes Problem befragt ein Schulleiter z. B. vorzugsweise seine Lehrerschaft, weil er sie kennt und sie ihm leicht zugänglich ist.
- b) Die Einstellung des Befragers: Eine Gruppe von Spezialisten wird dazu neigen, sich nur an einen solchen Personenkreis zu wenden, den sie über das Problem besonders gut informiert hält.
- c) Die Verwendung von verjäherten oder selektiven Informationsquellen. Beispiel: Willkürlich Namen aus einem Telefonbuch herausziehen, was naturgemäß eine Populationsschicht mit einem bestimmten sozio-ökonomischen Niveau gänzlich ausschaltet.
- d) Die unzureichende geographische Streuung. Wenn eine Umfrage beschränkt wird auf alle Schulen des gleichen Distrikts, der gleichen Stadt usw. läßt man Differenzen außer acht, die zwischen ländlichem und städtischen Milieu bzw. zwischen Arbeiterviertel und Wohngebieten des Mittelstandes bestehen können.
- e) Die nicht-definierten Quoten. Wenn man jeden Befragter damit beauftragt, eine bestimmte Anzahl von Lehrern in einem Bezirk zu befragen — wobei sich gewisse Personentypen aufgeschlossener zeigen als andere —, laufen die einzelnen Befragter Gefahr, sich nur an solche Probanden zu wenden, die weitgehend gleichartige psychologische Merkmale aufweisen.
- f) Die Reduzierung der Stichprobe im Verlauf der Arbeit. Die Stichprobe, nach der ein Fragebogen ausgerichtet ist, kann anfänglich Validität besitzen. Aber wenn man nur eine geringe Zahl von Antworten erhält, kann die Reduzierung einer Stichprobe die Allgemeingültigkeit gänzlich in Frage stellen.

37 Nach F. G. Cornell, *Sampling Methods* (Enc. of Ed. Res., op. cit., S. 1181).

C. Stichprobenverfahren nach dem Zufallsprinzip. Repräsentative Stichprobe:

Um verschiedene Fehlerquellen auszuschalten, von denen soeben die Rede war, wählt man die Stichproben nach dem Zufallsprinzip, d. h. derart, daß jede Person innerhalb der in Betracht gezogenen Gesamtheit die gleiche Chance hat, gewählt zu werden.

Die Wahrscheinlichkeitsrechnung erlaubt, den Sicherheitsgrad zu bestimmen, den das Verhältnis zwischen der Größe der Stichprobe und der Gesamtpopulation verlangt. Je umfangreicher die Stichprobe ist, um so zuverlässiger sind die Rückschlüsse.

Typenarten von Zufallsstichproben

a) Die einfache Stichprobe

Bei einer Untersuchung, die sich auf Schüler der Sexta der Gymnasien innerhalb einer Stadt erstreckt, hat man vielleicht den Wunsch, eine Stichprobe zu wählen, und zwar unter der einen Bedingung, jedem Schüler die gleiche Chance zur Beteiligung an der Umfrage sicherzustellen. Das in diesem Fall meist angewandte Verfahren geht so vor, daß man die Schüler in alphabetischer Reihenfolge mit fortlaufender Numerierung aufführt und sodann mit Hilfe einer „Tafel mit Zufallsfolgen von Ziffern“ wie z. B. von Fisher, auswählt³⁸.

b) Geschichtete Stichprobe

„Unter einer geschichteten Stichprobe versteht man die Aufteilung der Population in Unterpopulationen, „Schichten“ genannt, und die Wahl je einer Stichprobe aus diesen Schichten³⁹“. Jede Schicht wird sodann wie eine einfache Stichprobe behandelt⁴⁰.

Während der sehr umfangreichen in den Vereinigten Staaten durchgeführten Umfrage im Jahre 1960 zur Berechnung der Summe der Begabungsreserven von Jugendlichen im Alter von 15 Jahren⁴¹, erfolgte die

38 Anm. d. dtsh. Bearbeiters: Sie wird dargestellt in: Mittenecker, E.: *Planung und statistische Auswertung von Experimenten*. Wien 1969, S. 154. Diese Tafel bietet eine höhere Garantie als ein einfaches dem Zufall überlassenes Losen, bei dem, wenn es exakt sein soll, z. B. vollkommen homogene stets in der gleichen Art vermischte Kugeln erforderlich sind.

39 L. Kish, in L. Festinger u. D. Katz, op. cit., S. 223.

40 Wir unterscheiden hier nicht zwischen der Straufizierung im eigentlichen Sinn und der Gruppierung (*clustering*). Bei der letzteren sind die Stichprobeneinheiten bereits bestehende Gruppen (z. B. vollständige Schulen) (auch: Klumpenstichprobe).

41 Siehe G. de Landsheere, *L'operation Talent aux U.S.A. (Le Travail humain, Paris, Nr. 3-4, Juli-Dezember 1961)*.

Schichtung aufgrund von quantitativen und qualitativen geographischen Kriterien.

Man traf folgende Unterscheidungen:

1. 56 Gruppen, von denen 50 auf die Staaten insgesamt entfielen und 5 die größten Städte des Landes repräsentierten (man stellte somit sicher, daß alle Stadt- und Landbezirke in ihren erheblich unterschiedlichen sozio-ökonomischen Niveaus vertreten waren);
2. 4 Gruppen aus höheren Schulen, nach der Zahl der im letzten Jahr eingetragenen Schüler: Gruppen von 0 bis 24 Schüler, von 25 bis 99, von 100 bis 399 und darüber (angemessene Repräsentation je nach Größe der Schule);
3. Um dem Unterschied Rechnung zu tragen, der zwischen solchen Schulen besteht, die von der Mehrzahl der Schüler bis zur Erlangung der Abiturreife besucht wird und den übrigen Schulen, hat man auch Gruppen aufgestellt entsprechend dem Verhältnis zwischen der im ersten Jahr aufgenommenen Schülerzahl und der Zahl der Abiturienten;
4. Darüber hinaus wurde die Stichprobe durch die Repräsentation von Sonder- bzw. Versuchsschulen vervollständigt;
5. Schließlich wurden alle Jugendlichen im Alter von 15 Jahren, die sich nicht mehr im höheren Schulunterricht befanden, geprüft, und zwar in 1/10 der in der Stichprobe dargestellten Bezirke.

Dieses Beispiel zeigt den Grad der Komplexität, den ein geschichtetes Stichprobenverfahren erreichen kann, und vermittelt einen Eindruck vom Umfang der erforderlichen Vorarbeit. Natürlich variieren die Kriterien des Stichprobenverfahrens je nach den Erfordernissen der Untersuchung.

Jede Stichprobe bei einer Unterpopulation (Schicht) muß ihrerseits den Kriterien der *Repräsentativität* entsprechen. Doch es geschieht nicht selten, daß wegen der Stratifizierung bestimmte Gruppen nicht mehr genügend Schüler zählen, um statistisch *signifikante* Resultate zu erbringen.

c) Gewichtete geschichtete Stichprobe

Wetterhin wäre es in vielen Fällen falsch, jeder Schicht das gleiche Gewicht beizumessen. Wenn man beispielsweise nach den von Fourasté unterschiedenen drei Sektoren stratifiziert: dem primären, sekundären, tertiären, so sind diese drei Klassen mit Sicherheit quantitativ ungleich in der Population vorhanden. Man arbeitet daher mit proportionierten oder gewichteten geschichteten Stichproben.

Es sei jedoch darauf hingewiesen, daß man bei einer Untersuchung, die sich auf hohe Zahlen erstreckt, häufig damit rechnen kann, daß die Stichprobe sich selbst gewichtet (*self-weighting*).

Beispiel: Gewichtete geschichtete Stichprobe aufgrund der Beschäftigung des Familienvaters.

Nationale Verteilung bei den drei Sektoren	primär 15 %	sekundär 30 %	tertiär 55 %	total 100 %
Schülerzahl in den untersuchten Schulen	1 000	5 000	4 000	10 000
Stichprobe 1/10, nicht-gewichtet	100	500	400	1 000
Nach der nationalen Verteilung gewichtete Stichprobe	150	300	550	1 000

d) Situationen-Stichproben

Die Vorkehrungen, die wir soeben festgehalten haben, gründen die Repräsentativität auf die Wahl der Probanden.

E. Brunswick⁴² hat darauf aufmerksam gemacht, daß es auch von Wichtigkeit sei, die Situationen nach dem Stichprobenverfahren zu wählen, aus deren Anlaß die Beobachtungen erfolgen. Wenn die Umfrage sich auf die ganz allgemeine Verhaltensweise von Kindern erstreckt, kann man die Schlußfolgerungen natürlich nicht nur allein auf die in der Schule beobachteten Verhaltensweisen gründen. Es muß auch das Verhalten auf der Straße und das Benehmen zu Hause berücksichtigt werden, und es muß systematisch festgelegt werden, welchen Platz man den verschiedenen Beobachtungsvoraussetzungen innerhalb der Gesamtstudie vorbehält. Man stratifiziert gewissermaßen je nach den Situationen (*representative design*).

Eine solche Technik ist sicherlich nicht leicht anwendbar, aber sie kann zu wesentlichen Fortschritten führen.

D. Faktoren, die die Wahl des Stichprobenverfahrens bestimmen

Zusammenfassend kann man unter den Faktoren, die die Wahl des Stichprobenverfahrens bestimmen, folgende festhalten:

1. Die zu untersuchende Population;
2. Die Struktur dieser Population;
3. Die Zielsetzungen der Untersuchung;
4. Den für die Resultate gewünschten Präzisionsgrad;
5. Die zur Verfügung stehende Zeit und die Informationsquellen.

⁴² E. Brunswick, *Representative Design of Psychological Experiments*, Berkeley, Univ. of California Syllabus Series, Nr. 304, 1947, zitiert von R. Travers, *Introduction...*, S. 397-399).

