

De Landsheere
Einführung
in die
pädagogische
Forschung

beltz
com
pen
dium

G. De Landsheere

Einführung in die pädagogische Forschung



Gilbert De/Landsheere

Einführung in die pädagogische Forschung

Mit einem Vorwort von G. Mialaret
übersetzt und bearbeitet von
Hans Jörg unter Mitwirkung von Ulrich Raatz

Verlag Julius Beltz · Weinheim · Berlin · Basel

Dieser Band ist die Sonderausgabe des gleichnamigen Titels in der Reihe
 „Schriften zur Methodenlehre der empirischen Pädagogik“. Band 1.
 Herausgegeben von der Gesellschaft zur Förderung Pädagogischer Forschung e.V.

pac
001

Einführung

A 341 (2)



Dr. Gilbert De Landsheere, Ordinarius des Instituts für Psychologie und
 Erziehungswissenschaften der Universität Lüttich

Dr. G. Mialaret, Professor an der Humanwissenschaftlichen Fakultät der
 Universität Caen
 Präsident der Internationalen Gesellschaft für Experimentelle Pädagogik
 französischsprachender Länder

Dr. Hans Jörg, Professor an der Peter-Wust-Hochschule Saarbrücken

Dr. Ulrich Raatz, Leiter der Testabteilung am Deutschen Institut für
 Internationale Pädagogische Forschung, Frankfurt/Main

Titel des Originals: »Introduction à la recherche pédagogique«
 © 1964 Georges Thone, Éditeur a Liège

1. Auflage 1969
 2. Auflage 1971

© 1969 Verlag Julius Beltz · Weinheim · Berlin · Basel
 Gesamtherstellung: Offsetdruckerei Julius Beltz, Weinheim/Bergstr.
 Printed in Germany

ISBN: 3 407 12901 7

A 341 (1) | 37. DM

Inhaltsverzeichnis

Vorwort	9
Vorwort des Herausgebers	11
Einführung	13

Erster Teil: Allgemeines

1. Definition und Zielsetzung der Pädagogischen Forschung	21
2. Die Untersuchungsphasen	24
I. Problemstellung	26
II. Die Lösungshypothese	27
III. Der Forschungsplan	28
3. Systematik (Taxonomie) der pädagogischen Forschung	30
I. Die Forschung im allgemeinen	31
II. Die pädagogische Forschung	33

Zweiter Teil: Das Forschungsinstrumentarium

<i>Abschnitt I: Die allgemeinen Techniken</i>	51
Einführung	53
1. Der Fragebogen	54
Einführung	54
I. Aufbau des Fragebogens	56
II. Genauigkeitskontrolle der Antworten und ihre Auswertung	61
2. Die Unterhaltung oder das Interview	63
I. Klassifizierung	63
II. Vorbereitung des Interviews	67
III. Abwicklung des Interviews	68
IV. Notieren der Antworten	69
V. Kritische Analyse der gesammelten Information	69

3. Die „check lists“	70
I. Definition	70
II. Beispiele	70
III. Aufbau der „check lists“	73
IV. Zuverlässigkeit und Objektivität	73
4. Die „rating scales“	74
I. Definition	74
II. Klassifizierung	74
III. Der Aufbau der Beurteilungsskala	79
IV. Schwierigkeit der Beurteilung	79
V. Anwendung	81
VI. Eine spezielle Anwendung	82
5. Die Technik der kritischen Ereignisse (critical incidents)	83
I. Definition	83
II. Anwendungsbeispiele	83
III. Kritik an der Methode	84
6. Andere Techniken	84
I. Protokollierung der Beobachtungen	84
II. Die Zeitanalyse	87
III. Time Sample und Zufallsbeobachtungen	88
IV. Instrumentelle Ausrüstung	88

Abschnitt II: Die Tests und die Spezialtechniken

1. Die Tests	93
I. Definition	93
II. Testaufbau	94
III. Testklassifizierung	104
2. Das Kind und die Gemeinschaft. Die Soziometrie	121
I. Einführung	121
II. Definition der Soziometrie	122
3. Die Einstellungen und die Meinungen	146
Einführung	146
I. Methoden zur Messung der Einstellungen	147
II. Allgemeine Kritik	156
III. Die Einstellungsskalen und die pädagogische Forschung	156
4. Die Untersuchung von Werten	158
5. Die Identifizierung der Interessen	160

6. Die Bewertung der Arbeitsmethode der Schüler	171
I. Die Arbeitsmethode	171
II. Untersuchungsdaten	172
III. Einige Tests	172
7. Die Vorhersage und Bewertung der Leistungsfähigkeit der Lehrer	176
I. Die Kriterien aufgrund der Ergebnisse	178
II. Die Kriterien der mutmaßlichen Erwartung	179
III. Die Kriterien aufgrund der Unterrichtsverfahren oder der Unterrichtsfunktionen	184
IV. Schlußfolgerung	193
8. Die Bewertung der Schulen	194
9. Die Bewertung des sozio-ökonomischen Status	204
I. Bedeutung	204
II. Bewertung	208
III. Schlußfolgerung	215

Dritter Teil: Die Behandlung und die Analyse der Daten

1. Die Inhaltsanalyse	219
I. Definition	219
II. Anwendungsbereiche der Inhaltsanalyse	219
III. Kategorisierungssysteme	220
IV. Beispiele für Inhaltsanalyse	221
2. Die elektronische Datenverarbeitung	234
I. Die Lochkarten und ihre Behandlung	234
II. Elektronische Auswertungsmaschinen für Tests (Test Scoring Machines)	243
III. Elektronische Rechanlagen	248
3. Die Statistik	250
I. Die Versuchspläne	250
II. Die Stichproben	263
III. Die statistische Analyse	268

Anhang

1. Berufsethische Verpflichtungen (Deontologie)	292
2. Das dreidimensionale Modell des Intellekts	295

3. Taxonomie der Erziehungsziele — Erkenntnisbereich	298
I. Erkenntnis	299
II. Geistige Fähigkeiten und Fertigkeiten	300

Vorwort

Das Erziehen ist eine Kunst und wird auch immer eine Kunst bleiben. Diese Aussage steht keineswegs im Widerspruch zu derjenigen, daß ein Teil der Pädagogik in immer zunehmendem Maße „verwissenschaftlicht“ wird. Wir wissen heute sehr wohl, daß eine Kunst sich nur dann richtig ausüben läßt und es dem Künstler gestattet, sich völlig zu entfalten, wenn sie Bezug nehmen kann auf ein Minimum an objektiven Tatsachen, wenn Laborergebnisse mehr oder minder nutzbar gemacht werden oder wenn sie sich auf irgendwelche wissenschaftlichen Gegebenheiten stützt. Die Kunst des Arztes, beispielsweise, ist in diesem Zusammenhang sehr bezeichnend: Was wäre ein Mediziner, wenn er sich allein auf seine Intuition verlassen oder sich weigern würde, Laboruntersuchungen durchführen zu lassen oder von Röntgen-Einrichtungen Gebrauch zu machen, um seine Diagnose zu bestätigen und eine Therapie zu verordnen.

Ebenso verhält es sich auf dem Gebiet der Erziehung. Die Persönlichkeit des Lehrers ist einer der wesentlichen Faktoren in der Erziehungssituation, und seit langem hat man immer wieder festgestellt: „Man unterrichtet nicht das, was man weiß oder zu wissen glaubt, man unterrichtet das, was man ist“. Aber aller Scharfsinn und alle menschliche Intuition — wenngleich sie stets unentbehrlich sind — reichen nicht mehr aus. Biologische, soziologische und psychologische Kenntnisse sind für den Erzieher unentbehrlich, und er muß es lernen, seine Schüler zu beobachten, ihre Arbeit zu überprüfen und die erzielten Resultate objektiv zu beurteilen. Wir dürfen in der gegenwärtigen Situation keine Anstrengung unterlassen, die es der Schule ermöglicht, ihre Aufgaben besser zu erfüllen und ihre Wirksamkeit zu steigern. In diesem Bereich ist eine wissenschaftliche Einstellung unerlässlich, und derjenige, der sich mit experimenteller Pädagogik befaßt, muß über ein Instrumentarium verfügen, das geeignet erscheint, ihm objektive Resultate zu erbringen.

Die französischsprachigen Veröffentlichungen zu diesem Themenkomplex sind spärlich, und man muß auf die angelsächsischen Veröffentlichungen zurückgreifen, um sich die notwendigen Informationen und Untersuchungsergebnisse zu beschaffen. Außer in einigen wenigen Handbüchern, die der

pädagogischen Tatsachenforschung gewidmet sind, findet man nirgendwo die großen Leitlinien einer jeglichen Forschung und das Instrumentarium, das es erlaubt, sie erfolgreich durchzuführen: wie etwa die Ausarbeitung des experimentellen Untersuchungsplans, die zu verwendenden Tests, die Prüfung und Auswertung der Resultate, in einem Werk vereinigt. Das Verdienst von G. De Landsheere ist es, eine neue Gruppierung und Koordinierung der in mehreren Dutzend Werken verstreuten Forschungsergebnisse vorgenommen zu haben. Er stellt die neuesten Arbeiten und die modernsten Methoden zusammen, womit er jedem Erzieher und jedem Forscher wertvolle technische Informationen im Bereich der experimentellen Pädagogik an die Hand gibt. Von nun an steht allen denen, die die Fruchtbarkeit der wissenschaftlichen Einstellung klar erkannt haben, ein Handbuch zur Verfügung, und die Studierenden entdecken leichter den Reichtum und den Nutzen dessen, was viele noch als ein Steckenpferd von Spezialisten betrachten. Dank dem Werk von G. De Landsheere wird nunmehr eine progressive Einführung ermöglicht, und man darf hoffen, daß die zunehmende Zahl der durch diese neuen Aspekte interessierten Lehrer die Entwicklung der Pädagogik zu einem schnelleren Fortschritt führt. Das Buch von G. De Landsheere ist verständlich und für jeden Leser zugänglich, und selbst ein Leser ohne spezielle Vorbildung in den wissenschaftlichen Techniken kann sich seiner mühelos bedienen; es ist in der Tat das Werk eines Forschers und zugleich das eines Pädagogen.

Zu einem Zeitpunkt, in dem überall auf der Welt die pädagogische Forschung in der Entwicklung begriffen ist, fehlte in den französischsprachigen Ländern ein Werk, das es erlaubt, die verschiedenen Arbeitsmethoden und -techniken in einer einheitlichen Darstellung anzubieten. G. De Landsheere hat den Mut besessen, diese Lücke zu schließen. Es sei ihm an dieser Stelle gedankt und unsere Anerkennung ausgesprochen für den bedeutenden Beitrag, den er allen denen anbietet, die sich mit der Zukunft der Erziehung befassen.

G. Mialaret

Vorwort des Herausgebers

Anders als bei uns hat sich im Ausland die Tendenz, Erziehungswissenschaft auf empirischer Basis zu betreiben, immer mehr entwickelt. Vor allem in den englischsprachigen Ländern finden wir heute eine Fülle von auf empirischer Grundlage gewonnenen Ergebnissen und erprobten Verfahren vor, die es erlauben, das weite Feld erzieherischen Bemühens genauer zu untersuchen und uns wertvolle Einsichten in pädagogische Grundphänomene zu gewähren.

In jüngster Zeit schenken auch in Deutschland eine Reihe von Pädagogen der Empirie im Bereich der Erziehungswissenschaften wieder stärker Beachtung. Die Einstellung zur Pädagogik wandelt sich allmählich von einer literarisch-historischen in eine empirisch-methodische Betrachtungsweise. Das Bedürfnis nach Informationen über Methoden und Ergebnisse der empirischen Pädagogik wird immer stärker.

Leider gibt es im Deutschen kein Lehrbuch, das in die wichtigsten empirischen Methoden der pädagogischen Forschung einführt. Eine solche Sammlung wäre aber für alle, die pädagogisch arbeiten, sehr wichtig.

Deshalb hat es die *Gesellschaft zur Förderung pädagogischer Forschung e. V.* übernommen, das vorliegende Buch des belgischen Pädagogen *Gilbert de Landsheere* zu übersetzen und herauszugeben. Dieses Werk versucht, einen Einblick in den augenblicklichen Stand und die wichtigsten Verfahrensweisen pädagogischer Forschung unserer Tage zu vermitteln. Seit seiner ersten Veröffentlichung im Jahre 1962 hat es eine zweite Auflage in französischer Sprache 1964 erfahren, eine dritte ist in Vorbereitung. Die Übersetzung in fünf weitere Sprachen wird zur Zeit vorgenommen. Es wendet sich in erster Linie an Studierende der Pädagogik und an in der Praxis tätige Lehrer, die sich über empirisch-pädagogische Forschungsmethoden orientieren wollen.

Die hier vorliegende deutsche Übersetzung und Bearbeitung lehnt sich so eng wie möglich an den französischen Originaltext an, sie berücksichtigt außerdem bereits die Korrekturen, Ergänzungen und Texterweiterungen, die der Verfasser für die dritte Auflage des Werkes vorgesehen hat. Be-

stimmte Testverfahren, die auf eine amerikanische oder französische Population geeicht waren, wurden durch bereits ins Deutsche übertragene und auf deutsche Verhältnisse geeichte Verfahren ersetzt. Außerdem wurde überall dort, wo gleichwertige deutsche Tests vorliegen, auf dieselben hingewiesen. Eine Reihe von Berechnungsformeln im dritten Teil der Arbeit wurde auf in Deutschland gebräuchliche Formulierungen umgestellt. Dieser Teil wurde auch im Hinblick auf die übliche Terminologie etwas stärker bearbeitet. Die nun in deutscher Sprache vorliegende Fassung des Werkes wurde mit dem Autor abschließend persönlich überprüft und fand seine Zustimmung.

Wir möchten an dieser Stelle Herrn Professor Dr. *Hans Jörg* von der Pädagogischen Hochschule Saarbrücken für die Übersetzung und deutsche Bearbeitung des vorliegenden Werkes und Herrn Dr. *Ulrich Raatz* vom Deutschen Institut für Internationale Pädagogische Forschung in Frankfurt (Main) für die kritische Durchsicht der Übersetzung und die Bearbeitung des statistischen Teiles unseren Dank aussprechen.

Wir hoffen, mit der Herausgabe dieses Buches einen Beitrag zur dringend notwendigen Aktivierung der empirischen Pädagogik auch bei uns geleistet zu haben.

Im Auftrag der
Gesellschaft zur Förderung Pädagogischer Forschung e. V.
herausgegeben:

Prof. Dr. J. P. Ruppert

Einführung

1. In einem nicht genau abzugrenzenden Maße ist die Pädagogik noch eine Mischung von Wissenschaft, von Kunst und — um das harte Wort von Lauwerys und Travers¹ aufzugreifen — von volkstümlicher Tradition.

Eine reine Wissenschaft wird die Pädagogik niemals werden, denn die Komplexität der menschlichen Natur entzieht sich dem Determinismus, „dieser Faktorenordnung, in der jedes Element von gewissen anderen Elementen in einer solchen Weise abhängt, daß man es mit Sicherheit voraussagen, erzeugen oder verhindern kann, je nachdem, wie man diese kennt, wie man sie erzeugt oder sie verhindert“².

Aber diese Einschränkung, die übrigens für die Zukunft unserer Spezies Mensch erfreulich ist, rechtfertigt keineswegs ein blindes Sichhingeben an die Zufälligkeiten der Inspiration, an den „gesunden Menschenverstand“ und an die Tradition. Überall, wo die Wissenschaft unsere erzieherische Tätigkeit, d. h. unser einmütiges Bemühen um eine bessere Entfaltung des Kindes wirksam leiten kann, müssen wir sie zu Rate ziehen.

Nach einigen aufsehenerregenden Fortschritten, die übrigens mit den ersten Erfolgen der experimentellen Psychologie zusammenfallen, erfuhr die Pädagogische Forschung eine Phase der Zersplitterung, mehr noch der Zusammenhanglosigkeit.

Oftmals beschränkt auf systematische Studien über die Anwendung von methodologischen Verfahren oder Regeln, deren Gültigkeit selbst nicht angezweifelt worden war, hat sich die Pädagogische Forschung auf sich selbst zurückgezogen und sich somit selbst das Wasser abgegraben. Losgelöst von den übrigen Humanwissenschaften, hat die Pädagogik kaum eine Bedeutung. Es gibt keinen „*puer scholasticus*“, es gibt nur Kinder in ihrer Ganzheit, mit ihrem ganzen Reichtum an Intelligenz und Persönlichkeit, mit dem

1 J. A. Lauwerys, *The Philosophical Approach to Comparative Education, Thoughts on Comparative Education*, Festschrift für Pedro Rossello, Hamburg, Unesco Institute for Education, 1959, S. 29.

R. M. Travers, *An Introduction to Educational Research*, New York, 1958, S. 13.

2 A. Lalande, *Vocabulaire de la philosophie*, Paris, P.U.F., 1956, 7. Aufl. S. 222.

ganzen Bezugssystem, das sie — bewußt oder unbewußt — von draußen in die Schule hineinragen.

Seit einigen Jahren jedoch bemüht man sich um eine Neubelebung der Pädagogischen Forschung. Diese läuft parallel mit dem gegenwärtigen Aufschwung auf den Gebieten der Psychologie, Biologie, Soziologie, der Kulturanthropologie und der Statistik. Obwohl sich die Erkenntnisse — so verschieden wie wertvoll sie auch sein mögen — in der Erziehungswissenschaft anhäufen, ist noch keine moderne allgemeine Synthese daraus hervorgegangen. Bereits im Jahre 1955 drückte T.A. Lamke sein Bedauern hierüber aus³.

Die Weiterentwicklung der wissenschaftlichen Pädagogik stößt auf zwei wesentliche Schwierigkeiten:

1. Es mangelt ihr an Forschern: Innerhalb des sehr weit gespannten Wirkungsfeldes, das sie anbietet, wird recht wenig wirklich nutzbar gemacht. Wie hoch ist in jedem Land die Zahl der pädagogischen Forschungszentren? Wieviele europäische Universitäten sind in ausreichendem Maße mit entsprechenden wissenschaftlichen Mitarbeitern und Materialien ausgerüstet, um die umfangreichen Forschungsprogramme abzuwickeln, die man doch in unserer Zeit übereinstimmend als eine der besten Investitionen ansieht? Nun, besonders in der Grundlagenforschung ist die Zeit des Dilettantismus vorbei.

2. Der pädagogische Fortschritt krankt an dem mangelnden gegenseitigen Gedankenaustausch zwischen Forschern und praktizierenden Pädagogen, den Konsumenten der Forschungs-Produkte.

Solange die Lehrer noch keine umfassende Einführung in die Forschungsarbeit erhalten haben, kann man von ihnen nicht erwarten, daß sie ihre täglichen pädagogischen Techniken durch wirkliches Experimentieren überprüfen bzw. verbessern. Folglich, welchen Sinn hat die wissenschaftliche Forschung, wenn nicht jenen eines Spiels für Theoretiker?

Nicht nur die Ausbildung der Lehrer bedarf also einer Revision, sie müssen auch in den Stand versetzt werden, sich während ihrer gesamten Berufstätigkeit über neue pädagogische Erkenntnisse auf dem laufenden zu halten,

3 „Wir warten noch auf einen Kopernikus, der unsere Erklärungen vereinfacht, auf einen Newton, der einige Grundprinzipien formuliert, auf die sich unser Werk aufbauen könnte, für eine gewisse Zeit zumindest auf einen Mendel, der die Unmenge von zusammenhängend erscheinenden Daten ordnet, einen Descartes, einen Leibniz, einen Fisher, die uns mit mathematischen Modellen von der Wirklichkeit versorgen, so, wie wir sie sehen, Modelle, die speziell für unsere Arbeit konstruiert sind und nicht unbedingt für andere Disziplinen.“
(T.A. Lamke, Review of Educational Research, Nr. 3, Juni 1955, S. 192.)

und sie müssen ermutigt werden, diese Erkenntnisse in die Wirklichkeit umzusetzen⁴.

Ohne die Philosophie als Wegweiser ist die Pädagogische Forschung nur seelenlose Technik.

Man erzieht im Hinblick auf ein Ziel. Zweifellos verschaffen uns die Erfahrungen aus der Erziehungspraxis die Mittel und Wege, auf das Kind einzuwirken sowie die Möglichkeit, unsere Handlungen zu überwachen, aber sie geben uns keinen Aufschluß darüber, auf welches Ziel hin wir handeln müssen.

Der Philosophie fällt demnach die Aufgabe zu, der Erziehung ein Ziel zu weisen und die angewandten Mittel zu koordinieren.

Wir haben es bereits gesagt, daß die Persönlichkeit des Kindes eine komplexe Einheit darstellt, in der alle Komponenten eng miteinander verknüpft sind. Bevor der Forscher einen Versuch unternimmt, muß er eine möglichst vollkommene Kenntnis von den Kindern oder den Erziehern besitzen, auf die sich das Experiment erstreckt; er muß aufmerksam auf alles achten, was auf sie einwirkt und in ihnen vorgeht, und er muß rasch auch eine ungünstige Reaktion erklären können.

In der Tat wertet die moderne Erziehungswissenschaft in immer größerem Ausmaß Quellen und Erkenntnisse aus, die ihr die anderen Wissenschaften über den Menschen bieten. So erfährt man z. B. in diesem Buch darüber, wie sehr die Pädagogische Forschung von den Sozialwissenschaften beeinflusst wird. Natürlich kann der Fachmann in der Erziehungsforschung nicht die Absicht verfolgen, sich sämtlicher verfeinerter Techniken aller anderen Humanwissenschaften zu bedienen — wir warnen ihn häufig vor der Gefahr einer nicht erkannten Inkompetenz —, jedoch muß er über die Existenz der wichtigsten Forschungsmittel unterrichtet sein und ihre Möglichkeiten sowie ihre Handhabung kennen. Im übrigen ist es gut, wenn alle Erzieher sich einiger elementarer Forschungsmittel zu bedienen wissen.