III. Die statistische Analyse

Einführung

Der Pädagoge ist zwei vielleicht gleich großen Gefahren ausgesetzt: zum einen ist es die Annahme, daß die Statistik derart kompliziert sei, daß es vergeblich sei, sich damit vertraut zu machen, und zum andern die Annahme, daß einige selbst relativ komplexe Begriffe dabei das Alpha und Omega darstellen.

Die moderne Statistik ist eine Wissenschaft, die allein von Spezialisten in der Mathematik beherrscht wird. Diese Wissenschaft ist in ständiger Entwicklung begriffen: Um sich davon zu überzeugen, braucht man nur die Bibliographien eines Jahres durchzublättern. Darum ist es unumgänglich, daß der Plan, bevor man eine umfangreiche Untersuchung startet, einem Statistiker vorgelegt wird, und daß man gemeinsam mit ihm Überlegungen anstellt, wie die Informationen zu analysieren sind.

Dennoch leisten für den täglichen Gebrauch einige grundlegende Verfahrenswesen, die leicht zu handhaben sind, außerordentlich gute Dienste.

Leider werden die einfachsten statistischen Begriffe vielfach in einer schwerverständlichen Sprache und mit abschreckenden Allgemeinbetrachtungen dargeboten. So schreibt S. Corey in seinem Buch über die Handlungsforschung in der Pädagogik: „Eine der wirksamsten Arten, die Lehrer und ganz allgemein alle diejenigen zu entmutigen, die sich mit dem täglichen Leben in der Schule beschäftigen oder experimentelle Versuche unternehmen, ist die, daß man der theoretischen Statistik eine zu große Bedeutung beimißt⁴³“.

Den jungen Forschern sollte ein Handbuch zur Verfügung stehen, aus dem alle überflüssigen Erläuterungen entfernt sind und die wesentlichen statistischen Techniken in allgemeinverständlicher Sprache, und zwar als technische Hilfsmittel zusammen mit ihrem Anwendungsmodus beschrieben werden: Verfahrensweise und Anwendungsgebiet.

Gewiß ist man bei unzureichenden theoretischen Kenntnissen Fehlurteilen ausgesetzt. Aber die Erfahrung lehrt, daß von dem Augenblick an, da sie schon mit einigen Techniken vertraut sind, sowohl der Studierende als auch der junge Forscher ein gewisses Vertrauen in sich selbst entwickeln und ihre Kenntnisse in dem Maße bereichern, in dem sie eine echte Notwendigkeit dafür empfinden.

⁴³ S. M. Corey, *Action Research to Improve School Practices*, New York, Teachers College, Columbia Univ., 1953, S. 126.

Auf den folgenden Seiten, die in keiner Weise als ein statistischer Unterricht anzusehen sind⁴⁴, finden wir einige wichtige einfach formulierte Begriffe aufgezeichnet.

A. Die Häufigkeitsverteilung

Die folgenden Maßzahlen sind das Ergebnis einer Prüfung von 100 nicht-selektierten gleichaltrigen Schülern.

Punktwerte	Zahl der Schüler, die diese Punktwerte erreichten (Häufigkeit) (f_j)
26 I	1
30 I	1
32 I	1
34 II	2
36 II	2
38 III	4
40 III	5
42 III	5
44 III I	6
46 III II	7
48 III III	10
50 III III II	12
52 III III	10
54 III II	7
56 III I	6
58 III I	5
60 III	5
62 III	4
64 III	3
66 II	2
70 I	1
76 I	1

N = 100

⁴⁴ Lehrbücher sind zahlreich. Außer den bereits zitierten Werken sollte man nachschlagen: H. Fischer, *Les méthodes statistiques en psychologie et en pédagogie*, Neuchâtel, Delachaux u. Niestlé, 1955; J. Guilford, *Fundamental Statistics in Psychology and Education*, New York, Mc. Graw, 1956, 3. Aufl. A.L. Edwards, *Statistical Analysis for Students in Psychology and Education*, New York, Rinehart and Co., 1956, 9. Aufl. Anm. d. dtsh. Bearbeiters: Literatur in dtsh. Sprache: Helen M. Walker, *Statistische Methoden für Psychologen und Pädagogen*. Beltz, Weinheim. 9. Aufl. 1967. E. Mittenecker, *Planung und statistische Auswertung von Experimenten*. Deuticke, Wien, 6. Aufl. 1966. F. Hillebrandt, *Elementare Statistik für Pädagogen, Psychologen und Soziologen*. Reinhardt, München, 1965.

Aufstellung von Klassen. Das obige Beispiel enthält nur 22 verschiedene Werte; diese Zahl könnte natürlich sehr viel höher liegen.

Zur leichteren Handhabung, insbesondere im Hinblick auf bestimmte Berechnungen, denen wir begegnen, zieht man es vor, die Maßzahlen in eine niedrige Anzahl von Klassen aufzugliedern: 10 bis 20 genügen in der Mehrzahl der Fälle. Gewöhnlich wählt man eine ungerade Klassenzahl, um eine mittlere Klasse zu haben.

Verfahren:

1. Berechnung der Variationsbreite zwischen den beiden extremen Punktwerten: 76–26 = 50.
2. Dividieren dieser Streubreite durch 15^{45} : $50 : 15 = 3,33$.
3. Auswählen einer der beiden nächstliegenden ungeraden Zahlen als Intervallbreite: 3 oder 5. Man wählt hier die 5, angesichts der geringen Anzahl der Werte.
4. Einordnen der ranghöchsten Maßzahl in die Mitte des obersten Intervalls: Die höchste Klasse ist demnach: 74–75–76–77–78.

	Klassen	Häufigkeiten (f_k)	
1	74–78	I	1
2	69–73	I	1
3	64–68	III	5
4	59–63	III III	9
5	54–58	III III III III	18
6	49–53	III III III III II	22
7	44–48	III III III III III	23
8	39–43	III III	10
9	34–38	III III	8
10	29–33	II	2
11	24–28	I	1
11 Klassen	Intervallbreite (i) = 5	Anzahl der Werte (N) = 100	

B. Maße für die zentralen Tendenzen

1. Arithmetisches Mittel (\bar{X})

Einfachster Merksatz: Für die Beurteilung der allgemeinen Tendenz einer

⁴⁵ Diese Zahl ist beliebig. Die Praxis hat gezeigt, daß sie in der Mehrzahl der Fälle die bequemste ist.

Meßungsgruppe errechnet man das Mittel (\bar{X}), indem man die Summe (Σ) der Meßwerte (die X) durch ihre Anzahl (N) dividiert:

$$\bar{X} = \frac{\Sigma X}{N}$$

Der Mittelwert der als Beispiel angegebenen Werte ist 50. Weiter unten findet sich eine Kurzmethode zur Berechnung des Mittelwertes.

Der Mittelwert kann jedoch ein falsches Bild von einem Gesamtergebnis vermitteln. Bei einer Prüfung erhalten 10 Schüler 6 Punkte, ein Schüler 12 Punkte und ein Schüler 60 Punkte. Der Mittelwert ist 11 Punkte. Dennoch erreichen 10 von 12 Schülern dieses Niveau nicht. Die beiden folgenden Indizes erlauben eine feiner abgestufte Bewertung.

2. Medianwert (Mdn)

Das ist der Punkt, der eine Reihe von vorgegebenen Werten in zwei gleiche Teile teilt:

1–2–3

1–2 : 3–4

3. Modalwert (Md)

Dies ist der am häufigsten vorkommende Wert.

Wenn man bei einer Reihe von Ergebnissen verschiedentlich gleichhäufig auftretende Werte feststellt, hat man es mit einer Verteilung zu tun, die nicht dem Gesetz der Normalverteilung zu entsprechen braucht (zwei Modalwerte ergeben eine bimodale Kurve).

Bei einer vollkommenen Normalverteilung koinzidieren Mittelwert, Medianwert und Modalwert, desgleichen bei völlig symmetrischen Verteilungen.

C. Streuung oder Abweichung bezogen auf den Mittelwert

Das folgende Beispiel zeigt, daß die Indizes der zentralen Tendenz nicht ausreichen, um ein zuverlässiges Bild von der Wirklichkeitssituation zu geben.

Resultate einer Schülergruppe: 50–40–30–20–10.

Resultate einer zweiten Gruppe bei der gleichen Prüfung;

31–30,5–30–29,5–29.

In beiden Fällen ist:

N: 5
Mittelwert: 30
Medianwert: 30

Diese beiden Ergebnisreihen sind jedoch sehr verschieden. Während die erste Gruppe heterogen ist, überrascht die zweite durch ihre Homogenität. Außer der zentralen Tendenz muß daher auch die Streuung berücksichtigt werden.

Eine erste Vorstellung der Streuung vermittelt uns die Variationsbreite:

$$\text{Fall 1: } 50 - 10 = 40$$

$$\text{Fall 2: } 31 - 29 = 2.$$

Die Abweichung vom Mittelwert ergibt auch eine klare Indikation. Für die extremen Werte der ersten Gruppe liegt sie bei 20; für die Werte der zweiten Gruppe nur bei 1.

Der meist angewandte Index hingegen ist die Standardabweichung: Wurzel aus dem Durchschnitt der quadrierten Abweichungen vom Mittelwert.

Bei einer Normalverteilung bringt die Standardabweichung besonders interessante Anwendungsmöglichkeiten.