4 A. Die Forschungszentren sollten regelmäßig Veröffentlichungen von Synthesen herausbringen, die auf die Bedürfnisse des Unterrichtswesens abgestimmt sind.
B. Den Erziehern, die ein Forschungsprogramm oder ein Experiment ausführen, sollte eine ständige Beratergruppe zur Verfügung stehen.
C. Fortbildungslehrgänge sollten in den normalen Dienstplan der Lehrer aufgenommen werden, und denjenigen, die sich besonders verdient gemacht haben, sollte die Möglichkeit zu Ergänzungsstudien und Informationsreisen geboten werden.

Man kann gar nicht genug die interdisziplinäre Team-Arbeit herausstreichen; sie drängt sich mehr und mehr auf aufgrund ihrer Resultate, ihrer Sicherheit sowie der Erweiterung des Blickfeldes, die sie erlaubt.

Aber wenngleich es unerlässlich ist, sich der ungeheuren Komplexität der Phänomene bewußt zu werden, so sollte man doch nicht einem sterilen Perfektionismus verfallen. Die Sorge um eine allgemeine Wirksamkeit muß dem Wunsch nach einer idealen Lösung vorangehen. Man wird niemals allen nur möglichen Komponenten Rechnung tragen können. Ein gewisser Empirismus ist daher unvermeidlich; dieser bedarf aber der Klärung. Kluge Voraussicht und behutsames Vorgehen werden manche Unkenntnisse kompensieren.

2. Ziel des vorliegenden Buches ist es, diejenigen, die sich mit der Pädagogischen Forschung befassen wollen, mit ihren Grundideen, Methoden, ihrem Forschungsinstrumentarium sowie ihrem Wesen vertraut zu machen. Unsere Vorschläge in diesem Buch sind im wesentlichen praktischer Natur. Im ersten Teil geben wir eine Definition der Erziehungsforschung, ihrer Zielsetzung, der wichtigsten Entwicklungsstufen und entwerfen eine Taxonomie. Der zweite Teil ist dem Forschungsinstrumentarium gewidmet. Der erste Abschnitt analysiert die allgemeinen Forschungstechniken (oder Schätzinstrumente): Fragebogen, Interviews, rating scales, usw. Der zweite Abschnitt behandelt Tests sowie einige spezielle Anwendungstechniken und -gebiete, die für den Pädagogen besonders wichtig sind: Untersuchung der Arbeitsweise der Schüler, Beurteilung der Lehrer und der Schulen sowie Feststellung des sozio-ökonomischen Status.

Es ist sicher, daß eine bloße Kenntnis der Meßinstrumente und ihrer Handhabung nicht ausreicht. Man muß darüber hinaus und vor allem die Informationen, die sie erbringen, richtig interpretieren können.

Die wissenschaftliche Kenntnis der Phänomene stellt den Rahmen, das logische Bezugssystem dar, innerhalb dessen die Messung ihre Bedeutung und ihren Wert erlangt. So gesehen, findet sie nicht nur ihre praktische Rechtfertigung, indem sie bei der Lösung von echten Problemen hilft, sondern sie trägt gleichzeitig zum wissenschaftlichen Fortschritt bei, indem sie alte oder neue Hypothesen von Fall zu Fall bestätigt oder entkräftet.

Neben der quantitativen Analyse, die im Mittelpunkt unserer Arbeit steht, gibt es die qualitative oder klinische Analyse, die dem Gegenstande der Humanwissenschaften — nämlich dem Menschen — angemessen zu feinen Unterscheidungen führt, notwendige Synthesen erstellt, und zwar dort, wo die Zahlen Gefahr laufen, die Wirklichkeit zu entstellen, indem sie sie schematisieren und verfälschen.

Der gesamte Bereich der Pädagogik und der der Psychologie zeichnen sich daher auf den Seiten des vorliegenden Bandes filigranartig ab.

Aus dieser Sicht betrachtet ist unser Werk recht bescheiden. Es beschränkt sich auf die empirischen Instrumente, indem es nur dann den theoretischen Aspekt anschnidet, wenn dies zu deren Darstellung und Anwendung notwendig erscheint.

Der dritte und letzte Abschnitt ist der Verarbeitung der Forschungsdaten gewidmet: Inhaltsanalyse und Statistik einschließlich der Versuchspläne (*Research design*) sowie der elektronischen Datenverarbeitung. Wir glauben, auf letzteres Verfahren ausführlicher eingehen zu müssen, im Hinblick auf die große Bedeutung, die die Lochkarten und die elektrischen Datenverarbeitungsgeräte (Computer) in der Forschung von heute haben.

Zusammenfassend läßt sich sagen, daß im vorliegenden Buch der Versuch unternommen wird, Forschungsergebnisse und Tatsachen, die unseres Wissens bis heute nur verstreut in zahlreichen Büchern, Sonderdrucken und schwer zugänglichen wissenschaftlichen Veröffentlichungen aufzufinden sind, in einem Werk zusammenzufassen und zu ordnen. Wir haben alle historischen Aspekte ausgeschaltet, uns grundsätzlich nur an die Hilfsmittel unserer Tage gehalten und in einigen Fällen versucht, auch diejenigen zu berücksichtigen, deren wir uns in Zukunft bedienen werden.

Wie man auch immer bei einem solchen Versuch vorgehen mag, man umfaßt zu viel und zu wenig; zu viel, weil man eine große Anzahl von Techniken zitiert, die der Erzieher nicht alle beherrschen kann; zu wenig, weil die notwendige Kürze der Darstellung der Gefahr sinnverändernder Vereinfachungen ausgesetzt ist. Jede besondere Technik erfordert ein sorgfältiges Studium, wenn nicht sogar eine Spezialisierung. Wir erwähnen darum jedes Mal einige grundlegende Veröffentlichungen, die zum Weiterstudium geeignet erscheinen.

Bedarf es noch der Erwähnung, wie sehr diese Darstellung auch der Arbeit anderer verpflichtet ist? Aus meinen Zitaten wird man bald ersehen, von welchen Werken ich am meisten Anregung empfang.

Jedoch steht neben dem Einfluß des Schrifttums auch und vor allem derjenige einer Reihe von Persönlichkeiten, in erster Linie der meines Lehrers, Professor Dr. A. Clausse, dem — da er das Labyrinth der Techniken mit solcher Klarheit beherrscht, was man ihm einstimmig zuerkennt — diese Arbeit manche neue Betrachtungsweise verdankt.

Groß ist meine Dankbarkeit auch gegenüber Prof. Dr. A. Dubuisson-Brouha, der Präsidentin des Instituts für Psychologie und Erziehungswis-

senschaften der Universität Lüttich, die mich mit wertvollen Anregungen von Anbeginn meiner Arbeiten an unterstützt hat.

Weiterhin danke ich den Herren Dr. R. A. Osterrieth, Professor an der freien Universität Brüssel, Dr. R. Pasquasy, Professor an der Hochschule für Pädagogische Wissenschaften an der Universität Lüttich, Dr. P. Minon, Professor der Fakultäten Notre Dame de la Paix in Namur, F. Hotyat, Direktor des *Centre de Travaux* der Hochschule für Pädagogik der Provinz Hennegau, sowie Ingenieur P. Remouchamps von IBM Belgien, die lebenswürdigerweise die Überprüfung der sich auf ihr Spezialgebiet beziehenden Kapitel übernommen haben und mir dabei manche Anregung gaben.

Nicht zuletzt danke ich auch Herrn Dr. M. G. Mialaret, Professor an der Universität Caen und Präsident der internationalen Gesellschaft für Experimentelle Pädagogik französischsprachender Länder, dessen Untersuchungen für mich wegweisend waren und der mir die Ehre erwies, das Vorwort zu diesem Werk zu schreiben.

Erster Teil: Allgemeines

1. Definition und Zielsetzung der Pädagogischen Forschung

A. S. Barr definiert die Pädagogische Forschung im weiteren Sinne als „jedes systematische Bemühen um Verstehen, hervorgerufen durch ein Bedürfnis oder eine Schwierigkeit, die man erkannt hat und die sich auf die Erforschung eines komplexen Phänomens beziehen und eine Bedeutung besitzen, die über rein persönliche und unmittelbare Belange hinausreicht, wobei das Problem in Form einer Hypothese gestellt ist“¹.

Diese Definition hat den Vorzug, zwischen den Untersuchungen des Forschers und den tastenden und von den jeweiligen Umständen abhängigen Versuchen des Praktikers deutlich zu unterscheiden. Andererseits weist sie den Mangel auf, nicht spezifisch für die Pädagogik zu sein, denn sie liefert uns keine genauen Angaben über den Inhalt der Studien oder die verfolgten Ziele.

In der Tat interessiert sich die Pädagogische Forschung für jede Tätigkeit, die mit einem erzieherischen Vorgang in Zusammenhang steht und folgende Punkte umfaßt²:

1. Die experimentellen Untersuchungen, die das Ziel haben, bestehende bzw. vorgeschlagene Methoden und Verfahrensweisen einer kritischen Prüfung zu unterziehen. Wir kommen auf den Seiten xx und xx auf die Definition des „Experimentierens“ zurück, ein Wort, dem wir einen enger gefaßten Sinn geben, als dies R. Buyse tut, der die experimentelle Methode als „ein Bemühen um das Messen der pädagogischen Tatbestände, ein Erforschen ihrer Bedingungen und ein Bestimmen ihrer Gesetzmäßigkeiten“³ definiert. Wir möchten an dieser Stelle vor allem herausstellen, daß die Experimentelle Pädagogik nur ein Teil der Pädagogischen Forschung ist;
2. Die Umfragen und allgemeiner die Beobachtungen, die sich auf erzieherische Tatbestände erstrecken;
3. Die theoretischen, historischen, philosophischen bzw. interdisziplinären Beiträge, die einen allgemeinen Untersuchungsrahmen bilden sollen;
4. Kritische Durchsicht der Forschungsliteratur;

1 Siehe C. W. Harris, *Encyclopedia of Educational Research*, New York, Macmillan, 1960, 3. Aufl., S. 1166.

2 Siehe in diesem Zusammenhang: N. A. Fattu, *A Survey of Educational Research at Selected Universities*, in F. W. Banghardt, Edit., *First Annual Phi Delta Kappa Symposium on Educational Research*, Bloomington Ind., Phi Delta Kappa, 1960, S. 13–14.

3 R. Buyse *L'expérimentation en pédagogie*, Brüssel, Lamertin, 1935. Dieses Werk wird sehr empfohlen.

5. „Die Angewandte Forschung, konzentriert auf lokale Verfahrensweisen oder Programme, die sowohl dazu dienen, das Interesse für ein vertieftes Studium zu wecken als auch gleichzeitig das Lehrpersonal heranzubilden und gewisse unmittelbare Probleme zu klären. Diese Aktivität tendiert dahin, sich zur operationalen oder logistischen Forschung zu entwickeln...“⁴

Zu den Faktoren, die im Erziehungsprozeß miteinander in Beziehung treten, gehören einerseits die Kinder, die Eltern, die Lehrer und sonstigen Miterzieher, die direkt und willentlich in den Erziehungsprozeß eingreifen, und zum anderen das menschliche Milieu im weiteren Sinne sowie die natürliche Umwelt (Tiere, Pflanzen, Gesteine), der allgemeine Rahmen also, dessen Einfluß man nicht unterschätzen darf, dessen Studium letztlich gleichbedeutend mit der Erforschung des Universums ist.

Zur leichteren Handhabung der Probleme, mit denen sich die Pädagogische Forschung beschäftigt, muß sie ihren Blickwinkel reduzieren und ihre Untersuchungen auf die unmittelbaren Faktoren konzentrieren.

Man kann der Erziehungsforschung *fünf allgemeine Zielsetzungen* zuordnen, auf die der Pädagoge übrigens keineswegs den Anspruch erhebt, sie alleine zu erreichen. Er kann sich ihrer nur in enger Zusammenarbeit mit allen anderen Fachkundigen der Humanwissenschaften, Mathematik und Naturwissenschaften nähern: Mediziner, Biologe, Psychologe, Soziologe, Anthropologe, Wirtschaftswissenschaftler, Historiker, Philosoph, Statistiker usw.

Ziel I: Wissen um den Schüler als Kind und als Lernenden. – Von daher die Notwendigkeit der Überprüfung:

1. Wie das Kind sich verhält: Körperlicher und geistiger Gesundheitszustand, Wachstum und Entwicklung (learning), normale Kinder, psychisch und physisch gestörte Kinder;
2. Wie andere das Kind sehen: Eltern, Geschwister, sonstige Kinder, Lehrer, andere Erwachsene;
3. Wie das Kind sich selbst sieht.

Ziel II: Kenntnis der Erzieher und des Unterrichts

A. Die Erzieher:

1. Umwelt und menschliches Milieu: Anpassung an die Umwelt und an das

⁴ N. Fatu, op. cit. Wir definieren auf S. 30–50 die verwendeten Ausdrücke (Übersetzung des Originalzitates).

menschliche Milieu, Einwirkung auf die Umwelt und auf das menschliche Milieu;

2. Die Eltern: Die Erziehung in der Familie, die Vorbereitung der Eltern auf ihre Aufgabe als Erzieher, die Beziehung von Elternhaus und Schule;
3. Die Lehrer: Auswahl, Ausbildung, Status, Psychologie des Lehrers.

B. Der Unterricht:

1. Psychologie des Lehrens;
2. Methodologie;
3. Menschliche und materielle Hilfen;
4. Die Prüfungen: Dozimologie;
5. Guidance und Counseling (Schulberatung).

Ziel III: Kenntnis der zu vermittelnden Bildungsgüter

- A. Zielsetzungen
- B. Mittel: Unterrichtsstoff
- C. Stoffverteilung: Lehrpläne

Ziel IV: Kenntnis des Erziehungssystems

- A. Organisation: Untersuchung der Gesamtstruktur und spezieller Erziehungseinrichtungen (institutional research).
- B. Administration: Die pädagogische Hierarchie; die Schulverwaltung im eigentlichen Sinne; Schulgesetzgebung; Schulfinanzen: Kostenträger, Haushaltsplan, Unterhaltungskosten, Schulgebäude und -einrichtung; Schulstatistik: demographische Vorausplanung usw.

Ziel V: Kenntnis der durch andere herbeigeführten Lösungen.

- A. In der Vergangenheit: Geschichte der Pädagogik.
- B. In der Gegenwart: Studium der ausländischen Schulsysteme (Auslandspädagogik); vergleichende Untersuchungen: Vergleichende Erziehungswissenschaft.

Nach diesem ersten Versuch einer Definition der pädagogischen Forschung wollen wir ihre wichtigsten Entwicklungsstufen untersuchen, um ihr Wesen besser erfassen zu können.

2. Die Untersuchungsphasen

Es ist gleich, ob es sich um ein Beobachten oder Experimentieren handelt, die Forschung entsteht aus dem Vorhandensein eines Problems, das zu klären oder zu lösen ist.

Tatsächlich besteht zwischen dem *Problem-solving* (Problemlösungsverhalten) des täglichen Lebens und der Forschung kein absoluter Widerspruch: Es differieren nur der Grad des Bewußtwerdens —, die Bemühungen um Systematisierung und die Strenge der Verallgemeinerungen.

Kerschensteiner unterschied vier Phasen des gedanklichen Vorgehens^{5a}:

1. „Die aus dem Beobachten einer Tatsache sich selbst aufdrängenden Schwierigkeiten und Fragen an den Beobachter (Fragestellung).
2. Die vorläufigen Antworten, die gegeben werden, also die Vermutungen für die Lösung der Frage (Hypothesen).
3. Die Verfolgung der vermuteten Lösungen mit Hilfe des Wissens von anderen Tatsachen, Regeln, Gesetzen usw. durch Aufsuchen ihrer Konsequenzen für die Tatsachen selbst und für ihre Beziehungen untereinander (Verifizierung).
4. Die Bestätigung oder Nichtbestätigung der endgültig akzeptierten Schlußfolgerung durch Zusammenstellung mit anderen Tatsachen (Entscheidung).“

Oftmals wickelt sich dieser Prozeß so schnell ab, daß wir uns nicht aller Einzelheiten bewußt werden. So hat Dohrmann⁶ beispielsweise gezeigt, daß sich der Lesevorgang in der Vielfalt von Lösungsschritten vollzieht: sobald wir von den ersten Elementen eines Textes Kenntnis nehmen — mitunter genügt schon das Lesen des Titels — beginnen wir unbewußt, Hypothesen über den Inhalt aufzustellen, und das sinnentnehmende Lesen der Worte dient ebenso der Bestätigung unserer Vermutungen, wie der Erzeugung neuer Hypothesen. Man geht dabei von folgendem Schema aus:

5a „Sinn und Wert des naturwissenschaftlichen Unterrichts“. zit. bei:

6 P. Dohrmann, *Gedankliches Lesen, Theorie der Ganzheitsmethode*, Hannover, Hahnsche Buchhandlung, s. d. S. 52.

G. De. Landsheere, *La connaissance de la lettre, condition de la lecture globale* (Cahiers de Pédagogie et d'Orientation professionnelle, Universität Lüttich, Jan. 1960).

Erstes Wort oder Wortgruppe	Zweites Wort oder Wortgruppe	Drittes Wort oder Wortgruppe	Viertes Wort oder Wortgruppe
Fragestellung			
Hypothese	Fragestellung		
Verifizierung	Hypothese	Fragestellung	
Entscheidung	Verifizierung	Hypothese	Fragestellung
	Entscheidung	Verifizierung	Hypothese
		Entscheidung	Verifizierung
			Entscheidung

Die Forschung, die in Wirklichkeit nichts weiter ist als ein sich über einen längeren Zeitraum erstreckendes umfassendes und systematisches gedankliches Bemühen, impliziert die Existenz dieser vier allgemeinen Phasen. Das einfache Zusammentragen von Informationen oder die tabellarische Erfassung von statistischen Daten, die nicht von einer Hypothese geleitet werden und zu keiner Schlußfolgerung führen, können daher nicht als wissenschaftliche Forschung bezeichnet werden.

Aufzählen, selbst in sachkundiger Form, ist nicht Denken. Die Pädagogik hat unter der Verwechslung dieser beiden Vorgänge erheblich gelitten.

In „*How we think*“ entwickelt J. Dewey das Schema Kerschensteiners und ermöglicht dadurch eine verfeinerte Definition des Forschungsprozesses. H. Roth schreibt hierzu:

„Im Anschluß an Dewey hat die amerikanische Pädagogik ein Schema des Erkenntnisprozesses in der empirischen Forschung entwickelt, das als Rahmen immer noch Gültigkeit hat:

1. Untersuchungen beginnen meist aufgrund einer erlebten Not, eines Wunsches, eines besonderen Erfolges oder Mißerfolges. Alle diese Umstände werfen Fragen auf, denen man nachgeht.
2. Das Problem wird genauer formuliert, die Aufgabe wird ins Auge gefaßt. Es handelt sich darum, sicher zu werden, was man eigentlich will, was das Ziel ist, worauf die Untersuchung Antwort geben soll. Nun taucht ein ganzer Hof von Fragen auf, aber sie kommen langsam in eine hierarchische Ordnung nach dem Grad ihrer Wichtigkeit.
3. Nun werden Beobachtungen gesammelt, die zur Lösung der Frage beitragen können, Literatur wird eingesehen, Experten werden befragt, Quellen erkundet, Ratschläge eingeholt, Notizen gemacht, Auszüge angelegt, graphische Darstellungen versucht.

7 Siehe H. Roth in: H. Heckel, E. Lemberg, H. Roth, W. Schultze, F. Süllwold, *Pädagogische Forschung und pädagogische Praxis*, Heidelberg, Quelle und Meyer, 1958, S. 44–45.

4. Dabei tauchen nun schon begründete Vermutungen auf, Schätzungen, Erklärungsentwürfe, Hypothesen. Es handelt sich um die ersten Arbeitshypothesen. Die möglichen Ergebnisse werden überschlagen.
5. Strenge Überprüfung der Hypothese. Planung eines Experiments.
 - a) Vorversuch an einer kleinen ausgewählten Versuchspersonenzahl.
 - b) Versuch an einer repräsentativen Stichprobe.
6. Logische oder statistische Verarbeitung der Ergebnisse. Verfolgen der Konsequenzen. Sichtbarmachen des gesamten Bezugfeldes.
7. Versuch, das Ergebnis in einem allgemeinen Prinzip zu formulieren.
8. Einreihung der Ergebnisse in das gesonderte Ganze unseres Wissens. Grad der Allgemeingültigkeit.
9. Verfügbarmachen für die Praxis.“

Zu dieser Art des Vorgehens bemerkt H. Roth: „Diese Stufenreihe täuscht einen gradlinigen Fortschritt vor, den es in der Praxis nicht gibt. Ihr Weg gleicht mehr einer Zickzackkurve ...“

I. Problemstellung

Grundsätzlich kann man unterstellen, daß eine Untersuchung sich stets auf eine allgemeine Theorie stützen sollte. Wenngleich auch mitunter anfechtbar, ist sie bei der Untersuchung doch richtunggebend, und ohne sie läuft man Gefahr, die Grenzen der einfachen Technik nicht zu überwinden. Travers⁸ stellt mit Recht fest, daß bis heute die zusammenhängenden Theorien, selbst wenn sie sich noch unvollkommen auf die Erfahrung stützen, mehr zur Förderung der Erziehungswissenschaft beigetragen haben als die Vielzahl verschiedenartigster zusammenhangloser Untersuchungen. Trotz der Schwächen, die uns heute offensichtlich zu sein scheinen, haben die von Herbart und Dewey vorgezeichneten Grundideen zu einem erheblichen pädagogischen Fortschritt geführt.

Kein Forscher darf sich an die Lösung eines Problems heranwagen, ohne es vorher in seinen Gesamtzusammenhang einzuordnen und seine aktuelle Bedeutung klar hervorzuheben. Ohne diese Vorarbeit können der Nutzen und die Zweckmäßigkeit der Forschung nicht eindeutig hervortreten.

Wenn auch die Ausgangssituation übersichtlich und zuverlässig sein muß, so darf sich der junge Forscher dennoch nicht in übertriebene und peinlich genaue Bibliographien verlieren die ihn Tag um Tag vom Handeln abhalten; sonst läuft er Gefahr, nie schöpferisch tätig zu werden.

⁸ Op. cit. S. 7–8.

Das gewählte Problem wird mit der größten Sorgfalt dargelegt, jedes Wort trägt einen bestimmten Sinn ohne irgendwelche Zweideutigkeiten.

Das Bemühen um Information und Synthese zwecks Definition eines Arbeitsgegenstandes und dessen Eingliederung in einen sorgfältig gezielten Zusammenhang erlaubt es, etlichen Gefahren zu begegnen:

1. Dem falschen Problem, bei dem eine Wirkung an eine Ursache gebunden ist, die ihr fremd ist.
2. Dem zu weit gefaßten und infolgedessen mit den dem Studierenden und dem Forscher gewöhnlich zur Verfügung stehenden begrenzten Mitteln unmöglich zu behandelnden Problem.
3. Dem bereits gelösten Problem.

Bei den Humanwissenschaften sind die Lösungen jedoch selten definitiv und vollständig; die Wiederholung von Untersuchungen unter neuen Bedingungen führt oftmals zu einer wesentlichen Vervollkommnung der Erkenntnisse.

II. Die Lösungshypothese

Ohne eine leitende Hypothese verfällt ein Großteil der Untersuchungen einer sterilen Anhäufung von Daten und Informationen.

Im allgemeinen wird die Lösungshypothese nach einer nur vorläufigen Festlegung des Untersuchungsverfahrens unter mehreren ausgewählt. Oftmals wandelt sie sich im Verlauf der Untersuchung, sei es, daß sie umfassender wird oder sich präzisiert.

Der Reifungsprozeß des Gedankengutes verläuft langsam; jeder Schritt muß zunächst Gegenstand genauer Verifizierungen sein. Manch einer sieht in diesem Anfangsstadium nur ein lästiges Vorspiel, das er in seiner Ungeduld verkürzen möchte. In Wirklichkeit ist dies aber einer der entscheidenden Momente der Untersuchung. Die Intelligenz des Untersuchers, seine schöpferischen Fähigkeiten, sein Bildungsniveau und seine Erfahrung spielen dabei eine erhebliche Rolle.

Doch ist es zugleich eine undankbare, mitunter zeitraubende Angelegenheit, die den Eindruck eines „Stillstands“ entstehen läßt.

III. Der Forschungsplan

Aus der Lösungshypothese ergeben sich die Hauptlinien des Forschungsplanes — ein anderer Aspekt, der zu häufig mit vagen Abschätzungen behaftet ist.

Es ist selten möglich, einen Forschungsplan und einen Zeitplan restlos zu befolgen; sie stellen nichtsdestoweniger wertvolle Wegweiser dar und sind im übrigen unerlässlich für die Zusammenarbeit einer Forschungsgruppe.

Bevor man eine Untersuchung von einigem Ausmaß in Angriff nimmt, erweist es sich als sehr nützlich, einen Plan aufzustellen, der kurz und bündig abgefaßt ist und folgende Punkte berücksichtigen sollte:

Forschungsprojekt

1. Problemstellung
2. Inwieweit könnte die Lösung des Problems zum allgemeinen Fortschritt in der theoretischen und praktischen Pädagogik beitragen;
3. Lösungshypothese (n);
4. Definition der wichtigsten verwendeten Ausdrücke;
5. Erste Zusammenstellung der Literatur;
6. Beobachtungs- bzw. Untersuchungsstufen;
 - a) Was erforscht man und wie geht man dabei vor?
 - b) Welche Daten will man zusammenstellen?
 - c) Wie erfaßt man diese?
 - d) Wie verifiziert man ihre Validität?
 - e) Wie analysiert man sie?
7. Zeitplan.

Für bedeutende Untersuchungen und speziell für Auftragsforschung⁹ muß der Zeitplan häufig sehr präzise sein: Erreichen des gesteckten Zieles zu einem festgelegten Zeitpunkt; die Mitarbeiter können sich nur während einer bestimmten Zeit vom Dienst dispensieren lassen; internationale Koordinierung usw. Seit einigen Jahren ist die Ausarbeitung eines Zeitplans geradezu zu einer Wissenschaft geworden. Man verwendet die Methode des kritischen Weges, allgemein bekannt unter der amerikanischen Abkürzung P.E.R.T. (Program Evaluation Research Task).

„Das P.E.R.T. lenkt die Aufmerksamkeit auf die ‚kritischen Operationen‘, das sind solche, bei denen jede Verzögerung bei einer von ihnen sich gänzlich auf den abschließenden Zeitpunkt der Fertigstellung auswirkt. Die Folge der ‚kritischen Vorgänge‘ bestimmt den ‚kritischen Weg‘. Für die anderen Vorgänge bestimmt die Methode den Spielraum. Der Beginn einer jeden Operation wird festgestellt.“¹⁰ Man

⁹ Siehe Definition bei der Taxonomie.

¹⁰ Siehe Bull-Information, Nr. 10, 1963, S. 1–5.

Siehe auch: D. L. Cook, *P.E.R.T., Applications in Education*, Washington, Office of Education, 1966.

schließt mit einer graphischen Darstellung, die die Dauer eines jeden Untersuchungsprozesses (Untersuchungszeitdauer) anzeigt sowie die Reihenfolge, in der die Untersuchungen abzuwickeln sind (Zwang).

Es sei bemerkt, daß viele Institutionen und Stiftungen die Genehmigung für die finanzielle Unterstützung eines Projektes davon abhängig machen, ob die soeben erwähnten Punkte beachtet wurden.

Zur Erläuterung folgt nun ein Forschungsvorschlag, der den aktuellen Anforderungen des Erziehungsministeriums der Vereinigten Staaten entspricht.

*Untersuchung der beidohrig unterschiedlichen Reizung des Gehörs beim Unterrichten von lebenden Sprachen.*¹¹ (Anm. des Übersetzers: Hören der eigenen Stimme und der Stimme des Lehrers im Sprachlaborunterricht.)

Problemstellung: Diese Studie versucht zu klären, ob die beidohrig unterschiedliche Reizung des Gehörs (stimulation binaurale) beim Unterrichten von Fremdsprachen zur Verbesserung der Aussprache wirksam angewendet werden kann.

Hauptziel: Es ist zu bestimmen, inwieweit die Fähigkeit des Schülers, Unterschiede in der Aussprache zu vergleichen, durch den Tatbestand beeinflusst wird, daß er gleichzeitig mit dem einen Ohr seine eigene Stimme und mit dem anderen die des Lehrers hört.

Plan: Zwei Gruppen von 20 Studenten im ersten Jahr, die niemals Spanisch gelernt haben, sollen als Probanden dienen.

Sie sollen beide Unterricht bei demselben Lehrer erhalten, doch wird ihre Behandlung im Sprachlabor unterschiedlich gestaltet.

Die Versuchsgruppe wird nach der Methode der doppelten Gehörreizung behandelt: Die Schüler hören sich selbst mit dem einen Ohr, und mit dem anderen hören sie das von dem Lehrer besprochene Tonband. Die Kontrollgruppe verwendet denselben Unterrichtsstoff auf Band, jedoch hört sie mit beiden Ohren zu gleicher Zeit die Bandaufnahme und die eigene Stimme.

Die beiden Gruppen werden vor dem Experiment getestet, dann nach drei Monaten. Während dieser Zeit sollen sie wöchentlich zwei Laborübungen gemacht haben. Indem man die Fehler in der Aussprache als Grundlage nimmt, wird man eine Varianzanalyse durchführen und die Signifikanz der Differenzen errechnen.

Dauer: 1 Jahr und 1 Monat.

Autor: C. Van Riper, Western Michigan Univ.

¹¹ Siehe U.S. Department of Health, Education and Welfare, *Cooperative Research Projects*, Fiscal 1960, S. 29.

3. Systematik (Taxonomie) der pädagogischen Forschung

Ziel jeder Taxonomie ist es, präzise Definitionen einheitlich zusammengefaßt darzulegen, die für die Mehrzahl, wenn nicht sogar für die Gesamtheit derer, die sich ihrer bedienen müssen, leicht anwendbar und akzeptabel sind. Ohne eine solche Grundlage ist das Gespräch zwischen Forschern erheblich erschwert.

Der Erziehungsforschung fehlt in weitestem Maße eine in sich geschlossene Klassifizierung. J. W. Best — ihm folgen mehrere andere Autoren — geht z. B. nach der „deskriptiven“ Forschungsmethode vor, jedoch erkennt er an, daß: „Da alle Forschungsarten einen „deskriptiven“ Aspekt enthalten, diese Klassifizierung nicht völlig zufriedenstellend ist¹²“. Er stellt weiterhin fest, daß ein- und dieselbe Studie von einem namhaften Theoretiker als „Übersicht“, von einem anderen als „Fallstudie“ und von einem dritten als „Feldstudie¹³“ bezeichnet wird. Best wählt die zweite Formulierung. Hat er recht oder unrecht? Haben sie alle drei zugleich recht?

Im Jahre 1956 haben B. S. Bloom und Mitarbeiter die erste pädagogische Taxonomie veröffentlicht. Diese Tatsache ist insofern interessant, als sie den Willen zu wissenschaftlicher Exaktheit in der modernen Pädagogik kennzeichnet. Die zwei bis heute herausgegebenen Bände befassen sich jedoch nur mit den allgemeinen Erziehungszielen und beschränken sich auf den kognitiven und affektiven Bereich¹⁴.

Da bisher eine vollständige Taxonomie in der Erziehungsforschung fehlt, haben die folgenden Ausführungen kein anderes Bestreben, als einige aufklärende Beiträge zu leisten.

Die Pädagogik, als sich erst entwickelnde Wissenschaft, muß sich ihre Ordnungsprinzipien und Grundkonzeptionen bei älteren und zeitgenössischen Wissenschaften entleihen, die in ihrer Entwicklung schon weiter fortgeschritten sind als sie selbst — bei den alten Naturwissenschaften wie bei den jungen Humanwissenschaften.

12 J. W. Best, *Research in Education*, Englewood Cliffs, N. J., Prentice-Hall, 1959, S. 103.

13 Dgl. S. 106. Wir werden im Folgenden sehen, daß diese Klassifikationen sich nicht immer gegenseitig ausschließen.

14 B. S. Bloom et al., *Taxonomy of Educational Objectives, Cognitive Domain*, New York, Longmans, Green and Co., 1956.

D. R. Krathwohl, B. S. Bloom, B. B. Masia, *Taxonomy of Educational Objectives, Affective Domain*, New York, D. McKay Cy, 1964.

I. Die Forschung im allgemeinen

Bevor man in der Erziehungsforschung den Versuch einer Klassifizierung unternimmt, erweist es sich zweifellos als sinnvoll, zu beobachten, wie sich die Forschung im allgemeinen aufgliedert.

Diesem einleitenden Teil legen wir die vorläufige Ausgabe der „Taxonomie der wissenschaftlichen Forschung“, von Y. de Hemptinne für den Nationalrat für Politische Wissenschaft (Brüssel, Dez. 1960) und die Ausarbeitungen bedeutender amerikanischer Autoren zugrunde.

Abgesehen von der unterschiedlichen Terminologie, die von Fall zu Fall variiert, je nachdem, ob das Forschungsvorhaben aus eigener Initiative oder im Auftrag einer Organisation, ob es einzeln oder im Team usw. durchgeführt wird, unterscheidet man drei Forschungsrichtungen, die zur Auswertung der Forschungsergebnisse im täglichen Leben hinführen wollen.

A. Grundlagenforschung

(Basic Research; Fundamental Research; Recherche fondamentale)

Es handelt sich hierbei um die „Erforschung neuer Erkenntnisse und neuer Untersuchungsgebiete, ohne ein spezifisches praktisches Ziel“ (de Hemptinne, J. Conant, V. Bush).

Der Forscher bemüht sich, einen Forschungsgegenstand besser kennen- und verstehen zu lernen, ohne sich mit der praktischen Anwendung der neu erworbenen Erkenntnisse zu befassen.

B. Angewandte Forschung

(Applied Research; Recherche appliquée)

„Dies ist eine Forschung mit einem bestimmten praktischen Ziel, im Hinblick darauf, der Menschheit in einem ihrer Bedürfnisse nützlich zu sein“ (Unesco, de Hemptinne).

Man zielt hier auf die praktische Anwendung der wissenschaftlichen Erkenntnis hin; das ist das Zwischenstadium zwischen dem Auffinden und der Nutzbarmachung für den täglichen Bedarf, „die ersten Bemühungen, wissenschaftliche Erkenntnisse in die Technologie zu übertragen“ (President Science Advisory Committee, de Hemptinne).

C. Entwicklungsforschung

(Development Research; Recherche de développement technique)

„Systematische Angleichung der Informationen der angewandten Forschung und der empirischen Erkenntnisse hinsichtlich der Erstellung und des Einsatzes neuer Materialien, Instrumente, Methoden oder Verfahren ...“ (National Science Foundation USA, de Hemptinne).

Die Pädagogik beruft sich auf diese drei Forschungstypen. Man kann beispielsweise versuchen, den Lesevorgang als solchen zu verstehen (Grundlagenforschung), es sich sodann zur Aufgabe machen, im Hinblick auf den Unterricht zu bestimmen, auf welche Weise ein 6jähriges Kind am schnellsten und wirksamsten lesen lernt (Angewandte Forschung); und schließlich kann man den Versuch unternehmen, ein Lesebuch auszuarbeiten, indem man sich auf die Ergebnisse der vorausgegangenen Untersuchungen stützt (Entwicklungsforschung).

Es folgen nun einige Definitionen (diesmal von de Hemptinne ohne Modifizierung entlehnt), die auch für den Pädagogen nützlich sein können:

D. Teamforschung

(Team research, Group research; recherche collective)

Zusammenarbeit der Forscher im Hinblick auf die Lösung eines gemeinsamen Problems.

E. Konvergierende Forschung – Divergierende Forschung

Die Institutionen, die sich mit konvergierender Forschung befassen, führen Untersuchungen über eine zusammenhängende Problemgruppe durch (Beispiel: Schwerpunkt der Untersuchungen auf dem Gebiet der Probleme des technischen Unterrichts).

Die Universitäten, die auf zahlreichen, von einander unabhängigen Gebieten arbeiten, betreiben divergierende Forschung.

F. Auftragsforschung

(Sponsored Research; Recherche commanditée)

Eine Untersuchung, die auf Kosten der Organisation durchgeführt wird, die sie in Auftrag gibt, und die jeweils den Forschungsgegenstand und das Arbeitsprogramm oder aber den Gegenstand, das Arbeitsprogramm und das experimentelle Detail festlegt.

G. Aufwendige Forschung

Untersuchungen, die den Einsatz von aufwendigen Mitteln notwendig machen (wissenschaftliche Mitarbeiter, zahlreiche technische Hilfsmittel, Materialien mit hohen Selbstkosten, umfangreiche Verwaltung usw.).

H. Handlungsforschung

(Operational Research; Action Research; Recherche opérationnelle)

Anwendung der wissenschaftlichen Methode bei Organisationsproblemen, bei Problemen von Arbeits- oder Fabrikationsverfahren.

Über diesen Forschungstyp behalten wir uns eine besondere Bemerkung vor.

I. Reine Forschung

(Pure Research; Recherche pure)

Freie und nicht zweckgebundene Forschung, die ohne spezifisches praktisches Ziel auf das Verständnis des Universums einerseits und auf die Entdeckung neuer Forschungsgebiete andererseits abzielt.

II. Die Pädagogische Forschung

A. Definition der Beobachtung und des Experiments

Die wissenschaftliche Beobachtung ist die aufmerksame Feststellung von Phänomenen – ohne die Absicht, sie zu modifizieren – mit Hilfe der für diese Feststellung geeigneten Untersuchungs- und Forschungsmittel.

Demgegenüber ruft das Experiment eine Beobachtung künstlich hervor, in der Absicht, bestimmte Phänomene zu studieren, eine Idee zu überprüfen oder zu suggerieren¹⁵.

B. Die Beobachtung

1. Ziele

Für manche ist die Beobachtung beschreibend „deskriptiv“ und das Experiment voraussagend „prädiktiv“. Etliche moderne Autoren haben ihre

¹⁵ R. Robert, Dictionnaire de la langue française.

Klassifikation der Pädagogischen Forschung auf diese Gegensätzlichkeit aufgebaut. Nach unserer Ansicht ist dies eine irrtümliche Meinung, und das, was Poincaré hinsichtlich des Experiments schrieb, trifft auch für die Beobachtung zu: „Das gute Experiment ist dasjenige, welches uns mehr erkennen läßt als die isolierte Tatsache; es ist dasjenige, das uns erlaubt vorauszu- sehen, d. h. zu verallgemeinern“ (Die Wissenschaft und die Hypothese)¹⁶. Tatsächlich wird die Feststellung der Phänomene in ihrer spezifischen Bedeutung (ein gewisses Ereignis hat sich zu einer bestimmten Zeit zugetragen) erst in dem Moment zur wissenschaftlichen Forschung, wo die Relationen — kausaler oder nicht kausaler Natur — freigelegt werden und die Verallgemeinerungen — mögen sie auch sehr begrenzt sein — zur Voraussage führen.

Ziel der pädagogischen Beobachtung ist es demnach, Tatbestände bezüglich der Erziehungssysteme, -verfahren und -vorgänge kennenzulernen und vor- auszusagen.

2. Klassifikation

a) nach dem Gegenstand

b) nach dem Modus

1. *Direkte Beobachtung* — Die direkte Beobachtung ist diejenige, bei der der Untersuchende ein Phänomen mit seinen Sinnesorganen wahrnimmt:

— Das Zählen der Fehler in einer Abschrift;

— Das schrittweise Notieren des Gedankenganges eines Kindes, das sich laut denkend um die Lösung einer Aufgabe bemüht.

Die direkte Beobachtung schließt nicht die Verwendung von Instrumenten aus, die die Unvollkommenheiten und Unzulänglichkeiten unserer Sinne ausgleichen und die vom Beobachter erzielten Ergebnisse miteinander ver-

¹⁶ Der Unterschied zwischen Beobachtung und Experiment wird nicht immer eindeutig festgelegt. Und das ist verständlich. C. Ranzoli schreibt: „Der Widerspruch zwischen der Beobachtung und dem Experiment verschwindet oder erlangt zumindest zweit- rangige Bedeutung, wenn man die allgemeine Beobachtung von der methodischen Beobachtung unterscheidet. Die erstere ist nur eine spontane Übung der Sinne; die zweite ist der gleiche Vorgang, er wird vollzogen mit Hilfe von besonderen Mitteln, die die Tragweite (der Beobachtung) erhöhen, die Unvollkommenheit korrigieren aufgrund von Überlegungen, die den Vorgang nach feststehenden logischen Regeln und durch die Wahl von zweckmäßigen Untersuchungsgegenständen und Bedingungen hervor- heben. Wenn man von wissenschaftlicher Beobachtung spricht, kann man nur die zweite im Auge haben, die folglich niemals ganz und gar passiv ist, sondern die immer eine Intervention in die Tatbestände impliziert und die auch immer ein Experiment ist.“ (A. Lalande, Vocabulaire de la Philosophie, Paris, P.U.F. 1956, S. 708–709.)

gleichbar machen: Mikroskope, Kameras, Registriergeräte, Meßinstru- mente, usw.

2. *Indirekte Beobachtung* — Sei es, daß es sich um die Motivierung han- delt, um Fähigkeiten, um den Lernvorgang oder um erzieherische Schwierig- keiten, — die direkte Beobachtung ist tatsächlich selten möglich, weil ent- weder das Objekt als solches unseren Sinnen nicht direkt zugänglich ist (Beispiel: die Intelligenz) oder weil sich die determinierenden Faktoren nur indirekt in der Verhaltensweise äußern. Man hat in diesem Zusammenhang von *verborgenen Variablen* gesprochen.

Travers schreibt: „Die Beobachter neigen zu der Annahme, daß das, was man beobachten kann, d. h. die Reizbedingungen und das Ansprechen auf diese Anreize, eine hinreichende Basis für die Erklärung der Verhaltensweise darstellt. Das ist aber mit Ausnahme einiger weniger Situationen, die nur eine geringe Beziehung zur Erziehung haben, nicht der Fall¹⁷“.

Sobald wir die Schicht der direkten Beobachtung zugänglichen Verhal- tensäußerungen verlassen, arbeiten wir mit hypothetischen Begriffen (*con- structs*), so z. B. mit der Intelligenz, der Fähigkeit zu kritischem Denken, dem Anpassungsvermögen usw. Wir können Manifestationen dieser *Kon- strukte* feststellen, aber unsere Beobachtungen erstrecken sich niemals direkt auf diese selbst.

c) nach der Methode

1. *Die Monographie* — Die Monographie befaßt sich gründlich mit der Untersuchung eines bestimmten, verhältnismäßig eng gefaßten Problem- kreises. Sie behandelt einen einzelnen Gegenstand einer Wissenschaft.

Es ist im allgemeinen eine Grundlagenforschung, die eine Summe von inte- grierten Beobachtungen darstellt, auf die jeder je nach Bedarf zurückgreifen kann.

Litré schrieb: „Ich wollte nichts weiter als eine Monographie über jedes Wort geben, d. h., daß dem Leser eine Abhandlung über alles das vorgelegt würde, was man über ein Wort hinsichtlich seines Ursprungs, seiner Form, seiner Bedeutung und seiner Anwendung weiß¹⁸“.

Ebenso zeigt in der Pädagogik jeder Artikel der „Encyclopedia of Educa- tional Research“ den zu einem bestimmten Zeitpunkt erreichten wissen- schaftlichen Stand.

¹⁷ Op. cit. S. 205.

¹⁸ Wörterbuch, Vorwort, S. X.

Als Forschung in dem Sinne, wie sie von uns verstanden wird, darf eine Monographie keine einfache Beschreibung sein; sie muß vielmehr eine dynamische Untersuchung eines Gegenstandes darstellen, auf den sich der Autor bei seiner Untersuchung ausschließlich beschränkt.

2. *Fallstudie (Case-study)* — Die Fallstudie, die manchmal den Umfang einer sorgfältig ausgearbeiteten Monographie annimmt, unterscheidet sich von dieser nur in einem wesentlichen Punkt: Sie wird angesichts eines aktuellen Problems durchgeführt.