D. Die Normalverteilung und die Gauss'sche Kurve

1. Wahrscheinlichkeit

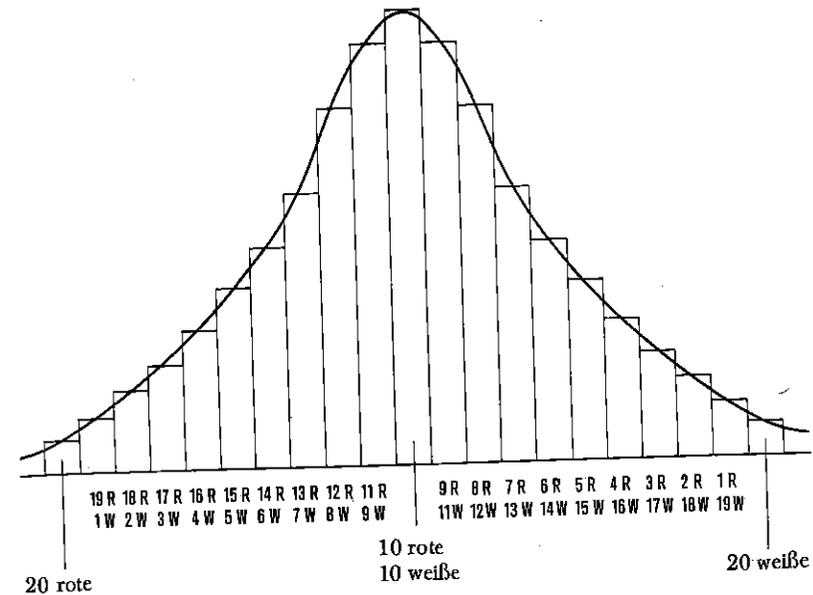
In einem Behälter befinden sich zwei in jeder Hinsicht gleiche Kugeln, doch ist die eine rot und die andere weiß.

Unter diesen Voraussetzungen besteht für beide Kugeln die Chance: 1 : 2, gezogen zu werden.

Bei 20 Ziehungen ist es demnach wenig wahrscheinlich, daß man 20 mal nacheinander die gleiche Farbe zieht; bei 200 Ziehungen ist die Chance, 200 mal hintereinander die gleiche Farbe zu ziehen, noch weitaus geringer usw.

Bei der folgenden auf einer Annahme beruhenden graphischen Darstellung entspricht jedes Rechteck 20 Ziehungen. Das erste und kleinste Rechteck stellt die Wahrscheinlichkeit dar, 20 mal hintereinander die rote Kugel zu ziehen. Die Chance, 19 mal die rote und 1 mal die weiße Kugel zu ziehen, ist schon größer, und so fort.

Die glockenförmige Gauss'sche Kurve verläuft, bezogen auf ihre Mittelachse, völlig symmetrisch und ist der Grenzfall der vorausgegangenen Verteilung.



2. Nutzen der Normalverteilung für die erziehungswissenschaftliche Forschung

Man hat die Feststellung gemacht, daß viele menschliche Eigenschaften dahin tendieren, sich nach diesem Zufallsprinzip zu verteilen. So hat man z. B. beobachtet, daß bei Gruppen, die nach dem Zufallsprinzip gewählt wurden, die Zahl der durchschnittlichen Probanden merklich größer ist als die der Probanden mit sehr hohem oder sehr niedrigem Niveau (Intelligenz, Leistungen auf verschiedenen Gebieten).

Da die Schul- oder Testergebnisse logischerweise dem Potential der Probanden entsprechen müssen, hat man daraus geschlossen, daß die Verteilung der Werte auch nach einer Gauss'schen Kurve erfolgen könnte.

Hier muß jedoch unmittelbar bemerkt werden, daß die „normale“ Verteilung bei weitem nicht alle untersuchten Phänomene charakterisiert, und daß man es vermeiden sollte, aus der Glockenkurve einen Lehrsatz aufzustellen. Aber vorbehaltlich dieser wichtigen Feststellung leistet die Theorie der Normali-

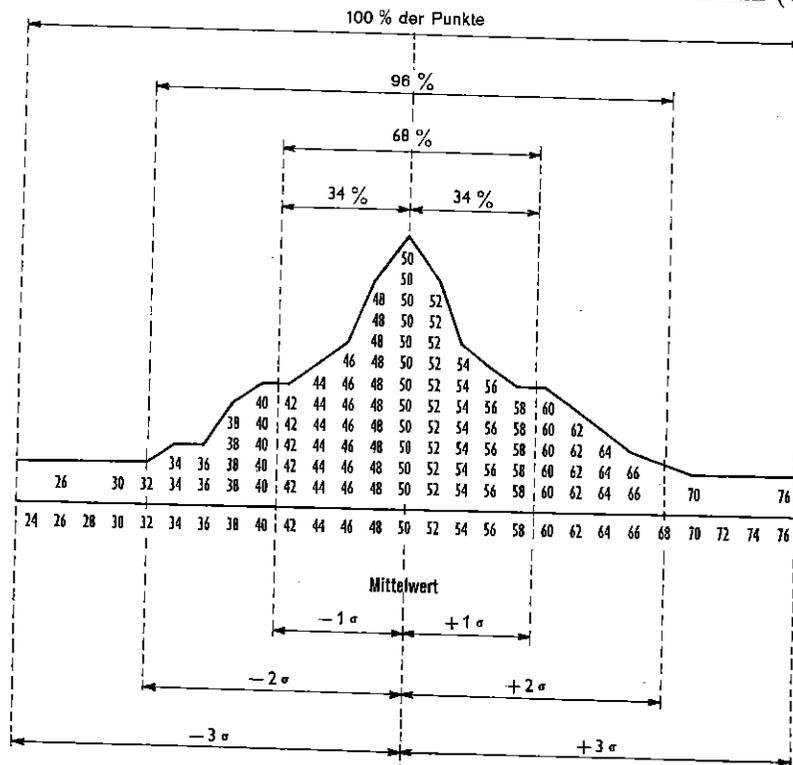
tät sicherlich unermessliche Dienste und stellt zweifellos eines der wichtigsten Instrumente für den Forscher dar.

Zwei Meßwerte kennzeichnen die Normalkurve: Der Mittelwert (μ), der der größten Höhe entspricht, und die Standardabweichung (σ), die die mehr oder weniger große Spannweite der Kurve – oder der Resultate – im Verhältnis zum Mittelwert angibt.

Die Standardabweichung kennzeichnet einen konstanten Flächenanteil der Kurve und somit stets den gleichen prozentualen Anteil der Resultate im Verhältnis zur Gesamtheit, ungeachtet ihrer Ausdehnung.

Die folgende Abbildung zeigt:

1. Daß die 100 Werte, die uns seit Beginn als Anschauungsmittel dienen, sich nicht nach einer vollkommenen Gauss'schen-Kurve verteilen (die



Skizze nach R.M. Thomas, Judging Student Progress, London, Longmans, Green and Co., 1960, Seite 504.

Zahl der Fälle ist nicht hoch genug), sondern nach einer dieser Kurve annähernd gleichen Form.

Man stellt die Hypothese auf, daß die Stichprobe von 100 Werten aus einer normalverteilten Gesamtheit stammen kann. Diese Hypothese konnte nicht widerlegt werden.

2. Welchen Prozentanteil der Fälle man erfaßt, je nachdem, ob man mit einem, zwei oder drei Standardabweichungen vom Mittelwert abweicht. Bei Vorliegen einer idealen Normalverteilung hätten sich 68,25%, 95,44% und 99,74% ergeben^{45a}.

E. Errechnung des Mittelwertes und der Standardabweichung

1. Ausführliche Methode:

Mittelwert:
$$\bar{X} = \frac{\sum x}{N}$$

Standardabweichung:
$$S = \sqrt{\frac{\sum d^2}{N-1}}$$
 wobei $d = X - \bar{X}$ ist.

2. Kurzmethode:

1. Zusammenfassen der Maßzahlen in Klassen;
2. Den Mittelpunkt von jeder Klasse (X_K) nehmen;
3. Häufigkeitstabelle (f_K);
4. Als angenommenen Mittelwert den Mittelpunkt der zentralen Klasse (\bar{X}_P) nehmen;
5. Die Abweichung der Klasse (d_K) errechnen. Die zentrale Klasse wird als Ausgangspunkt genommen:

1	+ 5
2	+ 4
3	+ 3
4	+ 2

^{45a} Anm. d. dtsh. Bearbeiters: Hier und in der Folge wurden die in Deutschland gebräuchlichen Berechnungsformeln eingesetzt.

5	+ 1
6	0
7	- 1
8	- 2
9	- 3
10	- 4
11	- 5

6. die (d_K) mit den Häufigkeiten multiplizieren: $(f_K) \times (d_K)$;
7. die Quadrate von (d_K) mit (f_K) multiplizieren;
8. Addition von $(f_K) \times (d_K)$;
9. Addition von $(f_K) \times (d_K)^2$;
10. den Korrekturwert errechnen:

$$C = \frac{\sum f_K \cdot d_K}{N}$$

Mittelwert:

$$\bar{X} = \bar{X}_p + i \cdot C$$

Standardabweichung:

$$S = i \cdot \sqrt{\frac{\sum f_K \cdot d_K^2 - C^2}{N - 1}}$$

Korrekturwert: $C = \frac{\sum f_K d_K}{N} = \frac{-20}{100} = -0,2$

Mittelwert: $\bar{X} = \bar{X}_p + i \cdot C = 51 + (5 \cdot (-0,2)) = 50$

Standardabweichung:

$$S = i \cdot \sqrt{\frac{\sum f_K (d_K)^2 - C^2}{N - 1}} = 5 \cdot \sqrt{\frac{332 - (0,2)^2}{99}} \approx + 9$$

Nr. der Klasse (K)	Klassengrenzen	Mittelwert der Klasse (\bar{X}_K)	Häufigkeit (f_K)	Abweichung der Klasse (d_K)	$(f_K) \times (d_K)$	$(f_K) \times (d_K)^2$
1	74-78	76	1	+ 5	+ 5	25
2	69-73	71	1	+ 4	+ 4	16
3	64-68	66	5	+ 3	+15	45
4	59-63	61	9	+ 2	+18	36
5	54-58	56	18	+ 1	+18	18
6	49-53	angen. Mittelwert (\bar{X}_p) 51	22	0	0	0
7	44-48	46	23	- 1	-23	23
8	39-43	41	10	- 2	-20	40
9	34-38	36	8	- 3	-24	72
10	29-33	31	2	- 4	- 8	32
11	24-28	26	1	- 5	- 5	25

Intervallbreite (i) = 5	N = 100	332
		+60
		-80
		-20

F. Die Eichung oder Berechnung der relativen Position

Einen Test eichen, heißt einen numerischen Rahmen liefern, der es erlaubt, die Ergebnisse eines Probanden bequem in ein Verhältnis zu den Resultaten der Population zu bringen, der er angehört.

1. Die Percentilberechnung

„Man bezeichnet als Percentile 0, 1, 2, ..., 100: die Werte einer Variablen bei denen 0%, 1%, 2% ..., 100% der Beobachtungen niedriger liegen als sie⁴⁶“.

Das erste Percentil ist daher das schlechteste und das 100. das beste⁴⁷, wenn ein niedriger Wert seine schlechte Leistung anzeigt.

46 H. Pierson, Vocabulaire de la psychologie.

47 Bei dem französischen System — das immer weniger angewandt wird — ist das erste Percentil das beste.

Errechnung:

- Zusammenfassen der Maßzahlen in Klassen (man beginnt mit der untersten Klasse);
- Die Häufigkeiten (f_K) angeben;
- Die kumulierten Häufigkeiten errechnen (F_K): man addiert jede Häufigkeit mit den ihr vorausgehenden;
- Folgende Formel ist anzuwenden (Abk. vgl. S. 279 u. 280):

$$C_X = U + \frac{(N \cdot C)/100 - F_{K-1}}{f_K} \cdot i$$

Klassen	Häufigkeiten (f_K)	Kumulierte Häufigkeiten (F_K)
24-28	1	1
29-33	2	3
34-38	8	11
39-43	10	21
44-48	23	44
49-53	22	66
54-58	18	84
59-63	9	93
64-68	5	98
69-73	1	99
74-78	1	100

	Errechnung des 50. Percentils	Errechnung des 16. Percentils	Errechnung des 84. Percentils
$N \cdot C$	100×50	100×16	100×84
100	100	100	100
U	$\frac{49 + 48}{2} = 48,5$	$\frac{39 + 38}{2} = 38,5$	$\frac{54 + 53}{2} = 53,5$
F_{K-1}	44	11	66
f_K	22	10	18
C_{PR}	$48,5 + \frac{50-44}{22} \cdot 5 = 49,9$	$38,5 + \frac{16-11}{10} \cdot 5 = 41$	$53,5 + \frac{84-66}{18} \cdot 5 = 58,5$

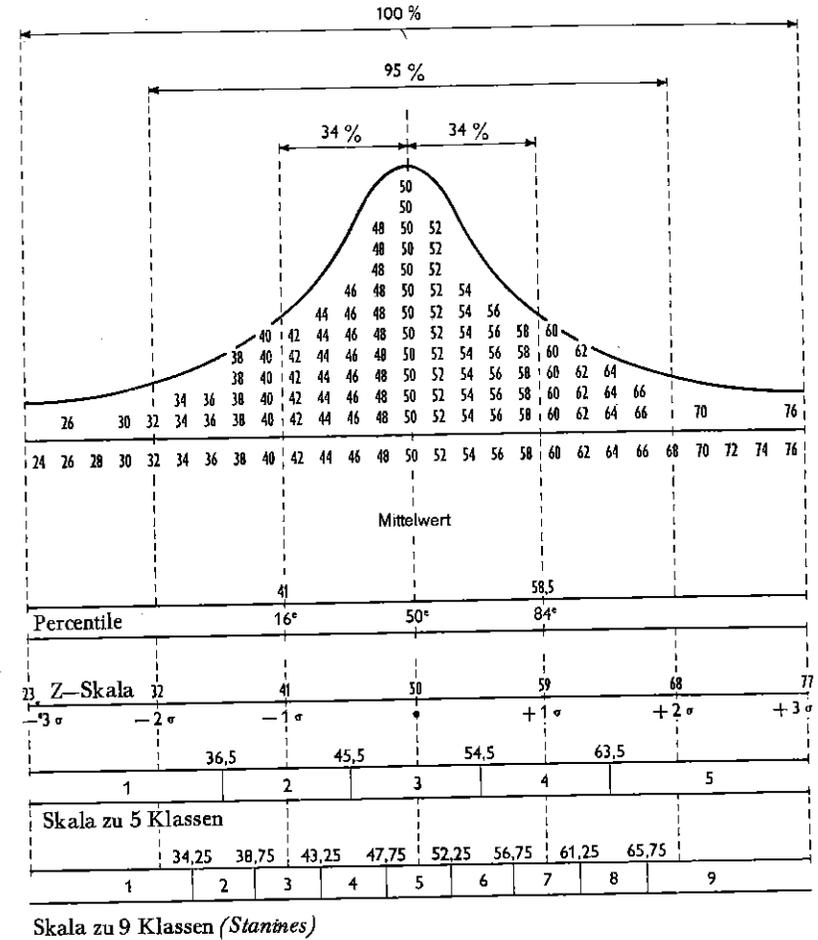
Mittlerer Quartilabstand ($C_{75} - C_{25}$)/2.

$$C_X = \frac{(N \cdot C)/100 - F_{K-1}}{f_K} \cdot i$$

Percentilwert von (X)

der Rang der Maßzahl, entsprechend einem beliebigen Percentil.

Beispiel: Das 50. Percentil entspricht dem Median, d.h. $N/2$. Bei den anderen handelt es sich um einen einfachen Dreisatz. Das 75. Percentil wird demnach sein: $(100 \cdot 75): 100 = 75$ (die Ziffern sind die gleichen, weil unser Beispiel genau 100 Maßzahlen enthält).



Skala zu 9 Klassen (Stanines)

U = der Mittelwert zwischen der unteren Grenze der Klasse, wo sich das gesuchte Perzentil befindet, und der oberen Grenze derjenigen Klasse, die ihr vorausgeht.

F_{K-1} = Kumulierte Häufigkeit der Klassen, die derjenigen mit dem gesuchten Perzentil vorausgehen.

f_K = die Häufigkeit der Fälle im Intervall, in dem sich das gesuchte Perzentil befindet.

i = Intervallbreite.

2. Die Normalisierung

Die Perzentilberechnung differenziert die mittleren Fälle schlecht. Tatsächlich liegt der größte Teil der in Perzentilen ausgedrückten Werte bei einer Normalverteilung zwischen -1σ und $+1\sigma$.

	-2σ	-1σ	0	$+1\sigma$	$+2\sigma$
Perzentile	C2	C16	C50	C84	C98

Um eine verfeinerte Klassifizierung zu erreichen, bezieht man sich auf Eichungen, bei denen die Anteile der Werte pro Klasse nicht mehr gleich sind, sondern den Häufigkeiten des Gesetzes der Normalverteilung entsprechen.

Die beiden zweckdienlichen Skalen sind folgende (vgl. Abb. S. 279):

a) Skala zu 5 Klassen.

Die 2., 3. und 4. Klasse sind eine Standardabweichung breit.

Die 3. Klasse ist auf den Mittelwert zentriert.

Die 1. und die 5. Klasse sind von unbegrenzter Ausdehnung.

	$-1,5\sigma$	$-0,5\sigma$	\bar{X}	$+0,5\sigma$	$+1,5\sigma$
	1	2	3	4	5
	7%	24%	38%	24%	7%

b) Die Skala zu 9 Klassen (Stanines)

Die 1. und die 9. Klasse sind von unbegrenzter Ausdehnung.

Die anderen Klassen umfassen eine halbe Standardabweichung, die 5.

Klasse ist auf den Mittelwert zentriert.

	$-1,75\sigma$	$-1,25\sigma$	$-0,75\sigma$	$-0,25\sigma$	\bar{X}	$+0,25\sigma$	$+0,75\sigma$	$+1,25\sigma$	$+1,75\sigma$
	1	2	3	4	5	6	7	8	9
	4%	7%	12%	17%	20%	17%	12%	7%	4%

G. Geeichte Werte oder Standardnormen

Ein Schüler hat die folgenden Gesamtpunktwerte erhalten⁴⁸: Rechnen: 22 von 25; Lesen: 72 von 100; Naturkunde: 26 von 50 Punkten.

Wenn der Lehrer im Lesen streng, jedoch im Rechnen weit weniger streng ist, ist das 72/100 im Lesen vielleicht dienstvoller als das 22/25 im Rechnen. Außerdem erlauben Rohwerte keinen Vergleich z.B. mit den Noten anderer Lehrer desselben Faches.

Um einen Vergleich anstellen zu können, drückt man diese Werte in Standardabweichungen als Einheit aus, wodurch ihre Verteilungen gleich werden (geeichte Werte oder Standardnormen).

Nehmen wir an, man erhält die folgenden Resultate:

	Mittelwert	Standardabweichung
Rechnen	15	2,5
Lesen	50	10
Naturkunde	29	5

Errechnung der Standardwerte: $z = (X - \bar{X}) : s$

Beispiel: $z = (22 - 15) : 2,5$

$z = 2,80$

Wenn z bekannt ist, kann man den Gesamtnotenwert finden:

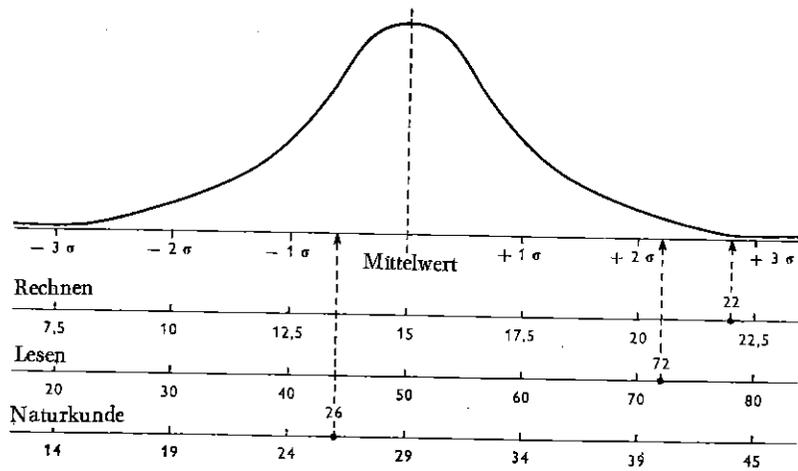
$X = \bar{X} + z \cdot s$

$X = 15 + (2,8 \cdot 2,5)$

$X = 22$

Graphische Darstellung:

⁴⁸ Nach R. Thomas, op. cit. S. 509.