In der Medizin handelt es sich um einen Krankheitsfall, der an dem betreffenden Patienten untersucht wird. In der Pädagogik sprechen wir von einer Untersuchung, die sich auf eine Person, auf eine Gruppe, eine Situation oder eine Erziehungsinstitution erstreckt, mit dem Ziel, eine bessere geistige, soziale oder physische Anpassung zu bewirken.

3. *Die Fallstudienmethode* — Man bezeichnet als Fallstudienmethode die Methode, die ihre Verallgemeinerung auf eine Reihe von „Fallstudien“ gründet.

4. *Die Übersichtsstudie (Survey)* — Die Übersichtsstudie (*Survey*) ist eine umfangreiche Untersuchung eines komplexen Ganzen, die zu einem bestimmten Zeitpunkt betrachtet wird. Es handelt sich also um eine Querschnittuntersuchung (oder horizontale Studie), dazu bestimmt, ein bekanntes Ganzes zu repräsentieren, wobei man insbesondere in statistischer Form der relativen Häufigkeit gewisser Züge oder charakteristischer Variablen Rechnung trägt.

Die Interaktionen oder Interrelationen werden nicht um ihrer selbst willen, sondern eher in ihren Auswirkungen betrachtet. Der *Survey* zieht die Bilanz der Ergebnisse, während die Fallstudie oder die Monographie auch die Ursachen und die Verfahren analysiert.

5. *Die Umfrage* — Die Umfrage unterscheidet sich von dem *Survey* durch den besonderen Ursprung der gesammelten Informationen: sie betrachtet nicht direkt die Tatbestände, sondern sammelt die sie betreffenden Meinungen und die Stellungnahmen.

Ein Teil der in einem *Survey* figurierenden Daten ist oftmals auf dem Wege einer Umfrage gesammelt worden.

Der *Survey* und die Umfrage sind im wesentlichen „deskriptiv“. Auch finden sie ihren eigentlichen Platz in der Forschung nur, wenn sie eine Entscheidung einer Handlung vorbereiten oder die notwendigen Elemente zu einem Schluß führen und wenn dieser formuliert wird.

Ohne eine exakte Definition der Forschungsziele ist es unmöglich, zu bestimmen, welche Informationen zusammengestellt werden sollen.

Umfragearten

- Einmalige und einfache Umfrage. Beispiel: Ein Fragebogen wird an Studenten gerichtet, um ihre Meinungen zu einem Problem zu erfahren.
- Einmalige und gewichtete Umfrage. Beispiel: Die Stichprobe enthält doppelt so viele als „gut“ betrachtete Studenten wie „mittelmäßige“. Somit kann man, ohne die Stichprobe zu sehr zu vergrößern, die Reaktionen der guten Studenten genauer ermitteln.
- Doppelumfrage: Vor- und nach einem Ereignis (*Before-and-after design*). Beispiel: Umfrage vor und nach der Reform der Lehrprogramme.
- Doppelumfrage mit nur einer konstanten Teilstichprobe. Beispiel: vorher: Stichprobe mit Schülern der Gruppen A und B; nachher: Stichprobe mit Schülern der Gruppen A und C.
- Wiederholung der Umfrage, ohne die Stichprobe zu ändern und ohne Kontrollgruppe (*Panel design*).
- Wiederholung der Umfrage mit einer Kontrollgruppe.

d) nach dem Ort

1. Feldstudie

2. *Beobachtung im Laboratorium* — Wir geben später den Unterschied zwischen Feld und Laboratorium an und zwar im Zusammenhang mit dem Experiment.

Im Augenblick beschränken wir uns darauf, den Unterschied zwischen der Feldstudie (*Field Study*) und der Umfrage klar zu machen, Methoden die leicht verwechselt werden.

Aus der Analyse, die D. Katz diesem Problem vorbehalten hat¹⁹, ergeben sich drei Gegenüberstellungen:

- Die Umfrage ist extensiv, während die Feldstudie intensiv ist;
- Die Umfrage zielt darauf ab, die Häufigkeit von Phänomenen in einer Gesamtheit zu bestimmen, von denen sie ein repräsentatives Bild zu geben versucht, während die Feldstudie exploratorisch ist und aus einer eingehenden Prüfung der Vorgänge besteht.
- Die Umfrage erstreckt sich auf die Endergebnisse der Vorgänge, während die Feldstudie an den Vorgängen selbst in ihrer Entwicklung interessiert ist.

19 L. Festinger u. D. Katz, *Les méthodes de la recherche dans les sciences sociales*, Paris, P.U.F., 1959, Bd. 2, S. 69–70.

e) nach der zeitlichen Orientierung

1. Längsschnittuntersuchungen oder Untersuchung des Entwicklungsverlaufs — Man verfolgt die Veränderungen, die Entwicklungen, die sich während einer kürzeren oder längeren Zeitdauer ereignen.

Die kurzen Längsschnittuntersuchungen — diese dauern manchmal nur einige Wochen — erstrecken sich z. B. auf etwas Erlerntes (Aneignung einer Rechentechnik) oder auf Änderung der Persönlichkeit des Schülers (Fluktuation der Interessen, der Einstellungen).

Demgegenüber können sich die langfristigen Studien auf ein ganzes Leben erstrecken. Welche Laufbahn schlägt ein Prüfling nach seiner Schulausbildung ein? Welches ist mittel- und langfristig gesehen die Wirkung einer bestimmten Behandlung (*follow-up*)?

Die lange Dauer gewisser Beobachtungen stellt selbstverständlich ein schweres „Handicap“ für den Untersuchenden dar, der — falls er nicht zu den schnelleren Methoden greift — in vielen Fällen das Resultat seiner Arbeit nicht erlebt. Ferner ist es sehr schwierig, während mehrerer Jahre in Verbindung mit allen Prüflingen einer experimentellen Gruppe zu bleiben (Veränderung des Wohnortes, Erlöschen des Interesses an der Untersuchung usw.). Der Umfang der von den Leitern der Untersuchung „Project Talent“ eingesetzten Mittel, die sich in den USA darum bemüht, Schüler, die im Alter von 15 Jahren getestet wurden, während der folgenden 20 Jahre weiter zu beobachten, zeugt von den Schwierigkeiten des Unterfangens: Ausstellung eines speziellen Ausweises, Herausgabe einer Zeitschrift, um den Kontakt und die Motivation aufrecht zu erhalten, sowie die Einrichtung eines Informationsdienstes.

2. Querschnittuntersuchungen — Es handelt sich um die Beobachtung von Bedingungen zu einem bestimmten Zeitpunkt. Statt den Werdegang derselben Prüflinge während eines ganzen Lebens zu verfolgen, studiert man vielmehr verschiedene Altersgruppen (*cross-sectional approach*). Wenn die Zahl der Probanden groß genug ist, kann man vermutlich bedeutsame Rückschlüsse ziehen.

Bei dieser Methode erhebt sich jedoch ein schwerwiegender Einwand. Je höher die Zahl älterer Probanden in der Stichprobe ist, um so mehr ist ihr repräsentativer Wert anfechtbar. Man kann in der Tat immerhin annehmen, daß eine Selektion progressiv mitgespielt hat, und daß nur solche Personen, die besondere Merkmale aufweisen, die anderen überlebt haben.

C. Das Experimentieren

1. Arten des Experiments

Streng genommen gibt es nur zwei Arten des Experiments:

- das Experiment im Laboratorium;
- das Feldexperiment.

Da die Ethik es in jedem Fall verbietet, Erziehungssituationen zu provozieren, die eine Gefahr für die Versuchsperson darstellen, versuchen die Versuchsleiter systematisch Ergebnisse oder Phänomene zu erschließen, die sie selbst nicht herbeigeführt haben, die aber durch ihren ungewöhnlichen Charakter mit Experimenten verglichen werden können. Nach den von uns übernommenen Definitionen handelt es sich hierbei um Beobachtungen und nicht mehr um Experimente im eigentlichen Sinne. Etliche andere Autoren unterscheiden nichtsdestoweniger:

a) Das nicht-kontrollierte oder natürliche Experiment.

Beispiele: Ein plötzlich auftretender Bevölkerungszuwachs (demographische Entwicklung) kann einen großen Mangel an Räumlichkeiten und Lehrern zur Folge haben und die Einführung von Halbtagsunterricht notwendig machen; während des Krieges konnte man die Auswirkungen des Mangels an Proteinen auf das Lernen feststellen;

b) Das Experiment *ex post facto*, bei dem der Versuchsleiter, um die Ursachen und Gründe aufzufinden von den Wirkungen eines Phänomens ausgeht, das er nicht provoziert hat.

2. Das Experiment im Laboratorium

Es erlaubt einerseits, eine Situation zu erzeugen, „welche die für die Forschung erforderlichen exakten Bedingungen vereinigen soll“ und andererseits, bestimmte Variablen zu überprüfen und andere zu modifizieren. Der Experimentierende ist somit imstande, die Wirkung der Manipulation unabhängiger Variablen auf abhängige Variablen zu messen, und zwar in einer Situation, in der das Einwirken anderer Faktoren (die tatsächlich vorhanden, jedoch der Untersuchung nicht direkt zugehörig sind) auf ein Minimum reduziert wird²⁰. Der Kommentar, den L. Festinger zu dieser Definition gibt, läßt sich auch auf die Pädagogik anwenden: „Es handelt sich hier jedoch um eine Definition, die die Dinge überaus vereinfacht. Aufgrund der dem Untersuchenden zur Verfügung stehenden experimentellen Tech-

²⁰ L. Festinger, in Festinger u. Katz, op. cit. S. 163–164.

niken kann er im besten Fall nur mit grober Annäherung die ideale Präzision erhalten, die die Definition impliziert. In dem Maße, wie die Techniken sich vervollkommen, wird natürlich eine strengere Überprüfung der Laborexperimente möglich sein. Aber zur Zeit umfaßt die Gruppe der Laborexperimente eine Vielzahl von Untersuchungen, deren Präzisions- und Kontrollgrad sehr unterschiedlich sind".

Um die Tragweite dieser Bemerkung zu ermessen, müssen wir die Forschung nach den Gesichtspunkten betrachten, nach denen sie sich in der Praxis abwickelt.

Stellen wir uns ein Experiment vor, das die Hypothese verifizieren soll, nach der ein Text besser gelesen wird, wenn die darin beschriebenen Tatbestände und Empfindungen kurz vor dem Lesen erlebt und nachempfunden worden sind. Es ist anzunehmen, daß der Schüler, sobald die ihm vertrauten Ereignisse angeschnitten werden, den Inhalt des Textes schneller erfassen wird als ein anderer und somit schnellere Fortschritte beim Lesen macht, was sich in einem mehr Worte umfassenden und schnellerem Hin- und Herwandern der Augen äußert. Man bildet eine Versuchsgruppe (die die Situation direkt vor dem Lesen erlebte) und eine Kontrollgruppe, die ohne besondere Vorbereitung liest. Das Hin- und Herwandern der Augen wird gefilmt.

Selbst wenn die gründlichsten Vorkehrungen getroffen werden, um eine möglichst gleichartige Zusammensetzung beider Gruppen zu garantieren (gleicher Intelligenzquotient, gleiches Leseniveau, gleiches soziokulturelles Niveau usw.) werden die Ergebnisse unvermeidlich nur annähernd bleiben: Einige Schüler werden im Augenblick der Untersuchung motivierter sein als andere; die wenig vertrauten Apparaturen beeinträchtigen und verwirren die Probanden in unterschiedlichem Maße; die vor dem Lesen erlebten Tatbestände haben bei den Kindern verschiedene Resonanz, je nachdem, ob sie für die einen vollkommen neu sind und für die anderen dagegen nicht und je nach der gefühlsmäßigen Wirkung, die sie erzeugen usw.

Wir werden feststellen, daß gewisse Versuchspläne es erlauben, die Wirkung dieser Faktoren z. T. auszuschalten. Trotz allem erreicht das Experiment bei den Humanwissenschaften niemals den Genauigkeitsgrad der Naturwissenschaften, da die Persönlichkeiten sehr verschieden und die psychischen Voraussetzungen fluktuierend sind.

Dennoch verhindert diese unvermeidliche Ungenauigkeit nicht den wesentlichen Fortschritt. Wenn man ein neues Medikament ausprobiert, hängen die beobachteten Reaktionen nicht nur von diesem selbst ab, sondern von dem mehr oder weniger guten Gesundheitszustand der Patienten, ihrer Ein-

stellung sowie von den Auswirkungen anderer vorher verabreichter Medikamente. Indessen läßt sich — bis zu einer gewissen statistisch berechenbaren Grenze — feststellen, ob das Heilmittel eine signifikante Verbesserung gebracht hat oder nicht. So schreiten die Medizin wie auch die Humanwissenschaften in ihrer Entwicklung voran, vorausgesetzt natürlich, daß man Erkenntnisse, die nur instrumental sind, niemals als definitive Lösungen betrachtet.

Das pädagogische Labor — Als pädagogisches Labor bezeichnen wir nicht nur die speziell ausgerüsteten Räumlichkeiten, außerhalb der Klassen (Meßinstrumente, Schalldämpfung usw.), die zur Prüfung von Einzelpersonen oder kleinen Gruppen bestimmt sind, sondern auch Klassen oder selbst ganze Schulen, so weit sie im Hinblick auf die Forschungsvorhaben eingerichtet worden sind. Man kann beispielsweise in einer Klasse Schüler in einem Verhältnis vereinigen, das der Verteilung der verschiedenen sozioökonomischen Gruppen in einem bestimmten Bezirk entspricht; man kann auch von Anfang an eine gewisse Verteilung nach dem Intelligenzquotienten berücksichtigen.

Das wesentliche Merkmal des pädagogischen Labors ist demnach die beabsichtigte und durchdachte Schaffung menschlicher und materieller Bedingungen zur Durchführung von Experimenten.

Begrenzter Wert der Experimente im Laboratorium — Im Idealfall soll sich das künstlich erzeugte Experiment ganz natürlich in das Leben der Schüler und Lehrer einordnen. Diese Bedingung ist kaum zu erreichen, es sei denn, daß das Labor die Klasse selbst ist. Abgesondert von seinen Klassenkameraden nimmt das Kind eine andere Verhaltensweise an, selbst wenn es guten Willens an dem Experiment teilnimmt.

Ziemlich schnell nimmt der Prüfling wahr, daß gewisse Übungen keinen „Ernstcharakter“ haben, und seine Motivation ändert sich. So kann man beispielsweise dieses Phänomen bei solchen Untersuchungen beobachten, bei denen man künstlich eine Belastungssituation (*stress*) erzeugt, um zu versuchen, deren Einfluß auf den Lernvorgang zu bestimmen. Die Reaktionen des Kindes sind häufig wenig repräsentativ —, einmal, weil eine experimentelle Belastung kurz ist, während sich die natürliche Belastung auf einen rezeptiven Bereich auswirkt, ziemlich lange dauern und eine Intensität erreichen kann, die willentlich zu erzeugen das elementarste Pflichtbewußtsein verbietet; und zum anderen, weil das Kind beispielsweise schnell den vorgetäuschten Zorn von der wirklichen Bedrohung unterscheidet, und

weil die zeitlich begrenzte Übung, deren Ergebnis nicht die monatliche Schulnote beeinflussen wird, rasch von der täglich „registrierten“ Arbeit, die gewertet wird, unterscheidet.

3. Die Feldstudie

Bei der Feldstudie variiert der Untersuchende bestimmte Variablen nach einem zuvor aufgestellten Plan, jedoch werden für diese Zwecke das menschliche und das materielle Milieu nicht künstlich erzeugt.

Die Mehrzahl der pädagogischen Experimente fällt in diese Untersuchungskategorie, meistens wickeln sie sich in Schulklassen ab, so wie sie sich gerade ergeben²¹.

Die Hauptschwierigkeit bei der Feldstudie liegt in der großen Zahl von verborgenen oder schwer meßbaren Variablen.

Es ist beinahe immer wünschenswert, daß der Feldstudie ein Laborversuch vorausgeht. Nicht nur, weil dort die Beobachtungen präziser sind, sondern vor allem, weil die Zahl der Kinder dort im allgemeinen niedrig ist, was eine wirksame Überwachung und eine frühzeitige Aufdeckung der eventuellen Schädlichkeit des Experiments erlaubt.

Im Rahmen des Möglichen sollte die Feldstudie von den Lehrern selbst ausgeführt und in die tägliche Arbeit mit einbezogen werden. Das läßt sich leider nicht immer verwirklichen; zum einen, weil die Erzieher nicht immer über die notwendige Zeit oder Spezialausbildung verfügen, zum anderen, weil sie bei bestimmten Experimenten persönlich mit einbezogen sind.

So hat auch G. Mialaret recht, wenn er auf der Notwendigkeit besteht, daß der Untersuchende von der Klasse, in der er arbeiten möchte, akzeptiert wird, sich in ihre Atmosphäre aufnehmen läßt und, wenn möglich, dort von Zeit zu Zeit Unterricht abhält²².

D. Anmerkung: Der Voraussagecharakter (prognostische Charakter) der Forschung

Wir haben es bereits gesagt, jede Forschung, die nicht bereits in ihrem Endergebnis eine gewisse Verallgemeinerung erlaubt und somit einen Voraussagewert besitzt, kann nicht als wissenschaftlich qualifiziert werden.

21 Wir bemerken, daß viele Schulen oder Klassen mißbräuchlich als experimentell bezeichnet werden, während man sich darauf beschränkt, einige Versuche der Lehrmethode oder der Organisation ohne strenge wissenschaftliche Kontrolle zu unternehmen.

22 G. Mialaret, *Nouvelle pédagogie scientifique*, Paris, P. U. F. 1954.

Ein Vergleich der Wirksamkeit von zwei Unterrichtsmethoden ist nur dann von wirklichem Nutzen, wenn die Schlussfolgerung vorauszusagen erlaubt, welche von beiden — unter annähernd gleichen Voraussetzungen — künftig die besten Dienste leisten wird. Ein *survey* über Schulgebäude soll vor allem dazu dienen, den künftigen Bedarf festzustellen, bzw. anzuzeigen, welche Ausrüstung für die Schulen, die man zu bauen oder einzurichten beabsichtigt, am geeignetsten zu sein scheint.

Die Geschichte der Wissenschaft zeigt, daß der Fortschritt zu einem bedeutenden Maße von der Qualität der Voraussagen abhängt, die der Mensch macht, und daß deren Wert wiederum von der umfassenden Kenntnis abhängt, die man von den beobachteten Phänomenen hat²³.

Denn es wird mitunter vergessen, daß man Vorhersagen bezüglich eines Phänomens machen kann, ohne sein Wesen näher zu kennen. Im Altertum konnte man in der Tat ein Gewitter — ohne dessen Mechanismus zu kennen — aufgrund einer Reihe von empirischen Beobachtungen ankündigen.

Eine nicht entwickelte Wissenschaft gründet ihre Voraussagen auf das einfache Wechselspiel von Versuch und Irrtum. Man stelle sich vor, daß die „Wunderdoktoren“ der Vergangenheit stark auf gut Glück eine Reihe von „Heilmitteln“ verabreichten, und daß sie danach, beeinflusst von zufälligen Erfolgen, Regeln für die Zukunft aufstellten.

Die Pädagogik geht heute noch bei so mancher Gelegenheit in ähnlicher Weise vor. Bestehen nicht viele Forschungen darin, eine bestimmte Anzahl Tests anzuwenden oder Beobachtungen in Menge zu sammeln, dann den Korrelationskoeffizienten zu berechnen, dem man einen Voraussagewert zuschreibt, ohne wirkliche Kenntnis der beobachteten Phänomene?

Die Genauigkeit der Statistik stellt nicht allein eine Garantie für wissenschaftliche Gültigkeit dar.

Travers macht die humorvolle Bemerkung, daß die Korrelation zwischen der Schönheit der Studentinnen und den Prüfungsergebnissen nicht bedeutet, daß eine Beziehung zwischen Intelligenz und Schönheit besteht, wohl aber zwischen der Schönheit der Studentin und der Großherzigkeit gewisser Prüfer...

Um auf die meteorologischen Vorhersagen zurückzukommen, hier weiß man, daß sie erst in dem Moment einen Grad beachtlicher Zuverlässigkeit erreicht haben, wo die umfassende Kenntnis der Phänomene den für die primären Beobachtungen notwendigen Interpretationsrahmen geboten hat. Wenn die auf den Zufall aufgebaute Untersuchung (*hit-or-miss approach*)

23 Die folgenden Anmerkungen sind von R. Travers beeinflusst, op. cit. S. 274–305.

manchmal zu Ergebnissen führt, so sind diese doch selten und kostspielig. Wir kommen damit auf ein schon weitreichendes Grundprinzip zurück: Die voraussagende Forschung, mit echtem wissenschaftlichem Charakter, soll sich auf eine theoretische Kenntnis von dem Wesen der vorauszusagenden Phänomene stützen, und diese theoretische Kenntnis legt das Aufstellen einer Arbeitshypothese nahe.

Es ist dennoch selbstverständlich, daß Prüfungen, die sich als prognostisch zu erkennen geben, auf den Weg zur Erklärung von grundsätzlichen Verfahrensweisen führen können. Wenn man feststellt, daß ein Test über das abstrakte Denken es erlaubt, diejenigen Studenten festzustellen, die später ihr Studium am besten abschließen werden, kann ein Forschungsprojekt über das abstrakte Denken selbst wichtige Aspekte des „Lernvorgangs“ ins Licht rücken.

Aber, denken wir daran, daß die Korrelation auch nur nach außen hin bestehen und in Wirklichkeit auf einem verborgenen dritten Faktor beruhen kann. Weiterhin kann sie auch nur umstandsbedingt sein: ein bestimmter Faktor ist mit dem Erfolg bestimmter Studien verbunden, weil sie sich gerade auf ihn beziehen (Beispiel: abstraktes Denken und Denksportaufgaben), was keineswegs bedeutet, daß das Messen dieses selben Faktors auch den Erfolg in anderen Unterrichtsfächern zeitigen würde.