H. Die Korrelation, Maß für den Zusammenhang

Der Korrelationskoeffizient (r) ist eine numerische Schätzung der zwischen zwei Variablen bestehenden Beziehung.

Der Korrelationskoeffizient liegt immer zwischen $+1$ und -1 : $1 =$ vollkommene positive Korrelation; $0 =$ Null-Korrelation; $-1 =$ vollkommene negative Korrelation.

Allgemeine Interpretation:

- $r < 0,20$: zu niedrige Korrelation;
- $r = 0,20-0,40$: niedrige Korrelation;
- $r = 0,40-0,60$: ausgeprägtere Korrelation;
- $r = 0,60-0,80$: gute Korrelation;
- $r > 0,80$: hohe Korrelation.

Anmerkung: Die angelsächsische Schreibweise von $0,20$ ist $.20$ usw.

Es gibt eine Vielzahl von Korrelationskoeffizienten⁴⁹, die jeweils nur unter bestimmten Voraussetzungen angewandt werden können.

Wir bringen als Anwendungsbeispiel den klassischsten Koeffizienten, und zwar den von Bravais-Pearson:

$$r = \frac{1}{(N-1)} \cdot \frac{\sum xy}{S_x \cdot S_y}$$

⁴⁹ Siehe die verschiedenen statistischen Handbücher.

wobei: x die Abweichungen zwischen den Werten der Variablen X und ihrem Mittelwert darstellt;
 y die Abweichungen zwischen den Werten der Variablen Y und ihrem Mittelwert darstellt.

Anwendung: Es wurde ein Intelligenztest und ein Test zur Messung des Gedächtnisses durchgeführt. Dabei hat man die in untenstehender Tabelle aufgeführten Werte erhalten. Besteht eine Korrelation zwischen diesen Resultaten? Wie hoch ist sie?

Nr.	Punkte für die Tests: Intelligenz/Gedächtnis		Abweichungen zwischen den Punkten und ihrem Mittelwert		x^2	y^2	xy
	X	Y	x	y			
1	7	8	+1,5	+2	2,25	4	+ 3
2	5	6	-0,5	0	0,25	0	0
3	3	2	-2,5	-4	6,25	16	+10
4	7	9	+1,5	+3	2,25	9	+ 4,5
5	4	6	-1,5	0	2,25	0	0
6	8	7	+2,5	+1	6,25	1	+ 2,5
7	6	8	+0,5	+2	0,25	4	+ 1
8	5	5	-0,5	-1	0,25	1	+ 0,5
9	6	7	+0,5	+1	0,25	1	+ 0,5
10	5	4	-0,5	-2	0,25	4	+ 1
11	6	7	+0,5	1	0,25	1	+ 0,5
12	5	6	-0,5	0	0,25	0	0
13	4	6	-1,5	0	2,25	0	0
14	8	7	+2,5	+1	6,25	1	+ 2,5
15	3	5	-2,5	-1	6,25	1	+ 2,5
16	4	6	-1,5	0	2,25	0	0
17	6	5	+0,5	-1	0,25	1	- 0,5
18	7	6	+1,5	0	0,25	0	0
19	8	7	+2,5	+1	6,25	1	+ 2,5
20	3	3	-2,5	-3	6,25	9	+ 7,5
Summe:	110	120			53	54	38
Mittelwerte:	$\bar{X} = 5,5$	$\bar{Y} = 6$					

$$S_x = \sqrt{\frac{\sum x^2}{N-1}} = \sqrt{\frac{53}{19}} = 1,7$$

$$S_y = \sqrt{\frac{\sum y^2}{N-1}} = \sqrt{\frac{54}{19}} = 1,7$$

$$r = \frac{1}{N-1} \cdot \frac{\sum xy}{S_x \cdot S_y} = \frac{1}{19} \cdot \frac{38}{1,7 \cdot 1,7} = 0,69$$

I. Die Bedeutung der Differenzen

1. Die Null-Hypothese

Eine Null-Hypothese aufstellen bedeutet die Annahme, daß die festgestellte Differenz zwischen zwei Maßen dem Zufallsprinzip zugeschrieben werden kann (Zufallsschwankungen bei Stichproben die aus einer gleichen Population entnommen sind, Meßfehler usw.).

Verifizierungstechniken erlauben es, die Nullhypothese beizubehalten oder sie zu verwerfen. Mit Hilfe von Formeln und Tabellen kann man tatsächlich feststellen, ob eine Wahrscheinlichkeit (P) in Höhe von maximal 5% ($P = 0,05$), 1% ($P = 0,01$) usw. besteht, daß die Differenz des Resultats oder der Kurven (im Falle der Normalitätskontrolle) auf bloßem Zufall beruhen kann.

Die Bedeutung einer solchen Verifizierung liegt auf der Hand.

Man darf diesen Test jedoch nicht überbewerten. Vergleicht man experimentell den Erfolg von zwei Unterrichtsmethoden, vorausgesetzt, die Chancen stehen 99% dafür, daß die Differenz des festgestellten Erfolgs nicht auf dem Zufallsprinzip beruht, so bedeutet das nicht *ipso facto*, daß diese Differenz ganz oder nur teilweise aus dem Übergang von der ersten zur zweiten Methode hervorgeht. Die statistische Kontrolle gibt an, daß die Differenz signifikant ist: das ist alles. Ob sie sich durch die neue Unterrichtsart, durch andere während des Experiments gewonnene Kenntnisse, durch einen Lehrerwechsel, durch eine Modifizierung der Motivation erklärt, wäre noch zu ergründen.

2. Prüfung der Signifikanz von Varianzunterschieden, bei der gleichen Population

Man wählt nach dem Zufallsprinzip zwei Schülergruppen einer gleichen Schülerpopulation und unterzieht sie der gleichen Prüfung.

Wenn die erhaltenen Werte für jede Gruppe dem Normalverteilungsgesetz entsprechen, ist es dennoch möglich und sogar wahrscheinlich, daß die

Kurven, die die Werte darstellen, nicht genau übereinstimmen indem sie z. B. verschiedene Standardabweichungen aufweisen.

Die Tabelle von Snedecor erlaubt es festzustellen, ob sich der Unterschied in den Standardabweichungen durch Schwankungen aufgrund des Zufallsprinzips erklären lassen kann oder aber durch eine signifikante Variation in der Zusammensetzung der Gruppen.

Die Prüfung wird folgendermaßen durchgeführt:

1. Man setzt jede Standardabweichung ins Quadrat (S_1^2 und S_2^2): Das Quadrat der Standardabweichung nennt man die Varianz. Die Varianzen sind hier mit v_1 und v_2 bezeichnet;
2. Man dividiert die höchste Varianz durch die andere: $F = v_1/v_2$ wenn $v_1 > v_2$;
3. Für jede Stichprobe ist die Zahl der Freiheitsgrade $N-1$:
 - für die größte Varianz: $(N_1-1) = v_1$;
 - für die andere: $(N_2-1) = v_2$.
4. Es genügt dann, sich die F-Tabelle von Snedecor zu vergegenwärtigen.

F-Tabelle von Snedecor (Auszug)

$v_1 \rightarrow$ $v_2 \downarrow$	1	2	3	4	500	∞
2	18,51 98,49	19 99	19,16 99,17	19,25 99,25			
3	10,13 34,12	9,55 30,82	9,28 29,46	9,12 28,71			
4	7,71 21,20	6,94 18	6,59 16,59	6,39 15,98			
1 000							
∞							

Beispiel:

$v_1 = 4$ und $v_2 = 3$.

Die Werte von F, die man bei dem Schnittpunkt der Spalte 4 und der Reihe 3 findet, geben an:

- Wenn $F = 9,12$ ist, so ist die Chance 5% ($P = 0,05$), daß diese Differenz der Varianzen zufällig zustande gekommen ist;
- Wenn $F = 28,71$ ist, ist die Wahrscheinlichkeit $P = 0,01$.

3. Prüfung der Signifikanz der Differenz zwischen zwei Mittelwerten bei korrelierenden Stichproben (kleine Stichproben von etwa 5 bis 60). Der t-Test von Student.

Der Index t in einer Tabelle erlaubt zu bestimmen, in welchem Maße eine festgestellte Differenz dem Zufallsprinzip zuzuschreiben ist oder nicht.

Formel:
$$t = \frac{\bar{D}}{S_D / \sqrt{N}}$$

wobei \bar{D} = die Differenz zwischen dem Mittelwert der Resultate nachher/vorher ist, S_D die Standardsabweichung der Einzeldifferenzen und N die Zahl der Wertepaare.

Beispiel: Man möchte untersuchen, ob die Verwendung von Lehrmaschinen während eines festgelegten Zeitabschnitts den Mittelwert der Ergebnisse im Rechnen verändert hat. 14 Schüler beteiligen sich an dem Experiment. Der Vergleich gründet sich auf zwei Parallelformen eines Rechentests.