Schließlich kann die Vorhersage tatsächlich nur unter zwei Voraussetzungen wirksam sein:

1. Das vorauszusagende Phänomen muß weitmöglichst homogene Ursachen haben. Man stelle sich vor, wie schwierig es beispielsweise ist, den beruflichen Erfolg vorauszusehen: dieser hängt nicht nur von den intellektuellen Fähigkeiten des Prüflings ab, sondern auch von seiner Erscheinung, seinen charakterlichen Qualitäten, seinem Sinn für menschliches Zusammenleben und von Unterstützungen, die ihm von anderer Seite gewährt werden usw.
2. Das vorauszusagende Phänomen muß wohl definiert werden. Um etwas über die Tüchtigkeit der Lehrer vorauszusagen, muß man sich zuerst klarmachen, was unter einem „tüchtigen“ Lehrer zu verstehen ist.

E. Die Handlungsforschung in der Pädagogik

Wir räumen der Handlungsforschung in der Pädagogik einen besonderen Platz ein, denn sie stellt nach unserer Meinung das wirksamste Verbindungsglied zwischen dem spezialisierten Forschen und dem Praktiker dar, zwischen Grundlagenforschung und Schulpraxis.

1. Definition

Der Begriff „educational action research“ erscheint in der amerikanischen Literatur erstmalig im Jahre 1948. Doch noch im Jahre 1953, als S. M. Corey sein heute allgemein bekanntes Buch veröffentlichte²⁴, drückte Kenneth Wann die Vermutung aus, daß die Methodologie dieser Untersuchungsart sich erst zu entwickeln beginne²⁵.

Tatsächlich aber wurde die allgemeine Methode, bevor sie in der pädagogischen Forschung Anwendung fand, bereits während des zweiten Weltkrieges in anderen Bereichen eingesetzt. John E. Magee definiert sie als: „die systematische Anwendung der wissenschaftlichen Methoden und Techniken (...) zur Untersuchung der Probleme der Betriebsführung, der öffentlichen Angelegenheiten, der militärischen Tätigkeiten. Sie verfolgt das Ziel, eine quantitative Illustration der wesentlichen Elemente, die einen bestimmten Vorgang darstellen, sowie der Faktoren, die das Ergebnis beeinflussen, zu liefern und somit eine solche Grundlage für die zu treffenden Entscheidung zu geben²⁶“.

Man erkennt sofort die Veranlassung zur Anwendung dieser Methode im schulischen Bereich. Um Maßnahmen aufgrund präziser Einzelheiten zu ergreifen, müßte die pädagogische Aktivität der Analyse unterzogen, und ihre verschiedenen Faktoren müßten quantitativ erfaßt werden. Dem gesunden Menschenverstand würde sich somit die wissenschaftliche Exaktheit, die in den Dienst täglichen Handelns gestellt wird, überlagern.

Man errät ohne Mühe die methodologischen Fortschritte, zu denen derartige Bemühungen um Präzision führen können.

2. Forschungsbeispiel

Hilda Taba und Elisabeth Noel, denen wir die zuverlässigste Methodologie der Handlungsforschung in der Pädagogik verdanken, haben im Jahre 1957 den Bericht über eine Untersuchung veröffentlicht. Diese macht den eingeschlagenen Weg deutlich, um ein spezielles Problem zu lösen und weist gleichzeitig den praktischen Nutzen nach, den die Erzieher daraus ziehen können.

24 S. M. Corey, *Action Research to Improve School Practice*, New York, Columbia Univ., 1953.

25 Siehe: *Review of Educational Research*, Bd. XXIII, Nr. 4, Okt. 1953, S. 342.

26 J. E. Magee u. A. D. Little, *Inter-Operation Research* (N. A. C. A. Bulletin, Juni 1954, S. 1252).

Vor einigen Jahren traten die Verantwortlichen des Unterrichtswesens von Yolo County (Kalifornien), beunruhigt durch die mangelnde Dynamik ihrer Schulen, mit der Bitte an Hilda Taba heran, mit ihnen zusammenzuarbeiten, um eine grundsätzliche Untersuchungsbasis zu schaffen, „die dazu dienen sollte, den Lehrern zu helfen, ihre praktischen Arbeitsmethoden in der Klasse zu modifizieren, ihre Probleme zu untersuchen und neue Lehrpläne zu erproben“²⁷.

Yolo County zählte zu der Zeit 39 Volksschulen und 5 Sekundarschulen (mehr als 400 Lehrer). Die oberste Leitung der Untersuchung umfaßte außer dem Direktor sieben pädagogische Berater und einen Lehrplanspezialisten.

Es zeigte sich schnell, daß man zuerst die Einstellung dieser neun Verantwortlichen, die bis dahin — von der Lehrerschaft allzu sehr distanziert — fast ausschließlich Führungsrollen gespielt hatten, grundlegend modifizieren mußte.

Für die Lehrer legte man grundsätzlich fest, daß allein die freiwillige Teilnahme am Experiment zulässig wäre, daß kein Teilnehmer von einem anderen vertreten werden dürfe und bestimmte Schulen ohne weiteres nicht am Unternehmen teilzunehmen brauchten.

Anfänglich meldeten sich fünfzehn Lehrer. Diese kleine Zahl ist nicht verwunderlich. Einerseits ist es nicht immer leicht, den Pädagogen dahin zu bringen, daß er absolut offen über seine Schwierigkeiten und über die Mißerfolge in seiner Klasse spricht. Andererseits läßt das abstrakte Erklären — wie wir es im Augenblick tun — die Handlungsforschung als eine ziemlich vage Methode erscheinen.

Nach zwei Jahren war die Teilnehmerzahl von fünfzehn auf mehr als hundert angestiegen; mehr als ein Viertel der ganzen Lehrerschaft von Yolo County war somit an einer wissenschaftlichen Untersuchung beteiligt, die sich mit spezifischen Problemen ihrer Schüler befaßte.

Die Methode der Handlungsforschung wurde peinlich genau angewandt; man ging aus von Problemen, die von den Lehrern selbst aufgeworfen wurden, um das Arbeitsprogramm auszuarbeiten.

Als wichtigste Fragegebiete wurden untersucht: Die lernschwierigen Schüler („slow learners“), die Nachzügler im Lesen, die Technik des Gesamtunterrichts, die Gruppenbildung in der Klasse, die für die Eltern angefertigten Schularbeitsberichte, die Identifizierung der schlecht angepaßten Kinder, die Untersuchung der Bedürfnisse der heranwachsenden Jugendlichen einer sozio-ökonomisch besonders gelagerten Gruppe und die menschlichen Beziehungen innerhalb der Klasse.

Wir geben hier zusammenfassend einen der von H. Taba²⁸ dargelegten Fälle wieder. Es handelt sich um eine Lehrerin der 3. Volksschulklasse, die den im Lesen erzielten Erfolg bei einem Großteil der Klasse als nicht ausreichend betrachtete.

1. Identifizierung des Problems — Die Lehrerin gab an, daß zwölf Kinder selbst bei sehr leichten Wörtern beim Lesen ins Stocken gerieten und das, was sie heute gelernt, am folgenden Tag bereits wieder vergessen hätten. Dazu bemerkt die Autorin, daß die Problemstellung ziemlich einfach war: Man stellte nicht nur ohne Schwierigkeiten fest, daß die angewandte Methode starke Mängel aufwies, sondern darüber

27 H. Taba, E. Noel, Action Research: A Case Study, Washington, N. E. A., 1957, S. 6.

28 Op. cit. S. 27 ff.

hinaus versuchte sich die Lehrerin selbst Rechenschaft über die Ursachen des Rückstandes im Lesen und ihre wahrscheinlich unzureichende Unterrichtstechnik zu geben. Ohne diese Einsicht hätte man damit beginnen müssen, die grundsätzliche Einstellung der Lehrerin zu modifizieren.

2. Problemanalyse und Zusammentragen der ersten objektiven Daten — Von den möglichen Ursachen, die einen Rückstand der Lesefertigkeit der Schüler begründeten, wurden vier festgehalten:

- a) Aneignung schlechter Lesegewohnheiten bei den Schülern im Laufe der ersten zwei Volksschuljahre;
- b) geistige Retardierung;
- c) affektive Schwierigkeiten in der Schule oder in der Familie, die einen großen Teil der Energie der Kinder absorbieren;
- d) ein Mangel an Eigenerlebnissen, der das Verständnis für einen gelesenen Text erschwert und ein Desinteresse hervorruft.

Die Lehrerin erklärte sich außerstande, allein zu entscheiden, welcher bzw. welche Faktoren eine entscheidende Rolle hierbei spielten. Der Berater half ihr, sich anhand der Untersuchung der schulischen Vergangenheit, der Bestimmung der sozialen Herkunft, der Charakteranalyse, der Diagnose der Intelligenz, der Beobachtung der Lesegewohnheiten und der Auswertung des exakten Rückstandes in diesem speziellen Fach zu orientieren (diese Untersuchung erstreckte sich hier auf eine Zeitdauer von einem Jahr sieben Monate auf zwei Jahre sechs Monate).

3. Die Hypothesen — Die Analysen zeigten zwei mögliche Ursachen an: Den Mangel an Reife für das Verständnis der gelesenen Texte und schlechte Lesegewohnheiten. Man stellte demzufolge die Hypothese auf, daß sich die Situation bessern würde, wenn:

- a) Den Schülern anschaulicher, interessanter Lesestoff vorgelegt würde;
- b) ein konkreter Lesestoff verwendet werden könnte, um eine engere Beziehung zwischen den Eigenerlebnissen und dem Textinhalt herzustellen;
- c) die ganze Gruppe außerhalb des Leseunterrichts gemeinsam etwas erleben würde, dessen Diskussion dazu beitrüge, die Beziehungen zwischen geschriebenem und dem gesprochenen Wort zu vertiefen;
- d) der Lesestoff abwechslungsreicher und weniger mühsam wäre²⁹.

4. Das Experiment — Wegen der ziemlich großen Klasse und der knapp bemessenen Zeit der Lehrerin wurde beschlossen, einen Film, dem ein kleiner Lesetext beigegeben war, und einen Film, der in stehenden Bildern die wichtigsten Ereignisse zusammenfaßte, zu verwenden. Obgleich die Lesbarkeitsformel von Flesch³⁰ anzeigte, daß der Textinhalt weit besser vom 5. Schuljahr an verstanden würde, entschied man sich dennoch dafür, ihn zu verwenden³¹.

29 H. Taba, op. cit. S. 29.

30 Siehe die Anmerkungen über diese Formel im Kapitel über die Inhaltsanalyse, S. 219.

31 Es handelt sich um Gray Squirrel, Encyclopedia Britannica Films, 1949. Lehrbuch von P. Witty, Boston, Heath, 1949.

Das kleine Lesebuch umfaßte 160 verschiedene Wörter. Man half der Lehrerin, die genaue Zahl der den Schülern unbekanntem Wörter festzustellen, es waren im ganzen 48.

Die Arbeit in allen Einzelheiten zu beschreiben, würde in diesem Rahmen zu weit führen und überdies für die erfahrenen Pädagogen, die dieses Buch lesen, nutzlos sein. Beschränken wir uns auf die Feststellung, daß, nachdem die gesamte Klasse den Film gesehen hatte, die zurückgebliebenen Schüler einem intensiven Training unterzogen wurden (Diskussionen über den Film, Zeichnungen, Modellieren, Lückentexte usw.) Der Film wurde jedesmal erneut gezeigt, sobald sich die Notwendigkeit hierfür ergab.

5. *Auswertung* — Während der Dauer des Experiments machte sich die Lehrerin sehr genaue Notizen über vier Kinder: das beste, das langsamste, ein durchschnittliches und ein schüchternes Kind. Der Lehrplanspezialist untersuchte seinerseits die Entwicklung der Schüler.

Sobald die erste Geschichte beendet war, ergab eine quantitative Kontrolle merkliche Fortschritte: Von zwölf Kindern versagten acht bei nur einem Wort pro Seite. Nach einem fünfmonatigen Training machten sich beim Lesen Fortschritte bemerkbar, die zwischen einem Monat und einem Jahr und drei Monaten lagen. Der beste Schüler hatte beispielsweise fünf Monate im aktiven Wortschatz und ein Jahr und drei Monate im Textverständnis gewonnen.

3. Warum die Handlungsforschung erfolgreich ist

Der Erzieher beteiligt sich an der Lösung einer Schwierigkeit, die ihn persönlich beschäftigt. Er kann somit die Hilfe feststellen, die die Forschung ihm gewähren kann.

Die menschlichen Beziehungen verbessern sich. Dank der gelösten Arbeitsatmosphäre spricht der Lehrer vorbehaltlos über seine Schwierigkeiten. Lehrer, Direktoren, Schulräte werden durch die Bewältigung eines gemeinsamen Vorhabens geschult, wobei man der Sachkenntnis den Vorrang vor der hierarchischen Ordnung gibt; jeder ist sich seiner eigenen Leistungsgrenze bewußt.

Zu einer Zeit, in der in allen Bereichen menschlicher Aktivität ein unaufhörlicher Apell an die Fachspezialisten ergeht, braucht die Intervention eines pädagogischen Beraters normalerweise keinen Autoritätskonflikt herbeizuführen.

Die Einschaltung des Beraters erübrigt sich übrigens in dem Maße, in dem sich die Erzieher mit den experimentellen Techniken vertraut machen.

Im Idealfall wird sich der höhere Mitarbeiterstab möglichst schnell bilden. Im Hintergrund bleibt der Berater, den die Amerikaner als „Informationsperson“ bezeichnen, d. h. die Person, die sich in der Entwicklung der Wissenschaft so weit wie irgend möglich auf dem laufenden hält und die bestrebt

ist, ‚Produzent‘ und ‚Konsument‘ der Forschung zu vereinigen. Ein Hauptmerkmal der Handlungsforschung — so wie sie hier verstanden wird — ist, wiederholen wir es nochmals, daß sich hierbei Vorgesetzte wie auch Untergebene zur gleichen Zeit bilden, was die Bezeichnung völlig rechtfertigt, die sich die Gruppe der Universität von Columbia vorbehalten hat: „Cooperative Action Research“.

4. Handlungsforschung, eine Methode zur Lehrerbildung

Wir haben bisher nur von der in Zusammenarbeit mit den ihren Beruf ausübenden Lehrern durchgeführten Arbeit gesprochen. Es ist jedoch sicher, daß die Methode auch bei Studierenden angewandt werden kann. In einer modernen Lehrerbildungsanstalt, die der wissenschaftlichen Forschung die ihr zustehende Bedeutung beimißt, sollte die Einführung in die Handlungsforschung einen Schwerpunkt in der Ausbildung darstellen. Die Erfahrungen, die wir gemacht haben, bestätigen auffallend, daß die Technik der „aktiven Forschung“ den Studierenden eine vertiefte Motivation vermittelt. Neue Begriffe wie Statistik, Soziometrie, die Technik, menschliche Beziehungen anzuknüpfen usw. werden dann funktionell eingeführt.

5. Schlußfolgerung

Die Handlungsforschung ist weder völlig noch grundsätzlich neu. Das, was neu an ihr ist, ist ihre Systematisierung und die Grenzen, die sie sich selbst eindeutig auferlegt: An die Stelle abstrakter und allgemeiner wissenschaftlicher Forderungen tritt das Bestreben nach einer rationellen Schul- bzw. Klassenführung, deren Leistung man steigern möchte.

Es wäre töricht zu glauben, daß die Handlungsforschung quasi von heute auf morgen alle Lösungen der schwierigen Unterrichtsprobleme bringt. Sie führt vornehmlich — wie R. Diez Hochleitner³² es hervorhebt — zu einer verfeinerten Analyse der erzieherischen Situation und bestimmt die Richtung für die zu treffenden Entscheidungen, indem sie routinemäßige Gepflogenheiten und Vorurteile aus dem Wege räumt.

Wir glauben, daß zum gegenwärtigen Zeitpunkt die Handlungsforschung unsere größte Hoffnung auf eine progressive und grundlegende Erneuerung unseres Unterrichtswesens erfüllt, als das einzige zur Verfügung stehende Mittel, um tatsächlich eine Brücke zwischen universell tendierender wissenschaftlicher Forschung und täglicher Praxis zu schlagen.

32 R. Diez Hochleitner, *Utilización de la educación comparada en el planeamiento integral de la educación* (Internationale Zeitschrift für Erziehungswissenschaft), Unesco, 1959, Nr. 3, S. 102.

Wir wollen uns nicht lange mit der Polemik aufhalten, die noch etliche amerikanische Fachleute beschäftigt bei der Erörterung der Frage, ob die Handlungsforschung es verdient oder nicht, als Forschung bezeichnet zu werden, und ob es nicht richtiger wäre, in ihr lediglich eine aktive Methode zur Heranbildung der Lehrerschaft und der Schulaufsichtsbeamten zu sehen.

Man kann der Handlungsforschung absolut den Rang einer wissenschaftlichen Forschung streitigmachen, denn sie hat nicht das Bestreben, ihre Schlußfolgerungen als allgemeingültig dahinzustellen. Dennoch haben wir es bei den Humanwissenschaften beobachtet, daß die Regeln und Gesetze entsprechend den jeweiligen Gruppen, auf die man sie anwenden möchte, neu interpretiert werden müssen.

Wie dem auch sei, nach einer gewissen Zeit sollte die Handlungsforschung normalerweise zu einer echten lokalen Wissenschaft führen, der einzigen, die letztlich in der Erziehung von Bedeutung ist. Man kann tatsächlich beobachten, daß von Jahr zu Jahr die Variationen in der sozioökonomischen und kulturellen Rekrutierung vieler unserer Schulen ziemlich gering sind. Der Lehrer würde wahrscheinlich während mehrerer aufeinanderfolgender Jahre gewisse auf einem begrenzten Gebiet entwickelte Techniken anwenden und aus diesem Vorsprung den Nutzen ziehen, um andere bisher noch nicht überwundene Schwierigkeiten in Angriff zu nehmen.

Zweiter Teil: Das Forschungsinstrumentarium

Abschnitt 1: Die allgemeinen Techniken

Einführung

Zwei Vorbemerkungen, die für fast alle Verfahren und Techniken der Pädagogischen Forschung gelten, müssen festgehalten werden:

1. Sie sind soviel wert, wie derjenige, der sich ihrer bedient, wert ist.
2. Getrennt angewandt, geben sie kein hinreichend deutliches Bild von der menschlichen Wirklichkeit. Im allgemeinen sind die untersuchten Verhaltensweisen komplex und erfordern daher sehr vielfältige, gleichzeitige Beobachtungen und Beurteilungen¹.

Oftmals sind sogar die Verhaltensformen so fließend und wechselnd, daß kein Instrument sie wiedergeben kann. In diesem Sinn wird der klinische Sinn, der Sinn für die Gesamtsituation beim Beobachter eine entscheidende Rolle spielen. Doch wird damit erneut dem Subjektivismus Tür und Tor geöffnet.

Dieser Gefahr kann man jedoch begegnen, indem man dasselbe Phänomen von mehreren Untersuchern gleichzeitig beobachten läßt. Der Beobachtungsgegenstand muß dabei selbstverständlich mit größter Sorgfalt definiert und eine standardisierte Methode für die Protokollierung der Beobachtungen festgelegt sein.

Die Beobachtungen variieren je nach Person, und zwar aus vielerlei Gründen: Unterschiedliche Vertrautheit mit dem Untersuchungsgegenstand, unterschiedliches Interesse, vorgefaßte Meinung über das in Frage stehende Problem, unterschiedliche Lebenserfahrung usw.

Darüber hinaus neigen manche Personen mehr, manche weniger dazu, das, was sie beobachten grundsätzlich zu verzerren (*high distorters und low distorters*). Diese Verfälschung mag sowohl durch Unzulänglichkeiten im sensorischen oder intellektuellen Bereich als auch durch Persönlichkeitsprobleme bedingt sein. Speziell im Hinblick auf diese läßt sich sagen, daß man niemals ohne Grund verzerrt.

Die Forschungsinstitute haben natürlich größtes Interesse daran, solche Mitarbeiter zu gewinnen, die für Beobachtungen besonders geeignet sind. Aber bis heute scheint es noch nicht gelungen zu sein, ein einfaches Ausleseverfahren zu schaffen, das es erlaubt, mit Sicherheit festzustellen, ob es sich bei

¹ Diese Vielfalt an Beobachtungen wird von E. Wolter besonders gut erläutert in: *Initiation à l'observation systématique des élèves. La méthode d'Albert Huth*, Louvain, Paris, Nauwelaerts 1957.

einer Person um eine konstante oder rein zufällige Tendenz zum Entstehen handelt².

Um die Verzerrungen zu reduzieren, verlangt man im allgemeinen von den Beobachtern, daß sie das, was sie feststellen, ohne irgend eine Interpretation notieren. Man berücksichtigt nur die übereinstimmenden Ansichten und erstellt erst später die Inhaltsanalyse.

Um persönliche Faktoren auszuschalten, versucht man gelegentlich die Beobachter in Unkenntnis darüber zu lassen, ob der Proband, den sie beobachten, der Untersuchungsgruppe angehört oder nicht.

Im Idealfall sollten selbst die Probanden nicht wissen, daß sie an einem Experiment teilnehmen. Andernfalls könnten sich ihre Motivationen und ihre Normalreaktionen grundlegend ändern^{2a}.

Man bezeichnet als „Hawthorne-Effekt“ die positiven oder negativen Ergebnisse, die nicht durch experimentelle Faktoren verursacht werden, sondern von dem psychologischen Effekt, den das Wissen darum, daß man an einer Untersuchung teilnimmt und Gegenstand einer besonderen Beobachtung ist, auf den Probanden ausübt.

1. Der Fragebogen

Einführung

Nachdem T. L. Kelly erklärt hat, daß der Fragebogen³ „das ärmste Meßinstrument ist, das in die hohe Wissenschaft eingeführt wurde“, gibt er zu, daß „dieses launische Kind der Wissenschaft, wie schwach es auch sein

2 Die Versuche, insbesondere die von J. McPerson, waren ziemlich enttäuschend. Siehe: *Predicting the Accuracy of Oral Reporting in Group Situation*. Lackland, Air Force Research Center, 1954, zitiert von Travers, op. cit. S. 203–204.

2a Aus diesem Grunde verwendet man manchmal zum Testen von Medikamenten die folgende Methode: Es ist weder den Patienten noch dem Pflegepersonal bekannt, welches Medikament verabreicht wird. Es werden zwei Patientengruppen, die die gleiche Krankheit haben, nach dem Zufallsprinzip gebildet: die eine erhält das Medikament, die andere ein „Placebo“, d. h. eine inaktive Substanz, die in der gleichen Form wie das Medikament verabreicht wird.

3 Etliche amerikanische Autoren unterscheiden das „schedule“ – das Formular wird in Gegenwart des Befragers ausgefüllt (eine Methode, die eine unmittelbare Rückgabe gewährleistet und es erlaubt, eventuell auftretende Unklarheiten in der Fragestellung zu klären) – von dem eigentlichen Fragebogen; diese Bezeichnung behält man sich nunmehr für zugesandte Fragebogen vor. Im allgemeinen ist diese Unterscheidung im deutschen nicht üblich.

mag, ein unentbehrliches Hilfsmittel bleiben wird“, bis die experimentelle Wissenschaft es uns erlaubt, auf menschliche Urteile zu verzichten oder bis sie unser Interesse für einmalige Ereignisse zum Erlöschen bringt⁴.

In mancherlei Hinsicht ist die Methode der Fragebogen unvollkommen. Zuerst einmal erfaßt der Fragebogen schwerlich alle Aspekte eines Problems. Eine schöpferische Voranalyse ist selten möglich, und die Zahl der Fragen darf nicht zu hoch sein.

Oftmals werden die Fragen in dem Sinne, wie der Befragte die Situation wahrnimmt, gestellt.

Auf die von dem Forscher mehr oder weniger subjektiv ausgewählten Aspekte äußern die Befragten ihrerseits eine gleichermaßen subjektive Meinung.

Tatsächlich ist es selten, daß die ganze Untersuchungsgruppe (Population), an die man sich wendet (der Fragebogen dient dazu, schnell eine große Zahl von Probanden zu erfassen), richtig informiert ist und über die aufgeworfenen Probleme hinreichend nachgedacht hat. Darüber hinaus versucht der Befragte häufig, sich aus Interessengründen oder der persönlichen Bequemlichkeit halber nach der Meinung des Befragers zu richten. In vielen Fällen sind die Antworten auch mehr das Echo von inneren Konflikten als der Ausdruck einer wohl überlegten Meinung, was K. Lewin übrigens zu der Bemerkung veranlaßte, daß „wir es lernen müßten, einen Fragebogen so zu handhaben, wie wir gewöhnlich ein projektives Verfahren handhaben“⁵.

Welche Vorkehrungen man auch immer getroffen haben mag, der Fragebogen ist ein wenig zuverlässiges Instrument, vor allem, wenn es sich um Meinungen handelt. F. Mosteller, der die gleichen Fragen in einem Abstand von drei Wochen wiederholt stellte, bestätigte eine Übereinstimmung von 96,5% bei Antworten bezüglich Tatsachen („Welches ist Ihre Wagenmarke?“) und von nur 79% bei Antworten, die eine Meinung bezüglich eines aktuellen Sachverhaltes ausdrücken („Glauben Sie, daß Roosevelt ein guter, ein durchschnittlicher oder ein schlechter Präsident ist?“⁶).

Eine andere Schwierigkeit: Bei Umfragen, die sich auf eine große Anzahl von Personen erstrecken, antworten die Befragten nicht alle – bei Fragebogen, die durch die Post geschickt werden, ist ein Verlust von 60 bis 80%

4 T. L. Kelly, *Scientific Method*, Columbus, Ohio State Univ. 1929, S. 39, zitierte von R. R. Rusk, *An Outline of Experimental Education*, London, Macmillan, 1960.

5 K. Lewin, *Field Theory in the Social Science*, Tavistock Publications, 1952, S. 16, zitiert von R. Rusk, S. 20.

6 Siehe H. Cantril, *Gauging Public Opinion*, Princeton Univ. Press, 1944.

nicht ungewöhnlich — so daß die erhaltenen Antworten Gefahr laufen, nicht mehr repräsentativ zu sein, trotz aller Vorsichtsmaßnahmen beim Erstellen einer Stichprobe. Wallace hat insbesondere gezeigt, daß der Teil der Bevölkerung, der gewöhnlich auf die Fragebogen antwortet, hinsichtlich des Bildungsniveaus einer recht homogenen Schicht angehört⁷.

Damit die Stichprobe ihren repräsentativen Charakter behält, muß ein Maximum an Antworten eingeholt werden. Mehrfache Aufforderungen sind häufig notwendig (direkte Kontaktaufnahme, Telefonanrufe, briefliche Verbindung: *callback principle*) Dies ist bei einer anonymen Umfrage selbstverständlich unmöglich.

Es ist weiterhin wahrscheinlich, daß — wenn man sich an eine homogene Population wendet (z. B. Gymnasiallehrer) — die Antworten am spontansten von solchen Personen gegeben werden, die besondere psychologische Merkmale aufweisen.

Schließlich muß man auch mit dem Überdruß rechnen, der sich aufgrund des Mißbrauchs von Fragebogen — „diese bequeme Art, Informationen zu sammeln“, wie es John Best ausdrückt⁸, zeigt. Es gibt da nicht nur eine Fragebogenmode, sondern es nehmen offenbar aus Gründen der leichten Handhabung junge Forscher bei der Suche nach einem Arbeitsgegenstand auch gerne zu diesem Hilfsmittel ihre Zuflucht.

I. Aufbau des Fragebogens

A. Fragearten

Man unterscheidet zwischen den Fragen mit „geschlossenen“ („closed“ oder begrenzten oder im voraus festgelegten) Antworten und den Fragen mit „offenen“ („open“ oder freien oder unbegrenzten) Antworten.

1. Fragen mit geschlossenen Antworten

Der Befragte muß lediglich aus mehreren ihm vorgelegten Antworten die nach seiner Meinung richtige Antwort wählen.

⁷ D. Wallace, A case for or against mail questionnaires (Public Opinion Quarterly, 1954, XVIII, S. 40–52).

⁸ Op. cit., S. 143.

Beispiele:

- Würden Sie gerne Englisch lernen? Ja? Nein?
- Von den unten angegebenen Fremdsprachen unterstreichen Sie diejenigen, die Sie gerne lernen würden: Englisch, Deutsch, Russisch.

Wenn er knapp zu formulierende Antworten vorsieht, wie die obigen, verzichtet der Untersuchende auf Informationen, die in bestimmten Fällen für ihn nützlich sein könnten. Es erweist sich oftmals als vorteilhaft, die Möglichkeit einer „offenen“ Antwort einzuräumen.

Beispiel:

- Warum üben Sie einen Sport aus? Um sich zu entspannen; um Ihre physische Konstitution zu behalten; weil Ihre Eltern es wünschen; anderer Grund (bitte spezifizieren).

Wenn man zum Verfahren mit geschlossenen Antworten greift, ist es zweckmäßig, zwischen den einzelnen Items einen ausreichenden Abstand vorzusehen, damit der Befragte seinen Kommentar hinzufügen kann, wozu oftmals der Wunsch besteht.

Vorteile bei geschlossenen Antworten: 1. Sie erlauben es, den Befragten zu lenken und ihm Möglichkeiten aufzuzeigen, die er vielleicht vergessen oder außer acht gelassen hätte. 2. Die einheitliche Form erleichtert die Auswertung. 3. Die geschlossenen Antworten eignen sich direkt für die Kodierung (Verwendung von Lochkarten).

2. Fragen mit offenen Antworten

Beispiel: Warum üben sie einen Sport aus? Antwort:

Der Befragte antwortet hierbei spontan und verwendet seinen eigenen Wortschatz, was tiefere Rückschlüsse auf seine Psyche und besonders auf sein kulturelles Niveau erlaubt.

Aber die Auswertung der offenen Antworten ist zeitraubend und wirft oft große Schwierigkeiten bei der Klassifizierung und der Kodierung auf.

Zur Feststellung der Wahlantworten, die für Fragen mit geschlossenen Antworten auszuwählen sind, beginnt man in bestimmten Fällen damit, einen Fragebogen mit offenen Antworten einer ersten Stichprobe zu geben, um die häufigsten Reaktionen zu identifizieren.

B. Einige praktische Ratschläge

Es gibt kein Rezept, das automatisch zu einem perfekten Fragebogen führt. Die folgenden Überlegungen werden jedoch dem Studenten helfen, häufig auftretende Fehler zu vermeiden.

Für alle Fälle gilt folgende Regel: Jeder Fragebogen muß vor seiner Verbreitung erprobt werden. Dieser Vortest deckt oftmals eine Reihe von Konstruktionsfehlern auf: schlecht formulierte Fragen, mangelhafte Anordnung usw.

1. Abfassung

Der Fragebogen soll so knapp wie möglich abgefaßt werden. Je mehr Zeit die Antworten erfordern, um so geringer sind die Chancen, daß sie gegeben werden. Insbesondere sollte man es vermeiden, nach Informationen zu fragen, die durch ein anderes Forschungsmittel einzuholen sind (z. B. durch Nachschlagen der Schulakten, der Einwohnermelderegister, Sachregister und der Jahrbücher). Die Abfassung muß besonders sorgfältig behandelt werden.

2. Plan

Einführung — Sie hat vor allem zum Ziel, den Befragten zu motivieren, indem man ihm erklärt, warum man gerade seine Mitarbeit wünscht, und indem man die generellen oder speziellen Vorteile aufzeigt, die sich aus der Umfrage ergeben können, und indem man ihm gleichzeitig alle notwendigen Garantien der Diskretion gibt.

Ein gesondertes Schreiben ersetzt manchmal diese einführende Erklärung.

Fragen — Man stellt im allgemeinen die Fragen (Items), die sich auf den gleichen Gegenstand oder den gleichen Aspekt beziehen, in Gruppen zusammen.

In jeder Gruppe werden die Fragen in logischer Anordnung dargelegt — einer Logik, wie sie vom Befragten, nicht aber vom Untersuchenden bevorzugt wird. Die amerikanischen Soziologen verwenden den Ausdruck „funnel approach“ (trichterförmiges Vorgehen), um den Verlauf von den allgemeinen zu den immer spezielleren Aspekten anzudeuten.

3. Auswahl der Fragen im Hinblick auf das verfolgte Ziel

Eine exakte Problemanalyse sollte der Ausarbeitung des Fragebogens vorausgehen, und es wäre zu wünschen, daß der Untersuchende sich bei jedem Item folgende Frage stellt: „Inwieweit fördert die Antwort die Untersuchung?“

4. Aufbau des Fragebogens im Hinblick auf die zu befragenden Personen⁹

Um den Befragten zum Mittelpunkt zu machen, ist es wichtig, daß der Frage-

⁹ Nach C. Cannell u. R. Kahn, in: L. Festinger u. D. Katz: *Les méthodes de recherche dans les sciences sociales*, Paris, P.U.F. 1959, S. 402–406.

bogen seine Sprache, sein Bezugssystem und sein Informationsniveau berücksichtigt; darüber hinaus müssen die Fragen von sozialen Gesichtspunkten her akzeptabel sein.

Die Sprache — Der in dieser Hinsicht am besten geeignete Fragebogen benutzt den Wortschatz und den Satzbau derjenigen Personen, an die er sich wendet. Man trägt sogar manchmal gewissen mundartlichen Merkmalen Rechnung.

Das Informationsniveau — Die befragten Personen sollen normalerweise alle an sie gestellten Fragen beantworten können.

Um peinliche Situationen oder eine Hemmung zu vermeiden, wozu das Zugeständnis, daß man etwas nicht weiß, führen würde, schlagen Cannell und Kahn etwa folgende Redeweise als Vorsichtsmaßnahme vor:

„Die Mehrzahl der Bevölkerung hat bisher keine Gelegenheit gehabt, sehr viel über die technischen Probleme zu erfahren, die durch den Umgang mit radioaktiven Stoffen aufgeworfen werden, aber einige Personen haben sich dennoch mehr oder weniger darüber informieren können. Wissen Sie vielleicht zufällig, ob ...?“

Das Bezugssystem — Jeder Mensch hat seine eigene „Logik“. Er interpretiert die Worte und Tatbestände nach seiner Erfahrung, seiner Persönlichkeit, seinen Kenntnissen und ganz allgemein nach seinem Wertesystem. Der Untersucher tut gut daran, das von ihm verfolgte Ziel klar zu präzisieren, wenn er befürchtet, daß der Befragte nicht in der gewünschten Weise auf seine Fragen eingeht. Ein kurzer Hinweis genügt im allgemeinen, um das „Bezugssystem“ deutlich zu machen.

Beispiel:

Man stellt häufig fest, daß eine Beziehung zwischen Schulleistung und familiären Voraussetzungen besteht: Wohnsitz, Fahrstrecken, eigenes Arbeitszimmer, Bildungsniveau der Eltern usw.

• Haben Sie ein eigenes Arbeitszimmer? Ja — Nein

Die einführende Bemerkung gibt hier an, daß die gestellte Frage nicht einer willkürlichen persönlichen Neugierde entspringt, sondern daß man sich aus rein pädagogischen Gesichtspunkten damit befaßt.

Fragen, die für den Befragten akzeptabel sind — Im allgemeinen antwortet der Befragte nicht auf die Frage, die er als „eine Bedrohung“ seiner Person¹⁰ ansieht und die ihn dazu veranlassen würde, einen Tatbestand anzuerkennen, bei dem er befürchtet, in Verruf zu kommen.

¹⁰ Op. cit., S. 406.

Wenn man einem Schüler die Pistole auf die Brust setzt und fragt, ob er bei Prüfungen manchmal schwindelt, läuft man sehr stark Gefahr, eine wertlose Antwort zu erhalten. Auch hier kann eine kurze vorbereitende Erklärung am Platz sein:

„Man findet selten Schüler, die — wenn sie sich einem Prüfungsthema unvorbereitet gegenübergestellt sahen — nicht mindestens einmal in ihrem Leben ‚gemogelt‘ haben.“

„Ist so etwas bei Ihnen schon einmal vorgekommen?“

„Wenn ja, unter welchen Umständen?“

Wenn der Fragebogen nicht anonym ist, muß der Befrager vorsorglich erklären, daß die Antworten vertraulich behandelt werden und diese Verpflichtung auch streng respektieren.

5. Einige „Schönheitsfehler“, die bei der Abfassung der Fragen zu vermeiden sind

Vage Begriffe — Man findet selten einen Fragebogen, der nicht irgendwelche Zweideutigkeiten enthält.

Beispiele:

- „Sollten nach Ihrer Meinung diejenigen Schüler, die eine Klasse wiederholen, Nutznießer einer besonderen Behandlung sein? Ja ... Nein.“

Bei dieser Fragestellung kann der Begriff „besondere Behandlung“ auf mancherlei Arten interpretiert werden: hinsichtlich der Disziplin, der Arbeitsorganisation, der Individualisierung des Unterrichts, Modifizierung des Unterrichtsprogrammes usw. Die Antwort: „Ja“ sagt dem Befrager praktisch gar nichts.

- „Beschäftigung?“ ...

Dieser Begriff ist zu vage. Handelt es sich um die augenblickliche Beschäftigung oder um diejenige, für die der Befragte qualifiziert ist? Der Befragte gibt im allgemeinen seine berufliche, seine „offizielle“ Beschäftigung an; aber vielleicht übt er andere, bezahlte Tätigkeiten aus, die den größten Teil der Zeit beanspruchen, die man durch Freizeitbeschäftigung ausgefüllt glaubt usw.

- Das Wort „Alter“ ist auch nicht prägnant genug. „Geburtsdatum“ wird jede Unklarheit beseitigen.

- Unklare Adjektive und Adverbien: *mittelmäßig, durchschnittlich, höher, häufig, selten, viel, wenig* ... Diese Begriffe drücken eine Relativität aus, wobei es praktisch unmöglich ist, die genaue Bedeutung nach der Vorstellung des Befragten zu erkennen. Um die Häufigkeit der Phänomene abzuschätzen, wird man sich eher auf folgende Fragen berufen: „Wieviel Kinovorstellungen besuchen Sie gewöhnlich pro Woche? — pro Monat?“

Zur besseren Abschätzung werden Bezugspunkte sinnerklärend wirken:

Ist das Musikprogramm Ihrer Schule:

Dürftig (etwa 0 bis 5 gelegentlich gelernte Lieder im Verlauf eines Jahres);
durchschnittlich (mindestens eine Stunde pro Woche; Vokal- und Instrumentalmusik);

umfangreich (mehr als eine Stunde pro Woche; Vokal- und Instrumentalmusik; rhythmische Übungen; Musikgeschichte).

Doppelte Verneinungen — Sie sind oftmals die Ursache von Schwierigkeiten und Mißverständnissen.

Beispiel:

„Würden Sie nicht den Schulen, die keine komplett ausgerüstete Turnhalle haben, finanzielle Unterstützung gewähren?“

Mutmaßliche Annahmen — Solche Vermutungen wirken hemmend und können verstümmen; so muß man z. B., bevor man eine Person über die Benutzung ihres Fernsehapparates befragt, sich zuerst vergewissern, ob sie tatsächlich einen besitzt.

Ein einziges Problem pro Frage — Eine Frage, die sich auf zwei verschiedene Aspekte bezieht (*double barreled question*), kann eine zweideutige Antwort zur Folge haben.

Beispiel:

„Sind Sie der Ansicht, daß die Schüler, die eine Klasse wiederholen, in einer Abteilung als Gruppe zusammengefaßt werden und einen individuellen Unterricht erhalten sollen? Ja — Nein.“

Man kann den ersten Vorschlag ablehnen und den zweiten akzeptieren. Wie soll man in diesem Fall reagieren bzw. antworten?

Die tendenziösen- oder Suggestiv-Fragen

Beispiel:

„Sind Sie für die traditionelle Erstlesemethode oder für die Ganzheitsmethode, die zumindest den Vorteil bietet, der Psychologie des Kindes besser zu entsprechen?“

Es ist offensichtlich, daß die beiden Formulierungen in dieser Frage nicht auf der gleichen Ebene liegen und daß der Befrager damit schon ein Werturteil vorschreibt.

II. Genauigkeitskontrolle der Antworten und ihre Auswertung

Die Kontrolle der Antworten ist nicht immer leicht. Einige Hilfsmittel führen dennoch zu guten Resultaten.



A. Konsistenzüberprüfung des Fragebogens

Man kann feststellen, ob die Antworten miteinander vereinbar sind: Ist das Alter der Kinder vereinbar mit demjenigen der Eltern, entspricht der angezeigte Stand dem sozio-ökonomischen Niveau, steht eine bestimmte Einstellung nicht im Widerspruch zu anderen Angaben usw.

Manchmal wird die gleiche Frage an verschiedenen Stellen und verschiedenen formuliert gestellt, um die Reaktionskonstanz zu überprüfen.

Ein Widerspruch braucht jedoch nicht in jedem Fall als ein Mangel an Aufrichtigkeit ausgelegt zu werden. Es kann eine Frage in der einen Form besser verstanden werden als in der anderen. Darüber hinaus weicht die „Logik“ des Befragten manchmal erheblich von der des Untersuchers ab.

B. Vergleichsgruppe

Eine reduzierte, besonders gut bekannte oder für die direkte Kontrolle besonders geeignete Eichstichprobe erlaubt es manchmal, Normen zu definieren. Antworten der untersuchten Gruppe, die außerhalb dieser Normen liegen, sollten Gegenstand besonderer Aufmerksamkeit sein.

C. Vergleiche mit anderen Informationen

Ein späteres Interview, Dokumente (Schulakten usw.) und das Ergebnis der Beobachtungen erlauben ebenfalls brauchbare Überprüfungen.

D. Auswertung¹¹

Man soll — wie gesagt — mit dem Fragebogen nicht nur einen Vortest durchführen, sondern auch die Auswertungsmethode erproben, bevor mit der Umfrage begonnen wird. Manche Schwierigkeiten und Mißverständnisse werden auf diese Weise vermieden.

Die Auswertungs- und Kodierungspläne müssen dennoch eine gewisse Flexibilität zulassen. Tatsächlich erfordern unvorhergesehene Reaktionen oder Modifikationen der Ansichten im Verlauf der Untersuchung Abänderungen des ursprünglichen Schemas.

11 Siehe auch: Kodierung der Fragebogen, im dritten Teil, S. 236.

2. Die Unterhaltung oder das Interview

Die meisten der hinsichtlich des Fragebogens gemachten Bemerkungen treffen auch für das Interview zu¹².

I. Klassifizierung

A. Nach der angewandten Methode

1. Die freie Unterhaltung oder das nicht-strukturierte Interview

Im Gegensatz zu der gelegentlichen Konversation wird dieses Interview mit dem Ziel auf eine ganz bestimmte Information durchgeführt. Da gibt es beispielsweise den Meinungsaustausch zwischen Lehrer- und Elternschaft, um ein aktuelles Problem zu lösen. Langdon und Stout schreiben hierzu: „Dies ist ein Interview, das allein der Lehrer unternehmen kann (...), denn er allein kann über die kleinen Dinge sprechen, die sich in einer Klasse zutragen, mit allen notwendigen affektiven Nuancen; tatsächlich stellt er eine mit in die Vorgänge verwickelte Person dar¹³“.

Die freie Unterhaltung eignet sich schwerlich für die Quantifizierung.

2. Das Tiefeninterview (non directive depth interview)

Es handelt sich hier um ein nicht-strukturiertes Interview, das durch die Psychoanalyse in den Vordergrund gestellt wurde. Anstatt eine Reihe von Fragen zu stellen, schneidet der Untersucher ein Thema an (Problem, Ereignis ...) und läßt den Befragten sich beliebig dazu äußern. Die Intervention des Prüfers beschränkt sich lediglich auf einige Zeichen der Ermutigung („aha, o ja, das ist interessant, und dann usw.) und gegen Ende der Unterhaltung auf einige Fragen zur Klärung gewisser Punkte.

Dieses Interview erlaubt ein besseres Verständnis der Motivationen, der Konflikte, der Einstellungen der Befragten, die — sobald das Vertrauen hergestellt ist — nach und nach ihre Besorgnis, ihre Frustrationen, ihre Gefühle, ihre Hoffnungen und ihre Vorurteile offenbaren.

12 Über das Interview siehe auch: P. Minon, *Initiation aux méthodes d'enquêtes sociales*, Brüssel, La Pensée catholique, Paris, Office General du Livre, 1959, 2. Aufl.