Nr.	Werte		Differenz zwischen den Werten nachher-vorher	Quadierte Differenz (D^2)
	nachher	vorher		
1	255	272	-17	289
2	226	222	4	16
3	254	246	8	64
4	246	221	25	625
5	296	287	9	81
6	222	224	- 2	4
7	293	275	18	324
8	247	246	1	1
9	240	246	- 6	36
10	258	269	-11	121
11	236	216	20	400
12	199	197	2	4
13	256	271	-15	225
14	220	203	17	289
$N = 14$	3 448	3 395		2 479

Mittelwert der Punkte „nachher“: $3\,448 : 14 = 246,3$.
 Mittelwert der Punkte „vorher“ : $3\,395 : 14 = 242,5$.
 $\bar{D} = 246,3 - 242,5 = 3,8$.

$$s = \sqrt{\frac{\sum D^2}{N-1}} = \sqrt{\frac{2\,479}{13}} = 13,8.$$

$$t = \frac{3,8}{13,8 / \sqrt{14}} = \frac{3,8 \sqrt{14}}{13,8} = 1,02.$$

Zahl der Freiheitsgrade: $N-1 = 14-1 = 13$.

t-Tabelle von Student (Auszug) ^{49a}

Freiheitsgrade	P = 0,1 (Chance 10%)	P = 0,05 (Chance 5%)	P = 0,01 (Chance 1%)
1	6,31	12,71	63,66
2	2,92	4,30	9,92
↓	↓	↓	↓
13	1,77	2,16	3,01
↓	↓	↓	↓
120	1,66	1,98	2,62
↓	↓	↓	↓
∞	1,64	1,96	2,58

Unser t (1,03) liegt unter 1,77. Man bezeichnet diese Differenz als nicht signifikant. Der Unterschied kann zufällig sein.

4. Prüfung der Signifikanz einer Differenz oder einer Reihe von Differenzen zwischen Häufigkeiten oder Prozentangaben.

Chi-Quadrat Test

Der grundsätzliche Aufbau ist der gleiche wie bei den beiden vorangegangenen Tests. Wir gehen daher direkt zu einer Anwendung über: Man führt einen Test zur Messung der Kenntnisse bei Schülern des zweiten Schuljahres in drei Volksschulen einer Stadt durch, wobei man annimmt: homogenes Milieu und Klassenzusammensetzung ergäben sich nach dem Zufallsprinzip. Man klassifiziert die Leistungen in drei Kategorien: Hoch, Mittel, Niedrig. Besteht eine signifikante Differenz zwischen den Resultaten der drei Schulen?

In diesem Fall verwendet man den Chi-Quadrat-Test (χ^2 -Test):

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

49a Vereinfachte Tabelle.

wobei o = die beobachtete Häufigkeit,
 e = die erwartete Häufigkeit ist.

Beobachtete Häufigkeiten (= Resultate so wie sie zusammengetragen wurden)

	Schule I	Schule II	Schule III	Gesamt
Hoch	8	10	7	25
Mittel	16	15	7	38
Niedrig	11	15	11	37
Gesamt	35	40	25	100

Globale oder theoretische Häufigkeiten. Auf 100 Kinder entfallen für die 3 hier zusammengefaßten Schulen insgesamt: 25 in die „höhere“ Klasse eingestufte d.h. 25/100; 38 in die „mittlere“ Klasse eingestufte, d.h. 38/100 und 37 in die „niedrige“ Klasse eingestufte Kinder, d.h. 37/100.

Erwartete Häufigkeiten

	Schule I	Schule II	Schule III	Gesamt
Hoch	$35 \times \frac{25}{100} = 8,75$	$40 \times \frac{25}{100} = 10$	$25 \times \frac{25}{100} = 6,25$	25
Mittel	$35 \times \frac{38}{100} = 13,3$	$40 \times \frac{38}{100} = 15,2$	$25 \times \frac{38}{100} = 9,5$	38
Niedrig	$35 \times \frac{37}{100} = 12,95$	$40 \times \frac{37}{100} = 14,8$	$25 \times \frac{37}{100} = 9,25$	37
Gesamt	35	40	25	100

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

$$= \frac{(8-8,75)^2}{8,75} + \frac{(10-10)^2}{10} + \frac{(7-6,25)^2}{6,25} + \frac{(16-13,3)^2}{13,3} + \frac{(15-15,2)^2}{15,2} + \frac{(11-9,5)^2}{9,5} + \frac{(15-12,95)^2}{12,95} + \frac{(11-14,8)^2}{14,8} + \frac{(11-9,25)^2}{9,25} = 2,98.$$

Zahl der Freiheitsgrade: v
 Allgemeine Formel:

$$v = (c-1)(z-1)$$

c ist hier gleich der Anzahl der Gruppen: 3

z ist hier gleich der Anzahl der Schulen: 3

daher: $(3-1)(3-1) = 4$.

χ^2 -Tabelle (Auszug)

Freiheitsgrade	P = 0,1 (Chance 10%)	P = 0,05 (Chance 5%)	P = 0,01 (Chance 1%)
1	2,71	3,84	6,64
2	4,60	5,99	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
↓	↓	↓	↓
30	40,26	43,77	50,89

2,98 liegt unter 7,78: Daher werden die Differenzen als nicht signifikant bezeichnet. Die Unterschiede können zufällig sein.

5. Simultan-Prüfung der Signifikanz der Differenzen zwischen mehr als zwei Mittelwerten, die sich auf mehrere nach dem Zufallsprinzip durchgeführte Stichproben bei einer gleichen Population beziehen. *Varianzanalyse*.

(1) Gesamtsumme der Quadrate (*total*):

$$TSS = \sum X^2 - \frac{(\sum X)^2}{N} = 673$$

TSS =

(2) Summe der Quadrate (SS) zwischen Gruppen (*between*):

$$BSS = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_2} + \frac{(\sum X_3)^2}{N_3} - \frac{\sum X^2}{N} = 287$$

(3) Summe der Quadrate (SS) innerhalb der Gruppen (*within*):

$$WSS = TSS - BSS = (1) - (2) = 386$$

Problemstellung: 3 Gruppen zu je 10 Schülern werfen mit verbundenen Augen einen Ball gegen eine Zielscheibe. Man übt Kritik an den Schülern

der ersten Gruppe, ermutigt die Schüler der zweiten Gruppe und gibt bei der dritten Gruppe keinerlei Kommentar. Man zählt pro Schüler die Zahl der bis zu einem Erfolg notwendigen Versuche. Besteht zwischen diesen drei Behandlungsmethoden eine signifikante Differenz in ihrer Wirkung? (nach D. Van Dalen, Understanding Educational Research, New York, McGraw, 1962, Seite 323).

10 Schüler -Kritik-N ₁		10 Schüler -Lob-N ₂		10 Schüler -ohne Kommentar-N ₃		Gesamt allgemein N = 30
Ver- suchs- zahl bis zum Erfolg X ₁	X ₁ ²	Ver- suchs- zahl bis zum Erfolg X ₂	X ₂ ²	Versuchs- zahl bis zum Erfolg X ₃	X ₃ ²	
25	625	16	256	25	625	
26	676	11	121	15	225	
17	289	11	121	21	441	
19	361	18	324	17	289	
26	676	15	225	22	484	
23	529	16	256	24	576	
24	576	15	225	15	225	
25	625	20	400	22	484	
17	289	16	256	26	676	
19	361	10	100	15	225	
Summen:	221 5 007	148 2 284	202 4 250		$\Sigma X = 571$ $\Sigma X^2 = 11 541$	

Zusammenfassung der Varianzanalyse

Es genügt nunmehr, sich die Tabelle von Snedecor zu vergegenwärtigen, wobei:

$v_1 = 2$ (Zahl der Freiheitsgrade, entsprechend dem größten mittleren Quadrat)

$v_2 = 27$ (Zahl der Freiheitsgrade, entsprechend dem kleinsten mittleren Quadrat) ist.

Variations- quelle	Summe der Quadrate	Freiheitsgrade	Mittlere Quadrate	F
Zwischen den Gruppen (Behand- lung)	287	Zahl der Behandlungen - 1 = 3 - 1 = 2	$\frac{287}{2} = 143,5$	$\frac{143,5}{14,3} = 10$
Innerhalb der Gruppen (interne Fehler)	386	$(N_1 - 1) + (N_2 - 1) + (N_3 - 1)$ $= (10 - 1) + (10 - 1) + (10 - 1)$ $= 27$	$\frac{386}{27} = 14,3$	
Gesamt	673	N - 1 = 29		

F-Tabelle von Snedecor (Auszug)

$v_1 \rightarrow$	1	2	500	∞
$v_2 \downarrow$					
2	18,51 98,49	19 99	→		
3	10,13 34,12	9,55 30,82	→		
↓	↓	↓			
27	4,24 7,77	3,38 5,57	→		
↓	↓	↓			
∞					

F liegt über 5,57. Daher ist die Wahrscheinlichkeit kleiner als 1% ($P = 0,01$), daß die Differenz auf dem Zufall beruht. Die Nullhypothese wird abgelehnt, und die Ergebnisse des Experiments werden als signifikant bezeichnet.

Anhang

1. Berufsethische Verpflichtungen (Deontologie)

Bei jedem erziehungswissenschaftlichen Forschungsvorhaben muß die Achtung vor der menschlichen Persönlichkeit oberstes Gesetz sein.

Deshalb ist zu beachten:

- Die Forschung muß mit Zustimmung aller Beteiligten vorgenommen werden (Studenten, Eltern, Lehrer, Schulverwaltungsbeamte);
- die Forschung darf weder der physischen noch psychischen Gesundheit der Probanden schaden noch sie, gleich auf welche Art und Weise auch immer, in ihrem Lernen behindern;
- die Forschung darf weder die Intimsphäre des Einzelnen noch die der Familie verletzen.
- Das Berufsgeheimnis muß peinlich genau beachtet werden. Ohne ausdrückliche Genehmigung dürfen Veröffentlichungen von Forschungsgeheimnissen nicht dazu führen, daß Teilnehmer identifiziert werden können. Die bei den Forschungszentren aufbewahrten Untersuchungsdokumente dürfen nicht der zufälligen Verteilung eines Archives überlassen werden und so indiskreten Personen in die Hände fallen.
- Es muß also ein System vorgesehen werden, nach dem die Akten vernichtet werden.
- Eine Kodeschrift zur Identifizierung der Akten ist von Beginn der Untersuchung an einer ausgeschriebenen Namensgabe vorzuziehen.

Hinsichtlich des Berufsgeheimnisses hat die französische Gesellschaft für Psychologie folgende Bestimmungen angenommen¹:

1 Der vollständige Text, der am 7. Mai 1961 angenommen, die Berufspflichten betreffenden Bestimmungen, wurde in den der Psychologie gewidmeten Nummern 2–3 der „Revue de L'Enseignement supérieur“ veröffentlicht, 1966, S. 172–174.