13 G. Langdon u. I. Stout, *Teacher-Parent Interview*, Englewood Cliffs, N. J., Prentice-Hall, 1960, 7. Aufl. S. 4.

3. Das laute Überlegen

Die Technik des lauten Überlegens kann als eine Art Interview bezeichnet werden.

Der Schüler wird aufgefordert, bei der Lösung einer Aufgabe seine Gedankengänge laut auszusprechen, wobei man den geistigen Abwicklungsprozeß und somit die Ursachen der Erfolge wie auch der Fehlleistungen untersuchen kann¹⁴.

4. Das strukturierte Interview oder die gelenkte Unterhaltung

Diese dient dazu, Informationen in standardisierter Form zu sammeln. Alle befragten Personen antworten auf die gleichen Fragen, erhalten die gleichen Erläuterungen, und die Unterhaltungen wickeln sich unter weitestgehend gleichen Bedingungen ab. Man erreicht somit eine Art von mündlich prä-sentierten Fragebogen.

5. Das halb-strukturierte Interview oder die gelenkte (oder zentrierte) Unterhaltung

Hierbei mißt der Befrager der Standardisierung weniger Bedeutung bei als vielmehr der Information selbst. Dennoch soll man am Ende der Unterhaltung eine Reihe präziser Ziele erreicht haben. „Ein Schema definiert die wichtigsten zu untersuchenden Themen und sieht eventuell bestimmte Fragen vor; über die Art, wie die Themen im Laufe der Unterhaltung behandelt werden, die Formulierung der Fragen sowie die Reihenfolge, in der Themen und Fragen erscheinen, wird nichts im voraus festgelegt¹⁵“.

B. Nach der Anzahl der Teilnehmer

1. Das Einzelinterview

2. Das Gruppeninterview¹⁶

Ein gut geführtes Gruppeninterview kann wertvolle Fingerzeige erbringen. Diejenigen Befragten, die den Mut aufbringen, sich als erste zu äußern (sie

14 Siehe auch in diesem Zusammenhang: A. M. de Moraes, Recherche Psychopédagogique sur la solution des problèmes d'arithmétique, Louvain, Nauwelaerts-Paris, Vrin, 1954, S. 9–19.

15 J. Maisonneuve u. M. Duclot, Les techniques de la psychologie sociale (Bulletin de Psychologie, 201, XV, 7. August 1962).

16 In dieser Rubrik verwechseln wir bewußt das Gruppeninterview und die Gruppendiskussion. Das Gruppeninterview – so wie es von einigen Soziologen verstanden wird – wobei jegliche Intervention durch den Interviewer erfolgen soll, selbst wenn es sich nicht an ihn richtet, ist derart künstlich, daß es für die pädagogische Forschung ohne Bedeutung ist.

verstehen es vielleicht am besten, ihre Meinung auszudrücken), ziehen die anderen mit. Dank diesem Anreiz treten manchmal Kritiken oder Vorschläge zutage, die die Einzelunterhaltung niemals ergeben hätte. Man weiß auch, wie sehr das Wechselspiel der Ideenassoziation eine Diskussion fördert.

Das Gruppeninterview verfolgt im allgemeinen gleichzeitig zwei Ziele: 1. Faktuelle Informationen zu vereinigen (z. B.: Vorschläge hinsichtlich der Organisation des Schulwesens) und 2. die Verhaltensweise der befragten Personen zu beobachten. Zu diesem Zweck notieren Beobachter, wie die Teilnehmer intervenieren, und welches das Hauptmerkmal ihrer Teilnahme ist: konstruktiv, synthetisch usw. Die gesammelten Einzelnotierungen lassen eine Anzahl Rückschlüsse auf die Intelligenz und die Persönlichkeit der Teilnehmer zu.

Die Technik des Gruppeninterviews wurde manchmal – wie es scheint mit Erfolg – für die Selektion künftiger Lehrer angewandt.

Formen:

a) freie Diskussion

Ein ziemlich allgemeines Thema wird vorgeschlagen, und die Gruppe diskutiert ohne Unterbrechung durch den Befrager, der die Rolle eines „Zeugen“ übernimmt. Themenbeispiel: „Die Schulerziehung kann ohne die Mitwirkung des Elternhauses nicht zum vollen Erfolg führen“.

b) gelenkte Diskussion

Auch sie konzentriert sich auf ein allgemeines Thema, jedoch leitet ein gewandter Gesprächsführer die Diskussion, wirft einige Bemerkungen dazwischen, die das Interesse in Gang halten und die passiven Probanden zur Teilnahme anregen sollen.

Im folgenden bringen wir eine Darstellung eines Beobachtungsplanes der Interaktionen während eines Gruppeninterviews¹⁷.

17 Siehe R. F. Bales, Interaction Process Analysis: A Method for the Small Groups, Cambridge, Mass., Addison-Wesley Press, 1950.

Kategorien,
die sich bezie-
hen auf:
Positive sozio-
emotionale
Interaktionen

Neutrale
Beiträge

Negative
sozio-emotio-
nale Inter-
aktionen

1. Zeigt Solidarität, hebt den Status der anderen, hilft, belohnt.
2. Zeigt Entspannung, scherzt, lacht, zeigt seine Zufriedenheit.
3. Zeigt sein Einverständnis, zeigt passive Billigung, paßt sich an.
4. Macht Vorschläge, gibt Hinweise, die die Selbständigkeit der anderen implizieren.
5. Tut seine Meinung kund, beurteilt, analysiert, drückt seine Gefühle und Wünsche aus.
6. Gibt Aufklärung, Auskunft, erklärt, bestätigt.
7. Bittet um Aufklärung, Auskunft, Wiederholung, Bestätigung.
8. Fragt andere nach ihrer Meinung, Beurteilung, Analyse, ihren Äußerungen.
9. Bittet um Vorschläge, Anweisungen, Verhaltensrichtlinien.
10. Ist nicht einverstanden, lehnt passiv ab, hilft nicht.
11. Zeigt Spannung, bittet um Hilfe, geht aus dem Felde.
12. Zeigt antagonistische Einstellung, erniedrigt die anderen, denkt nur daran, sich selbst zu behaupten.

a b c d e f

- Schlüssel: a Probleme der Kommunikation
 b Probleme der Beurteilung
 c Probleme der Kontrolle
 d Probleme der Entscheidung
 e Probleme des Spannungszustandes
 f Probleme der Reintegration
 A Positive Reaktionen
 B Antworten
 C Fragen
 D Negative Reaktionen

Anmerkung: Über die Technik der Gruppendiskussion im eigentlichen Sinne siehe insbesondere: W. Mangold, Gegenstand und Methode des Gruppendiskussionsverfahrens, Frankfurt/M., Europäische Verlagsanstalt, 1960.

II. Vorbereitung des Interviews

A. Ausbildung des Befragers

Gewisse Menschen scheinen eine besondere Begabung für menschliche Kontaktaufnahme zu besitzen, d. h. außergewöhnlich kontaktfreudig zu sein; mit guter Ausbildung stellen sie erstklassige Interviewer dar.

Dennoch lehrt die Erfahrung, daß eine adäquate Vorbereitung es auch vielen Untersuchern mit weniger guten Voraussetzungen erlaubt, fruchtbare Unterhaltungen herbeizuführen.

Eine der in den USA am häufigsten angewendeten Ausbildungstechniken ist die des *role playing* (Rollenspiel): In einem fingierten Interview spielen der Befragter und die Studenten abwechselnd die Rolle des Befragten und die des Befragers. Die Ergebnisse des *role playing* sind nicht unbeachtlich, jedoch eignen sich nach unserer Erfahrung die amerikanischen Studenten besser für diese Übungen als die europäischen.

Das beste Training ist und bleibt das der Wirklichkeitssituationen. Die Analyse und die Kritik der registrierten Interviews im Seminar ergeben auch gute Resultate.

Schließlich ist die Beobachtung von Interviewern mit langjähriger Erfahrung immer lehrreich.

B. Unmittelbare Vorbereitung

1. Des Befragers

Das gesamte Interview soll sorgfältig durchdacht und vorbereitet werden, wobei sich der Befragter mit den wichtigsten Punkten des zu untersuchenden Problems vertraut macht. Ohne das tiefere Verständnis der verfolgten Ziele ist es kaum möglich, aus der Unterhaltung das herauszuholen, worauf es ankommt.

Je weniger das Interview strukturiert ist, um so stärker sind die Geschicklichkeit des Befragers, sein Scharfsinn und sein psychologisches Feingefühl von Bedeutung.

2. Des Befragten

Abgesehen von wenigen Ausnahmen, beruft sich die pädagogische Forschung nicht auf Unterhaltungen *ex abrupto*, wobei der Befragte absichtlich nicht vorbereitet wird. Es ist im Gegenteil in den meisten Fällen wichtig,

daß der Befragte gut vorbereitet ist, um die Fragen, die an ihn gerichtet werden, beantworten zu können.

Das zuverlässigste Mittel, seine Mitarbeit zu bewirken, besteht darin, daß man ihn vorher von der Wichtigkeit seiner Antworten überzeugt sowie auf die Vorteile hinweist, die sich dadurch für ihn direkt oder indirekt ergeben können.

Sobald der Befragte einmal davon überzeugt ist und man sein Vertrauen gewonnen hat, übertrifft seine Mitarbeit oftmals die optimistischsten Erwartungen.

III. Abwicklung des Interviews¹⁸⁾

Ebensowenig wie für den Fragebogen gibt es kein perfektes Rezept für das Interview. Wir wollen daher in den folgenden Ausführungen keineswegs Regeln aufstellen, sondern machen lediglich ganz schlichte, allgemeine Angaben.

1. „Wenn der Befragte sich zum erstenmal irgendeiner Versuchsperson vorstellt, wird er beobachten, daß die Situation Form annimmt, noch bevor er angefangen hat zu sprechen.“¹⁹⁾
2. Wichtig dabei ist, daß man eine Atmosphäre der Sympathie und des gegenseitigen Vertrauens schafft, indem man dabei ganz natürlich, aufrichtig und aufgeschlossen bleibt. Der Befragte wird jedoch eine möglichst neutrale Haltung einnehmen: Es ist nicht Aufgabe, im Namen der Moral, der Pädagogik oder der Psychologie zu sprechen, sondern sich zu informieren.
3. Auch die Ausdrucksweise des Interviewers muß neutral sein: weder pedantisch noch zu technisch nach dem Niveau des Gesprächspartners falsch angepaßt. Niemand, selbst wenn er noch so ungebildet ist, erwartet von dem Befragten, daß er sich geistlos, fade oder unkorrekt ausdrückt. Eine schlichte und angemessene Haltung weckt das Vertrauen, demagogische Lässigkeit zerstört es.
4. Glücklicherweise, ihre Meinung über Erziehungsprobleme ausdrücken zu können, darüber hinaus oftmals für Ratschläge und Rechtfertigungen sehr empfänglich, antworten die Eltern mitunter weitschweifend.

¹⁸⁾ Interessant zu lesen ist der Interviewvorschlag für ein unartiges oder nervöses Kind von R. Cattell: A Guide to Mental Testing, London, Univ. of London Press, 1954, S. 408 bis 410.

¹⁹⁾ Festinger u. Katz, op. cit. S. 415.

Es ist dabei wichtig, nichts zu überstürzen, entgegenkommend und aufgeschlossen zu bleiben und das Gespräch geschickt auf die eigentlichen Ziele hinzulenken.

5. Bei Erziehungsfragen berührt man leicht die Intimssphäre der Einzelpersonen und der Familien. Die Zurückhaltung, die der Befragte gewissen Punkten gegenüber wahrnimmt, soll respektiert werden, selbst wenn sie einen wichtigen Aspekt der Untersuchung verbirgt.
6. Indiskretion oder autoritäres Verhalten führen zu verständlichen Hemmungen und schaden daher der Untersuchung.
7. Während der Diskussion hartnäckig auf einer Sache zu bestehen, führt zu nichts, dies erzeugt unnötige Spannungen und Gleichgültigkeit. Aus Angst, sein Gesicht zu verlieren, nimmt der Befragte in bestimmten Fällen unbedenklich seine Zuflucht zu Trugschlüssen. Wenn die Frage wichtig ist, kann sie im weiteren Verlauf des Gesprächs noch einmal, anders formuliert, gestellt werden.

IV. Notieren der Antworten

Wenn man die Film- und/oder Tonbandaufnahmen, die ohne Wissen des Befragten vorgenommen werden, ausschaltet, — dieses Verfahren ruft schwerwiegende Einwendungen hervor — verbleiben kaum mehr als zwei unvollkommene technische Hilfsmittel:

— sich im Laufe der Unterhaltung — mit der Zustimmung des Befragten — Notizen zu machen oder in diskreter, aber nicht verdächtiger Form Protokoll zu führen, was die Atmosphäre des Interviews in einer schwer zu beschreibenden Weise stört.

— die Antworten nach der Unterhaltung so schnell wie möglich zu protokollieren, was unvermeidlich zu einer Verstümmelung und Verunstaltung führt.

V. Kritische Analyse der gesammelten Informationen

Absolut aufrichtige Antworten sind selten, vor allem, wenn die Fragen die Persönlichkeit betreffen.

Getzels²⁰⁾ hat gezeigt, daß der Interviewte in dem Augenblick, da er befragt wird,

²⁰⁾ J. W. Getzels, The Question-Answer Process: A Conceptualization and Some Derives Hypothesis for Empirical Examination (Public Quaterly, VIII, 1954, S. 79–91); siehe Travers, Introd., S. 243.

1. sich unmittelbar im Innern eine Antwort bildet, deren er sich mehr oder weniger klar bewußt wird;
 2. sich nach der jeweiligen Situation richtet, in der er sich gerade befindet;
 3. schließlich seine Antwort ausspricht, die übrigens vielfach so formuliert ist, wie sie nach seiner Meinung dem Befrager am meisten zusagt.
- Man begreift, wie schwierig es ist, die so entstandenen Verunstaltungen genau einzustufen. Es sind auf diesem Gebiet noch manche Untersuchungen anzustellen.

3. Die „check lists“

I. Definition

Die „check lists“, die einfachste Form des Fragebogens, ist ein einfaches Inventarblatt (Kontroll- oder Richtblatt), das dazu dient, die Beobachtung zu lenken und zu systematisieren. Sie soll das Vorhandensein oder das Fehlen eines Gegenstandes feststellen, ohne eine Bewertung oder eine Beurteilung zu formulieren.

Man verwendet die „check lists“, um die Arbeitsmethoden, die Reaktionen der Schüler, die Tätigkeit der Lehrer, die behandelten Unterrichtsstoffe, die Bücher, Radio- und Fernsehprogramme usw. zu analysieren.

II. Beispiele

Es gibt eine Unzahl von „check lists“, die alle auf dem gleichen Prinzip beruhen. Die folgenden Beispiele unterscheiden sich nur durch den Beobachtungsgegenstand.

A. Analyse der Arbeitsmethode

In der folgenden Aufstellung finden sich verschiedene Lernmethoden, welche die Schüler allgemein anwenden, um ihr englisches Vokabular zu lernen.

Macht ein Kreuz an das Verfahren, das Ihr gewöhnlich anwendet:

- ... Ich lese die Wörter ein- oder zweimal, dann wiederhole ich sie auswendig.
- ... Ich schreibe die Wörter ab, dann wiederhole ich sie auswendig.
- ... Ich versuche, die Wörter aus dem Englischen zu übersetzen und verweile nur bei solchen, die ich nicht weiß.

- ... Ich versuche, die Wörter aus dem Deutschen zu übersetzen und verweile nur bei solchen, die ich nicht weiß.
- ... Jemand läßt mich mündlich die Wörter übersetzen.
- Usw.

B. Analyse der Verhaltensweise des Kindes

Discoll Identification sheet ²¹

Schule: Schuljahr: Lehrer:
Aufgrund Ihrer Beobachtungen schreiben Sie unter die folgenden Sätze die Namen derjenigen Schüler, die sich häufig folgendermaßen verhalten:

I

- a) Richtet sich gewöhnlich nach der Klassenordnung und akzeptiert sie;
- b) Muß häufig zur Ordnung aufgerufen werden;
- c) Unberechenbares Verhalten.

II

- a) Arbeitet ununterbrochen an einer gestellten Aufgabe;
- b) Leicht abgelenkt von der gestellten Aufgabe;
- c) Sucht zuviel Aufmerksamkeit und zuviel Hilfestellung von seiten des Lehrers.

III

- a) Bringt spontan eigene Ideen;
- b) Bringt niemals von sich aus Ideen, es sei denn, daß er dazu aufgefordert wird;
- c) Unregelmäßig;
- d) Sehr selbstsicher, wenn er Ideen bringt.

IV

- a) Scheint ausgezeichnet und ein fleißiger Arbeiter zu sein;
- b) Scheint schwerfällig im Begreifen.

V

- a) Beliebt bei anderen Kindern;
- b) Von anderen Kindern gemieden oder nicht beachtet.

VI

- a) Sucht ständig Kontakt zu anderen Kindern;
- b) Sucht zu sehr die Aufmerksamkeit Erwachsener auf sich zu lenken;
- c) Ergreift selten die Initiative zum Kontakt mit anderen Kindern;
- d) Achtet nicht auf das Entgegenkommen anderer Kinder.

²¹ Siehe M. Almy, Ways of Studying children, New York, Columbia Univ., 1966, S. 58.

VII

- a) Scheint gewöhnlich glücklich;
- b) Scheint Spannung zu zeigen, leicht aus der Fassung zu bringen;
- c) Hat nervöse Angewohnheiten;
 - Nägelkauen;
 - Daumenlutschen;
 - Grimassenschneiden;
 - Onanie;
 - Tagträumen.

VIII

- a) Besucht die Schule regelmäßig;
- b) Häufig abwesend wegen geringfügiger Erkrankung;
- c) In der Schule anwesend, hört aber nicht zu, müde, blaß, nicht wohl.

IX

- a) Seelisch ausgeglichen;
- b) Überhaupt nicht ausgeglichen und ohne Ausdauer.

X

Hat Sprachfehler:

- a) Schlechte Aussprache (spricht wie ein Baby);
- b) Flüstert;
- c) Stottert;
- d) Vertauscht Buchstaben.

C. Analyse der elterlichen Verhältnisse

„Check list“ für die Analyse der Eltern-Kind-Relationen, von R. Cattell²².

1. Zuneigung — Gleichgültigkeit

1. Die Eltern²³ sprechen selten mit dem Kind, ohne dabei zu lächeln.
2. Die Eltern versuchen, sich in die Lage des Kindes zu versetzen, wenn man seine Verhältnisse bespricht.
3. Die Eltern machen dem Kind kleine Geschenke außerhalb sonst üblicher Anlässe.
4. Die Eltern nehmen das Kind zum Ausflug oder auf die Reise mit, selbst wenn dies für sie gewisse Unbequemlichkeiten mit sich bringt.
5. Die Eltern beklagen sich, daß das Kind sie ermüdet und sie ständig kritisiert (negativ).
6. Die Eltern heißen das Kind, aus dem Haus oder ihnen aus dem Weg zu bleiben (negativ).

²² In: A Guide to Mental Measurement, London, Univ. of London Press, 1953, 3. Aufl. S. 362—364.

²³ Wir übersetzen mit „Eltern“, während in Wirklichkeit zwei getrennte Aufstellungen für die Mutter und den Vater erstellt werden müssen.

7. Die Eltern richten es sich nicht systematisch ein, bestimmte Zeiten am Tag mit ihrem Kind zusammen zu verbringen (negativ).
8. Die Eltern küssen oder begrüßen das Kind nicht nach einer längeren Abwesenheit.

21. Die Eltern stehen auf seiten dieses oder jenes Kindes oder bekunden ihm ihre Gunst (Geschwisterrivalitäten).
22. Die Eltern erlauben dem Kind, Gegenstände, die der Familie gehören, uneingeschränkt und unkontrolliert zu benutzen.
23. Die Eltern zeigen wenig Mitgefühl, wenn dem Kind gehörende Gegenstände versehentlich zerstört werden (Beispiel: Lieblingspuppe, Buch, Spielzeug) (negativ).

III. Aufbau der „check lists“

Man beginnt im allgemeinen damit, nach Erfahrung und nach Sachkenntnis eine „check list“ zu entwerfen; man vermeidet somit, sich gleich zu Beginn von früheren Arbeiten zu stark beeinflussen zu lassen.

Dieser erste Entwurf muß sodann erweitert werden, unter dem Aspekt der bereits vorhandenen Literatur kontrolliert (deskriptive Untersuchungen, Analyse der Verfahren usw.) und mit anderen Forschern diskutiert werden. Man sollte auch nicht die häufig wertvollen Anregungen außer acht lassen, die die Befragten hinsichtlich des Beobachtungsgegenstandes geben.

Übrigens ist auch hier der Vortest angezeigt.

Während bestimmte „check lists“ nur Gedächtnisstützen sind, eignen sich andere dazu, echte Forschungs- oder Diagnoseinstrumente zu sein, sie können dann den Einsatz von derartig komplexen Techniken wie die der Faktorenanalyse erforderlich machen.

IV. Zuverlässigkeit und Objektivität

Wenn der Befragte selbst die Liste ausfüllt ist es wünschenswert, das Experiment nach einigen Tagen zu wiederholen. Wenn die Listen durch Interviewer vorgelegt werden, wird man berücksichtigen (Wrightstone, Justman, Robbins), daß mindestens drei Interviewer gleichzeitig, jedoch unabhängig voneinander, den gleichen Probanden prüfen sollen.

4. Die „rating scales“²⁴

I. Definition

Während man sich bei der „check list“ damit begnügt, das Vorhandensein oder Fehlen einer Sache oder eines Phänomens festzustellen, fügt die Beurteilungsskala dieser Beurteilung eine qualitative oder quantitative Einstufung hinzu und erlaubt somit eine Klassifizierung. Häufig jedoch gibt sie eher ein subjektives Urteil als einen objektiven Meßwert wieder. Aus diesem Grund können die Beurteilungsskalen nicht nur dazu dienen, die Personen oder die Sachverhalte besser kennenzulernen dank der Mitwirkung der Beurteiler, die ihr Urteil in einer vorgeschriebenen Form abgeben, sondern auch dazu, die Gefühle und Wertmaßstäbe der Beurteiler hinsichtlich der Einstufung, die sie vornehmen, zu überprüfen.