Anm. d. dtsh. Bearbeiters: Auch der „Berufsverband deutscher Psychologen“ (BdP), Hamburg, hat solche „Berufsethische Verpflichtungen“ erarbeitet, die Bestimmungen über die Schweigepflicht, das Verhalten des Psychologen gegenüber der Öffentlichkeit, dem Klienten, dem Auftraggeber und den Kollegen enthält.

- „2.2. Besonders wird darauf hingewiesen, daß sich die Geheimhaltungspflicht im privaten Bereich der Klienten auf alles, was der Psychologe gesehen, gehört oder im Laufe seiner Praxis oder seiner Untersuchungen verstanden hat, erstreckt.
- 2.3. Das Berufsgeheimnis muß sowohl hinsichtlich des Wortlautes als auch bei der Aufbewahrung oder Weitergabe der Dokumente gewahrt bleiben. Der Psychologe hat so vorzugehen, daß die aus seiner Arbeit stammenden Dokumente (Beschlüsse, Zusammenfassungen, Meldungen, Berichte usw.) immer so abgefaßt, vorgetragen und geordnet sind, daß das Berufsgeheimnis gewahrt bleibt.
- 2.4. Außer in zwingenden Rechtsfällen kann der Psychologe von niemandem, selbst nicht von denjenigen, die die Geheimhaltung betrifft, von seiner Schweigepflicht entbunden werden.

Im Prinzip und soweit kein Gegensatz zu der Bestimmung der Wahrung des Berufsgeheimnisses besteht, haben diejenigen, die an einer pädagogischen Untersuchung teilnehmen, auch das Recht, deren Resultate zu kennen.”

- In jedem Fall muß die Mitteilung der Resultate mit Takt und Vernunft geschehen. Die brutale Aufdeckung von Unzulänglichkeiten, besonders von erniedrigenden Situationen kann schwere Schocks hervorrufen und schwerwiegende Folgerungen bewirken.
- Damit falsche Interpretationen vermieden werden, sollte der Forscher sich einer Ausdrucksweise bedienen, die dem kulturellen Niveau und den Kenntnisstand der ihn befragenden Personen angepaßt ist. Im Zweifelsfall ist eine Unterredung einer schriftlichen Mitteilung vorzuziehen.
- Versprechen müssen gehalten werden.

Beispiel: Bei einem soziometrischen Test ermuntert man oft die Schüler zum Mit-tun, indem man ihnen verspricht, die Ergebnisse als Grundlage für die Bildung von Arbeits- oder Spielgruppen zu benutzen. Ein solches Versprechen nicht ein-zuhalten, wäre nicht nur ein Betrug, sondern würde auch das Mißtrauen gegen-über den Forschern hervorrufen. Bei vielen Gelegenheiten kommt eine solche Haltung einer Politik der verbrannten Erde gleich.
Ein Großteil der pädagogischen Forschung vollzieht sich in Schulen. Oft fühlen sich die Lehrer selbst betroffen.

- Es ist äußerst wünschenswert, daß die Lehrer über die verfolgten Ab-sichten und die angewandten Mittel informiert werden und aktiv an den Arbeiten teilnehmen. Wenn dies nicht möglich ist, ist es wichtig, klar zu erklären, warum.
- Außer wenn es von Anfang an besonders abgesprochen ist, müssen die Lehrer über den Verlauf und die Resultate der Untersuchung auf dem

Laufenden gehalten werden (dies besonders, damit es keine Gegensätze zu den vorher erwähnten Bestimmungen gibt).

- Mehr als von jedem anderen erwartet man vom Forscher vollkommene wissenschaftliche Redlichkeit.
- Eine Untersuchung darf niemals, so wenig es auch sein mag, wissenschaftlich verfälscht werden.
- So trügerisch die Resultate auch sein mögen, sie müssen getreu wiedergegeben werden.
- Die von anderen Forschern erzielten Resultate sind ihr striktes Eigentum. Es ist besser, einmal zuviel seine Quellen anzugeben, als sich das Gut anderer, so wenig es auch sein mag, anzueignen.

Diesbezüglich hat die französische Gesellschaft für Psychologie folgende Bestimmungen angenommen.

- 4.0. Jeder Psychologe, welches auch immer sein Spezialgebiet ist (Forschung, Praxis, Unterrichtswesen), muß sich laufend über die in seiner Disziplin erzielten Fortschritte informieren. Er legt sie seiner Arbeit zugrunde und bemüht sich, zu ihrem Fortschritt beizutragen.

Er muß daher alle Regeln, Bedingungen und Unbequemlichkeiten die wissenschaftliche Arbeiten einem auferlegen, akzeptieren.

- 4.1. Jeder Psychologe bemüht sich, Kriterien und wissenschaftliche Methoden, die mittelbar und kontrollierbar sind, zu suchen und anzuwenden. Er trägt dadurch dazu bei, die Zuflucht zu autoritären Entscheidungen zu beschränken. „Nach wissenschaftlicher Gepflogenheit trägt er dafür Sorge, sein Wissen so vollständig wie möglich in einem von Genauigkeit und Wahrhaftigkeit bestimmten Geist weiterzugeben.“

Berufliche Freiheit

Wir beschränken uns darauf, die drei an der französischen Gesellschaft für Psychologie anerkannten Bestimmungen wiederzugeben.

„6.0. Der Psychologe darf keine Arbeitsbedingungen annehmen, die seine berufliche Unabhängigkeit beeinträchtigen würden, d. h., die ihn hindern würden, den in vorliegendem Kodex aufgeführten Berufspflichten nachzukommen.

6.1. Er muß seine berufliche Unabhängigkeit wahren, welches auch immer seine hierarchische Stellung bei seiner Arbeitsdurchführung sei.

6.2. Jeder Psychologe ist verpflichtet, seine Kollegen bei der Verteidigung ihrer Unabhängigkeit zu unterstützen.“

2. Das dreidimensionale Modell des Intellekts

Die Feststellung, daß die Intelligenz einer der wesentlichen Faktoren ist, die die pädagogische Forschung zu beachten hat, ist eine allgemein anerkannte Wahrheit. Die Faktoretheorie von J. P. Guilford² eröffnet uns in zahlreichen Gebieten neue und fruchtbare Perspektiven. Hier die wesentlichen Grundzüge:

J. P. Guilford versteht jegliches geistiges Verhalten als ein dreidimensional strukturiertes Gebilde: Operationen – Ergebnisse – Inhalte³.

Er unterscheidet fünf Arten von Operationen:

1. Gedächtnis
2. Kognition: Feststellung von Faktoren und Patterns
3. Divergierendes Denken: Möglichst viele verschiedene Lösungen zu einem vorgegebenen Problem finden.
4. Konvergierendes Denken: Die einzig richtige Lösung aus der Reihe möglicher Lösungen zu einem vorgegebenen Problem finden.
5. Werten.

Jede Operation kann zu sechs verschiedenen Arten von Ergebnissen führen:

1. Einheiten
2. Klassen
3. Beziehungen
4. Systeme
5. Transformationen
6. Implikationen

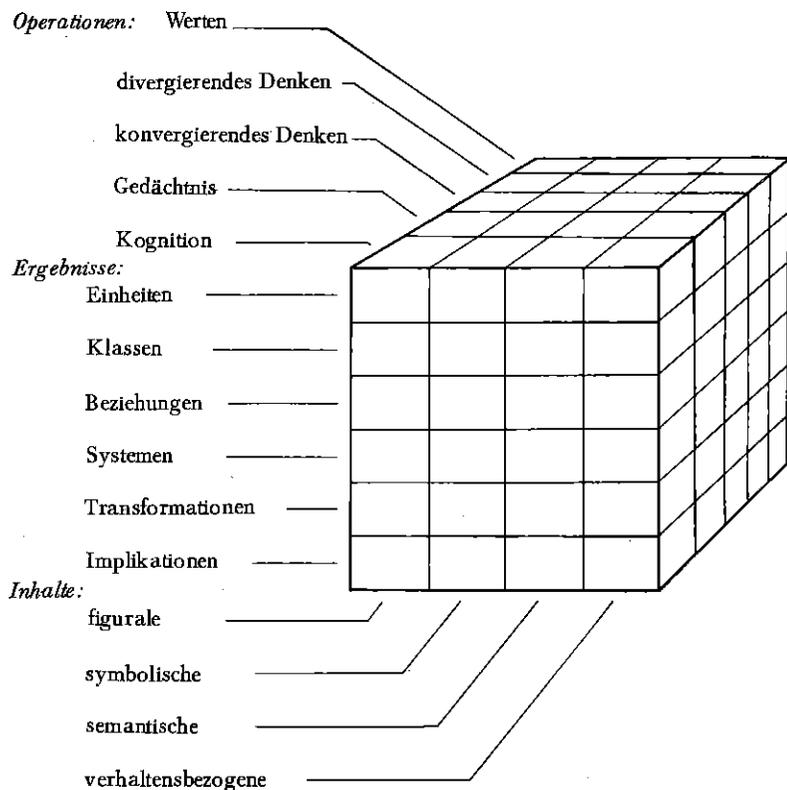
Jedes Ergebnis kann seinerseits abhängig sein von vier Inhaltskategorien:

1. Figuralen: Konkrete Gegenstände, wie sie von den Sinnen wahrgenommen werden; diese Gegenstände stellen nur sich selbst dar.
2. Symbolischen: Buchstaben und Zahlen; allgemein gebräuchliche Zeichen.
3. Semantischen oder Verbalen.
4. Verhaltensbezogenen: Eine Kategorie, die auf rein theoretischer Grundlage beigefügt wurde, um den allgemeinen Bereich, der gelegentlich als „soziale Intelligenz“ bezeichnet wird, darzustellen.

² Die Theorien von Guilford befinden sich noch in voller Entwicklung. Man vergleiche: J. P. Guilford: Three Faces of Intellect (The American Psychologist: Bd. 14, 1958, Nr. 8, S. 469–79) und die mit den Untersuchungen parallel laufenden Veröffentlichungen (U.S.C.).

³ Anm. d. dtsh. Bearbeiters: Die wichtigsten Grundzüge des Verfahrens von Guilford sind dargestellt in: Weinert, Fr.: Pädagogische Psychologie, Köln – Berlin 1957, S. 118 ff.

Guilford gelangt so zu folgender Darstellung:



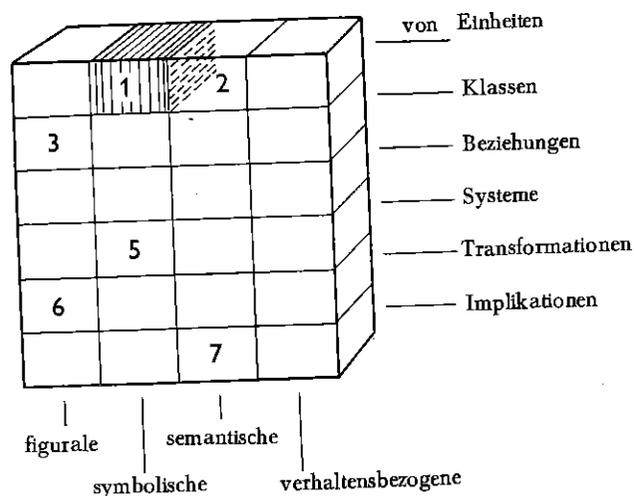
Dieses Modell bietet 120 Kombinationsmöglichkeiten, von denen jede einzelne eine Funktionsart darstellt, die durch einen Test untersucht werden kann.

Eine der von Guilford aufgezeigten Analysen soll uns zur Erläuterung dienen. Wir haben das divergierende Denken ausgewählt, da dieser Bereich einen besonderen Einblick erlaubt und sich im ganzen gesehen zunehmender Interessen erfreut⁴.

Um die Darstellung so klar wie möglich zu machen, isolieren wir zunächst den Abschnitt des Modells, den wir besprechen wollen.

⁴ Vgl. auch: G. de Landsheere: Für eine Pädagogik der Vielseitigkeit, Bericht Nr. 204 vom Mai 1963.
J. P. Guilford, Persönlichkeit, aus dem Amerikanischen übertragen von H. Kottenhoff und Ursula Agrell, Weinheim, 1964.

Divergierendes Denken



Block 1 stellt die „Fähigkeit zu divergierendem Denken von symbolischen Einheiten“ dar. Dieser Faktor wird beispielsweise in der Untersuchung der Wortflüssigkeit nach dem PMA von Thurstone (es sind möglichst viele Wörter zu schreiben, die mit S beginnen) getestet. Es ist zu bemerken, daß die von Thurstone als primär angesehene Fähigkeit nur 1/120 des Modells von Guilford darstellt.

Block 2 ist der, der die „Fähigkeit zu divergierendem Denken von semantischen Einheiten“ darstellt. Es ist die alte Vorstellungsfähigkeit (Test: Aufzählen einer größtmöglichen Anzahl von Gegenständen, die rund und essbar sind).

Block 3 Fähigkeit zu divergierendem Denken von figuralen Klassen wird erläutert durch einen Versuch, bei welchem der Proband aufgefordert wird, die größtmögliche Anzahl von Verwendungszwecken von einem Ziegelstein aufzuzählen.

Wenn der Proband antwortet: „Ein Haus, eine Scheune, eine Garage, eine Schule bauen...“, beweist er damit eine gewisse Vorstellungsfähigkeit, er beschränkt sich jedoch auf eine einzige Klasse (Bauen) und offenbart damit seine Armut an spontaner Flexibilität.

Antworten wie: „Einen Türeingang bauen, eine Papierpresse daraus machen, eine Katze damit ertränken, rotes Pulver davon ma-

chen ...", beziehen sich auf verschiedene Klassen und beweisen damit eine größere Vielseitigkeit.

Block 4 (Fähigkeit zu divergierendem Denken von semantischen Beziehungen) stellt die assoziative Flüssigkeit dar.

Test: Eine Liste von Wörtern aufstellen, die etwa dasselbe wie „gut“ bedeuten.

Block 5 (Fähigkeit zu divergierendem Denken von symbolischen Systemen) entspricht der Flüssigkeit des Ausdrucks.

Test: Bilden Sie, von den folgenden Anfangsbuchstaben von Wörtern ausgehend, eine möglichst große Anzahl von Sätzen:

N..... m..... u..... t.....

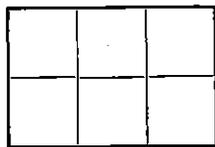
Beispiele von Antworten: Not macht uns traurig.

Nikotin macht uns träge.

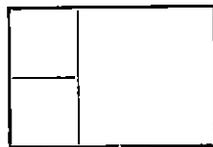
Nikolaus macht unser Tor.

Block 6 (Fähigkeit zu divergierendem Denken von figuralen Transformationen) beinhalten die adaptive Flexibilität.

Test: Streichholzaufgabe



A



B

Entfernen Sie von A vier Streichhölzer, so daß drei Quadrate übrigbleiben.

Block 7 (Fähigkeit zu divergierendem Denken von semantischen Implikationen) wird erläutert durch folgenden Test:

Test: Man legt dem Probanden die Skizze eines Planes vor. Der Proband muß diesen so detailliert wie möglich vervollständigen, so daß der Plan getreu verwirklicht werden kann.

3. Taxonomie der Erziehungsziele-Erkenntnisbereich

Sowohl für den Unterricht als auch für die Forschung ist ein klares Bewußtsein der Ausbildungsziele nötig. B. Bloom und seine Mitarbeiter haben eine Rangliste aufgestellt, in der das niedrigste Ziel das Auswendiglernen und die Wiederfindung von Einzeltatsachen ist und das höchste Ziel die Fähigkeit

Urteile zu fällen aufgrund innerer Evidenz und äußerer Kriterien. Trotz ihrer großen Bedeutung ist es uns nicht möglich, die Systematik Blooms in allen ihren Feinheiten und Einzelheiten zu studieren. Eine allgemeine Übersicht kann dennoch den Forscher anleiten und dürfte im besonderen erlauben, ein Gesamturteil abzugeben über die Verteilung der einzelnen Prüfungen von Kenntnissen, ob es sich um standardisierte Tests oder um einfachere Ergebnisprüfungen handelt.

I. Erkenntnis

1.00 Erkenntnis

Erinnerung an das, was spezifisch oder allgemein ist, an Methoden oder Verfahren, an Vorbildern, Strukturen oder Rahmen in Zeit und Raum.

1.10 Kenntnis spezifischer Gegebenheiten.

1.11 Kenntnis der Terminologie.

Beispiel: Bedeutung des Vokabulars der Symbole.

1.12 Kenntnis spezifischer Tatsachen.

Beispiel: Genauere oder annähernde Kenntnis von Daten, Orten, Personen usw.

1.20 Kenntnis, wie spezifische Gegebenheiten zu behandeln sind.

1.21 Kenntnis des Üblichen.

Beispiel: Der gute Ausdruck in der Muttersprache.

1.22 Kenntnis der Tendenzen und Folgen.

Beispiel: Die Kenntnis von der Entwicklung der Eisenindustrie.

1.23 Kenntnis der Einteilung und der Kategorien.

1.24 Kenntnis der Kriterien.

1.25 Kenntnis der Methoden.

(Beispiel dieser Untersuchung wird die Kenntnis selbst betont und nicht die Fähigkeit, die Methoden anzuwenden).

1.30 Kenntnis der allgemeinen Gegebenheiten und der Abstraktionen.

1.31 Kenntnis der Prinzipien und Gesetze.

1.32 Kenntnis der Theorien und Strukturen.

II. Geistige Fähigkeiten und Fertigkeiten

Diese Ziele weisen auf die geistigen Vorgänge hin, die es erlauben, Materialien zu sammeln oder neu zu ordnen, um ein bestimmtes Ziel zu erreichen. Die Materialien können von außen zur Verfügung gestellt oder durch das Gedächtnis geliefert werden.

2.00 Verständnis: Es handelt sich um das Verständnis der tiefsten Stufe. Der Prüfling kann das benutzen, was er kennt oder was ihm mitgeteilt wird, ohne notwendigerweise die Beziehungen zu anderen Materialien oder alle Folgerungen zu ersehen.

2.10 Übersetzung: Der Inhalt der ursprünglichen Mitteilung muß in der ursprünglichen Art beibehalten werden, aber die Form wird verändert.
Beispiel: Mathematische Symbole in Worte umsetzen können.

2.20 Interpretation: Erklärung oder Zusammenfassung einer Mitteilung. Die Interpretation enthält eine neue Anordnung, eine neue Sicht des Materials.

3.00 Anwendung: Benutzung von Abstraktionen in besonderen konkreten Situationen.

4.00 Analyse:

4.10 Die Analyse von Elementen.

4.20 Die Analyse von Beziehungen.

4.30 Die Analyse von Ordnungsprinzipien.

Beispiel: Die Fähigkeit, die allgemeinen Techniken festzustellen, die in Materialien zur Überredung benutzt werden: Reklame, Propaganda usw.

5.00 Synthese:

5.10 Erstellung einer einheitlichen Mitteilung.

5.20 Erstellung eines Planes, eines Vorschlags für ein planmäßiges Vorgehen.

5.30 Ableitung einer Reihe von abstrakten Beziehungen.

6.00 Schätzung:

6.10 Urteile fällen aufgrund innerer Evidenz.
Abschätzen der Genauigkeit einer Mitteilung aufgrund von Kriterien wie etwa der logischen Geschlossenheit.

6.20 Urteile fällen aufgrund äußerer Kriterien.

Beispiel: Eine Arbeit beurteilen durch Vergleich mit Vorbildern.