Wie dem auch sei, die Schätzskaalen sind nur ziemlich dürftige Instrumente; denn obwohl sie auch ein Einordnen erlauben, bringen sie doch keine genauen Angaben über die Intervallwerte, die die einzelnen Grade trennen.

II. Klassifizierung²⁵

A. Die graphische Skala

In ihrer einfachsten Form besteht die graphische Skala aus einer einfachen Linie, die alle Grade eines Kontinuums darstellt. Der Beurteiler markiert seine Bewertung durch ein Zeichen auf der Linie. Beispiel: Ist dieser Schüler in Ihrem Unterricht aufmerksam?

24 P. Pichot, *Les tests mentaux*, „Que sals-je?“, Paris, P.U.F. 1962, S. 13, übersetzt „rating scale“ mit Beurteilungsskala.

25 Einige Autoren, u. a. Wrightstone, Justman und Robbins, fügen den fünf von uns zitierten Typen folgende hinzu:

1. Die Technik der forcierten Wahlen, wobei der Beurteiler sich jedesmal zugunsten eines der Vorschläge, die ihm paarweise oder in Dreiergruppen vorgelegt werden, äußern muß usw.

Beispiel I.

A. Dieser Lehrer spricht gut.

B. Dieser Lehrer vermeidet überflüssiges Gerede.

Beispiel II.

A. Er nimmt die Bemerkungen der Schüler zu Kenntnis.

B. Er verfolgt seine Idee ohne Rücksicht auf die Reaktion der Klasse.

2. Die Rangordnungsmethode, wobei die Beurteiler aufgefordert werden, Eigenschaften, charakteristische Merkmale usw. in der Reihenfolge ihrer Vorliebe zu ordnen.

äußerst aufmerksam

sehr wenig aufmerksam

B. Die numerische Skala

Die Bewertung wird hier entsprechend eines zu Beginn festgelegten Wertmaßstabes durch eine Zahl ausgedrückt.

Beispiel: Gesamtbewertung der Schüler

1 = sehr gut; 2 = gut; 3 = befriedigend; 4 = ausreichend; 5 = nicht genügend

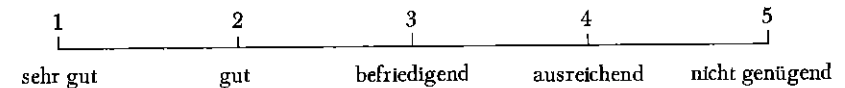
Unterstreichen Sie die zutreffende Note:

Fleiß	1	2	3	4	5
Verantwortungsgefühl	1	2	3	4	5
Betragen in der Schule	1	2	3	4	5
Führeigenschaften	1	2	3	4	5

C. Kombination von graphischer und numerischer Skala

Jeder oben angegebene Notenwert kann auch graphisch dargestellt werden.

Beispiel: „Fleiß“



Bemerkung: Die Beurteiler neigen im allgemeinen dazu, Extremwerte zu vermeiden und sich bei ihrer Beurteilung auf die Durchschnittswerte zu konzentrieren.

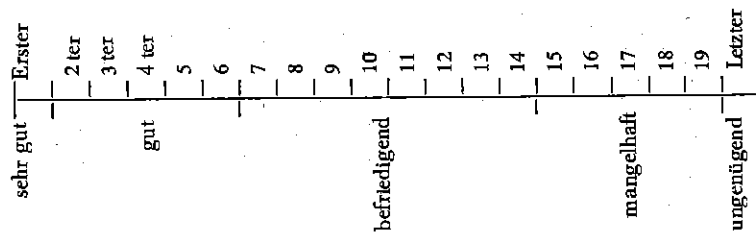
Man versucht, dieses zu verhindern, indem man:

1. die Hypothese aufstellt, daß die beobachteten Probanden sich nach einer Normalverteilung aufgliedern, und indem man die Skala entsprechend dieser Normalverteilung aufstellt,

Beispiel:²⁶

Allgemeine Beurteilung des Schülers: „Wo ordnen Sie den Schüler in einer Stichprobe von 20 Schülern ein, die alle auf der gleichen Schulstufe stehen und sich zusammensetzen aus einer Gruppierung, in der es einen sehr guten, fünf gute, acht mittelmäßige, fünf schwache und einen ungenügenden Schüler gibt?“

26 R. Pasquasy, *Une méthode de contrôle en orientation scolaire et professionnelle*, Brüssel, Editest, 1962, Annexe B.



2. die Beurteiler dazu auffordert, die Probanden auf dem Kontinuum einzu-
stufen unter Berücksichtigung der Proportionen einer Normalverteilung
(Abschnitt A wird 5%, B 25%, C 40%, D 25% und E 5% der Schüler um-
fassen) (Einstufung aufgrund einer vorgeschriebenen Verteilung).

D. Die deskriptive Skala

In ihrer einfachen Form ist die deskriptive Skala nur eine Aufstellung von
Beurteilungen, wie sie die Lehrer alltäglich abgeben.

Beispiel:

	Immer	Meistens	Manchmal	Niemals
— Paßt auf, wenn man ihn fragt				
— Drückt sich klar aus				
— Hat die Hausaufgaben sorgfältig gemacht				
usw.				

Neben einer solchen ziemlich simplen Skala gibt es jedoch auch komplexe
deskriptive Skalen, die eine exaktere Beurteilung erlauben. Um diese Exakt-
heit zu erreichen, ist es einerseits wichtig, daß der Untersuchungsgegenstand,
über den das Urteil abgegeben werden soll, eindeutig festgelegt ist, und
andererseits, daß die einzelnen Skalenstufen genau definiert sind.

1. Definition des Untersuchungsgegenstandes²⁷

Wenn man z. B. verlangt, ohne nähere Bestimmung den „Mut“ von Mit-

²⁷ Es liegt auf der Hand, daß für alle Skalen der Untersuchungsgegenstand eindeutig defi-
niert werden muß.

gliedern einer bestimmten Gruppe zu beurteilen, so führt das zu nahezu
völlig unbrauchbaren Antworten. Welche Bedeutung würde der Beurteiler
jeweils dem Wort „Mut“ geben: Standhaftigkeit, Unerschrockenheit, Gleich-
mut, Dreistigkeit, Zähigkeit, Draufgängertum, Waghalsigkeit?

Und selbst, wenn wir genau angeben, daß wir unter „Mut“ die Standhaftig-
keit vor der Gefahr verstehen, unterscheiden wir dann dabei auch zwischen
Beherztheit und der Tollkühnheit?

Um diese Ungenauigkeiten zu beseitigen, kann man der Skala eine mög-
lichst genaue Beschreibung des Gegenstandes oder der zu beurteilenden
Eigenschaft hinzufügen und die Definition mit Hilfe typischer Situations-
beschreibungen veranschaulichen.

So stellt Schonell²⁸ die Eigenschaft „Selbstvertrauen“ dar:

Selbstvertrauen

Äußerst star- kes Selbstver- trauen. Fast selbstsicher	Stark auf die eigenen Kräfte vertrauend	Selbstver- trauen	Mangel an Selbstver- trauen, schüchtern	Äußerster Man- gel an Selbstver- trauen. Hängt vom Urteil an- derer ab. Lehnt Verantwortung ab.
---	---	----------------------	--	---

Eigenschaftsbeschreibung

„In ihrer positiven Form wird diese Eigenschaft folgendermaßen gekennzeichnet:
Die Person baut auf sich selbst, ist in der Lage, Schwierigkeiten entgegenzutreten,
ist selbstsicher, selbständig und bereit, Verantwortung zu übernehmen.“

„Das Kind mit Selbstvertrauen versucht, mit einem Minimum an Unterstützung
voranzukommen; dasjenige ohne Selbstvertrauen bedarf ständiger Hilfe. Das
erstere möchte gerne ausprobieren, was es selber schaffen und leisten kann, wenn es
klare Anweisungen bekommen hat, das zweite möchte, daß man ihm die Aufgabe
erleichtert, daß man ihm ständig bei der Ausführung hilft.“

Typische Erprobungssituationen

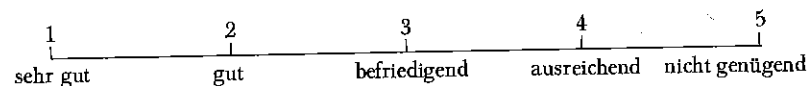
1. Hat es Angst vor der Dunkelheit?
2. Kann es auf sich selbst und seine Sachen achtgeben oder muß ständig jemand
aufpassen?
3. Fährt es allein in der Straßenbahn oder im Bus? (für Kinder über 9 Jahre)
4. Spricht es unbefangen mit fremden Besuchern?

²⁸ F. J. Schonell, Backwardness in the Backward Subjects, zitiert von F. Warburne,
Measurement of Personality (Educational Research, November 1961, S. 9).

5. Spielt es gut? Kann es schwimmen?
6. Antwortet es ungezwungen und sicher bei mündlichen Prüfungen?
7. Liest es gut, trägt es einen Text vor der Klasse lebendig vor?
8. Nimmt es bereitwillig neue Aufgaben in Angriff oder richtet es ständig Fragen an seine Mitschüler und Lehrer?

2. Definition der Skalenwerte

Beispiel: Organisation des Lesunterrichts²⁹



- 1 - Alle Schüler lesen den Text gemeinsam. Keine Gruppenarbeit.
- 2 - Wie 1, aber manchmal erhält ein sehr langsamer Schüler einige Aufgaben weniger als die anderen.
- 3 - Aufstellung von zwei oder drei Gruppen nach den erzielten Leseleistungen. Wenig Beweglichkeit in der Gruppierung.
- 4 - Gruppierung nach Leistungen. Beweglichkeit in der Gruppierung.
- 5 - Gruppierung nach eingehender Prüfung der Lesefähigkeit und der auftretenden Schwierigkeiten. Große Beweglichkeit in der Gruppierung.

Dieses letzte Beispiel kombiniert, wie man sieht, die graphischen numerischen und deskriptiven Skalen.

E. Die konkrete Skala

Anstatt der verschiedenen Grade der Skala — siehe oben — *ad abstractum* festzulegen, kann man die Beurteiler auffordern, die Einstufung nach der Ähnlichkeit mit bestimmten Bezugspersonen oder -objekten vorzunehmen:

1. *Man-to-man-rating*: Es ist anzugeben, welchem der fünf Schüler, die als Vergleichspersonen verwendet werden, die verschiedenen Mitglieder einer Gruppe am ähnlichsten sind.
2. Skalierung der in der Schule hergestellten Werkstücke:
Eine Bewertung der im Werkunterricht angefertigten Gegenstände ist anhand einer bestimmten Anzahl von Vergleichsmodellen vorzunehmen (*products rating scale*).

²⁹ Guide for Evaluating and Improving Nebraska Elementary Schools, 1955, S. 90.

3. Skalierung von Mustern:

Die Beurteilung ist durch Vergleichen mit anderen Mustern vorzunehmen (*scaled specimens*).

- a) Beurteilung des Entwicklungsniveaus nach der zeichnerischen Darstellung einer Person (durch Vergleichen mit einem Muster, das als typisch für eine bestimmte Entwicklungsstufe angesehen wird).
- b) Handschriftskala von R. Piscart (abgestufte Handschriftenmuster dienen als Beurteilungsgrundlage)^{29a}.

III. Der Aufbau der Beurteilungsskala

Viele Forscher haben sich durch die scheinbare Leichtigkeit, mit der man eine Beurteilungsskala erstellen kann, zu deren Mißbrauch verleiten lassen. Tatsächlich ist oftmals ein gründliches Studium notwendig, um den Untersuchungsgegenstand und die Skaleneinteilung festzulegen.

Es ist darüber hinaus schwierig, die charakteristischen Aspekte eines komplexen Phänomens oder die persönlichen Faktoren, die eine bestimmte Situation entscheidend beeinflussen, richtig zu identifizieren.

In bestimmten Fällen, ist die Anwendung so komplexer Techniken, wie z. B. die Faktorenanalyse, notwendig.

IV. Schwierigkeit bei der Beurteilung

A. Begrenzte Anzahl der verwendeten Skalengrade

Theoretisch kann eine Beurteilungsskala unbegrenzt viele Unterteilungen haben. Aber die Beurteiler unterscheiden dagegen nur eine relativ niedrige Zahl. Miller³⁰ schätzt, daß man kaum neun Gruppen überschreiten kann, und einige andere Autoren, u. a. Guilford, stimmten mit ihm darin überein. In den meisten Fällen beschränkt man sich auf fünf bis sieben Unterteilungen.

^{29a} R. Piscart, Echelle objective d'écriture pour écoliers belges, Louvain Nauwelaerts — Paris, Vrin, 1950.

³⁰ G. A. Miller, The magical number seven, plus or minus two: some limits of our capacity of processing information (Psychological Review, 1956, Nr. 63, S. 81—97).

B. Beeinflussung durch die Persönlichkeit der Beurteiler

Trotz mancherlei Vorkehrungen, die man getroffen hat, bleibt es dennoch stets schwierig, den Einfluß des persönlichen Geschmacks und der persönlichen Vorurteile auszuschalten. Dies macht sich um so deutlicher bemerkbar je unbestimmter die Skalengrade sind (ausgezeichnet, sehr gut usw.). Auch bei einer systematischen Untersuchungsarbeit präzisiert man nicht nur die Gruppen (siehe oben!) soweit als möglich, sondern zieht auch eine große Anzahl von Beobachtern mit heran.

C. Die Tendenz zur Mitte

Wir haben bereits darauf hingewiesen, daß die Beurteiler dazu neigen, die extremen Skalenwerte zu vermeiden und die Skala dadurch zu reduzieren. Da diese Ungenauigkeit nicht bei allen gleich stark ausgeprägt ist, kommt man schließlich zu verschiedenen Bewertungsskalen. Um die Bewertungen vergleichbar zu machen, kann man — wenn die beobachteten Phänomene komplex genug sind, um sich nach dem Gesetz der Normalverteilung zu verteilen — auch Standardwerte berechnen (siehe statistischer Teil!).

D. Die Beeinflussung der Angaben (contamination)

Man versteht darunter den Einfluß, den die Kenntnis von bestimmten Resultaten auf spätere Bewertungen ausübt, die eigentlich unabhängig davon abgegeben werden sollten.

E. Stereotype und Halo Effekt

Ein Stereotyp ist — so könnte man sagen — eine permanente Beeinflussung. Zum Beispiel: Nach einigen Erfahrungen oder Beobachtungen auf einem relativ begrenzten Gebiet beurteilt ein Lehrer (nicht immer bewußt) einen Schüler als einen insgesamt gesehen guten, schlechten oder mittelmäßigen Schüler. Dieses Urteil wird schwerlich revidiert.

Der Halo Effekt ist eine stereotype Reaktion vorwiegend affektiver Natur: Ein artiges Kind, höflich, hübsch aussehend, wird häufig für intelligenter gehalten als es in Wirklichkeit ist; das Gegenteil ist der Fall, wenn es sich um ein unartiges, undiszipliniertes Kind handelt.

Mittel und Wege, die Beeinflussung, die stereotypen Ungenauigkeiten und die Halo Effekte zu reduzieren, sind u.a. folgende:

- a) Man läßt einen hinreichend langen Zeitraum (mehrere Stunden oder sogar mehrere Tage) zwischen der Bewertung der einzelnen Eigenschaften bei ein- und demselben Prüfling verstreichen. Diese Vorkehrung wird im allgemeinen als wesentlich angesehen;
- b) Man schaltet eine möglichst große Anzahl von Beobachtern ein;
- c) Man verwendet nicht nur Skalen, die in einer Richtung, z. B. von schlecht nach gut, aufgebaut sind, sondern teilt nach Zufallsprinzipien auf.

V. Anwendung

Die Beurteilungsskala wird unter zwei verschiedenen Aspekten angewendet:

1. Um Beobachtungen möglichst schnell festhalten zu können (z. B. vom Schulrat, der einer Unterrichtsstunde beiwohnt; bei der Verhaltensbeobachtung der Kinder während der Pause usw.).
2. Um die Bewertungen im Verlauf der Untersuchungen zu systematisieren. Ohne noch länger bei den zahlreichen bereits erwähnten Möglichkeiten für Ungenauigkeiten zu verweilen, glauben wir, daß man zumindest im zweiten Fall mit einer Art „photographischen“ Registrierung der Tatbestände und Verhaltensweisen beginnen sollte, und zwar, indem man zunächst versucht, die Wirklichkeitssituation möglichst genau festzuhalten, und erst anschließend wertet. Diese Methode bietet folgende Vorteile:
 - a) Die Beobachtung wird fixiert, was eine Bewertungskontrolle erlaubt.
 - b) Man kann somit der Wechselwirkung und der Komplexität der Verhaltensweisen sowie der Dauer der Interaktionen besser Rechnung tragen.
 - c) Die Registrierung der Verhaltensweisen ist leichter wertfrei vorzunehmen als bei direkter Bewertung.
 - d) Da die Registrierungen weiterhin bestehen bleiben, kann man sie auf verschiedene Arten und nach unterschiedlichen Gesichtspunkten untersuchen.

VI. Eine spezielle Anwendung^{a)}

Beurteilungsskala als Quartalzeugnis

GEWOHNHEITEN UND HALTUNG

Erklärung der Zeichen, die die Einzelfortschritte anzeigen:

- (1) sehr gut - sehr fleißig
- (2) befriedigend - fleißig, bemüht sich, versucht, die in ihn gesetzten Erwartungen zu erfüllen
- (3) Müßte sich bessern - könnte mehr leisten aufgrund seiner Befähigung

Arbeitsgewohnheiten

Beginnt unmittelbar und führt die Arbeit bis zu Ende durch
 Arbeitet gut mit seinen Mitschülern zusammen
 Befolgt gut die gegebenen Anweisungen
 Arbeitet akkurat - sauber
 Behandelt seine Bücher und sein Unterrichtsmaterial sorgfältig
 Gestaltet seine Freizeit sinnvoll
 Zeigt eigene Initiative
 Arbeitet schnell

Soziale Einstellungen

Hört aufmerksam zu, wenn seine Mitschüler vor der Klasse sprechen
 Beweist eine faire Haltung
 Höflich in seinen Worten und Handlungen
 Versteht und respektiert die Regeln der Disziplin
 Beweist Führerqualitäten
 Übernimmt Verantwortung bei der Pflege und Ordnung der Klasse

Schulhaushalt

Zahl der versäumten Halbtage
 Verspätungen

UNTERRICHTSFÄCHER

Erklärung der Zeichen

- Individuelle Fortschritte: Siehe oben!
 Leistungen im Vergleich zu (den Leistungen) der gesamten Klasse
 (I) überdurchschnittlich
 (II) durchschnittlich
 (III) unter dem Durchschnitt liegend

Individuelle Fortschritte		Leistungen im Vergleich zu der gesamten Klasse
	Lesen	
	Liest mit Verständnis	
	Findet den Sinn von unbekanntem Wörtern und verwendet sie	
	Bildet sich durch persönliche Lektüre weiter	
	Geschichte - Geographie	
	Sachkenntnisse	
	Verwendet das Unterrichtsmaterial und teilt es mit den anderen	
	Respektiert die Lebensweise und Gefühle anderer und bringt ihnen Verständnis entgegen	
	Mündliche und schriftliche Ausdrucksweise	
	Spricht gut vor einer Gruppe	
	Hat eine deutliche Aussprache	
	Drückt seine Ideen deutlich in Worten aus	
	Drückt sich im schriftlichen deutlich aus	
	Orthographie	
	Orthographische Übungen	
	Orthographie bei anderen schriftlichen Arbeiten	
	Handschrift	
	Schreibt sauber und leserlich	
	Rechnen	
	Kenntnisse im Rechnen	
	Anwendung auf erlebte Situationen	
	Genauigkeit beim Rechnen	
	Naturkunde	
	Qualitäten als Beobachter	
	Kenntnis der wissenschaftlichen Begriffe	
	Music	
	Malen	
	Handarbeit/Werkarbeit	

a) In Anlehnung an den *Niagara Falls Report, Grades 4 - 5 - 6*, siehe R. M. THOMAS, *Judgin Progress, New York, Longmann, Green & Co., 2. Aufl., S. 382 - 383.*

5. Die Technik der kritischen Ereignisse (critical incidents)

I. Definition

Nach der Methode der kritischen Ereignisse versucht man, die funktionellen Anforderungen (z. B. diejenigen des Gymnasiallehrers) zu bestimmen, indem man die „Ereignisse“ untersucht, die auffallenden Züge, die speziellen Verhaltensweisen, die dem Beobachter auffallen und die bemerkenswerten Erfolge oder Fehlleistungen zu erklären scheinen.

Die Methode kann zu zwei verschiedenen Zwecken angewendet werden:

- Um objektiv zu klären, welche die charakteristischen Verhaltensweisen der Lehrer, Schüler, Eltern usw. sind, die als gut oder schlecht angesehen werden.
 - Um den Gedankengang der Beurteiler zu analysieren (z. B. der Schullehrer) und in Erfahrung zu bringen, worauf sie ihre Bewertung gründen.
- J. Flanagan scheint der erste gewesen zu sein, der diese Technik in der Pädagogischen Forschung angewendet hat³¹.

Die Methode der kritischen Fälle ist bisher wenig untersucht worden. Trotz ihrer Mängel, die wir in der Folge aufzeigen werden, scheint es, daß sie lehrreich sein könnte. Es wäre zu wünschen, daß sie Gegenstand eingehender Untersuchungen würde.

II. Anwendungsbeispiele

Wir verdanken N. Limbosch einen interessanten Beitrag zur Methode der kritischen Fälle für die Analyse der Funktion des Erziehers³².

Wir zitieren die sechs von N. Limbosch an Schulleiter gerichteten Fragen, weil sie ein klares Bild von der allgemeinen Technik wiedergeben.

31 Siehe J. C. Flanagan, *La Technique de l'Incident Critique (Revue de Psychologie appliquee, IV, Nr. 2, 1954)*.
 Siehe auch: A. C. Jensen, *Determining critical requirements of teachers (Journal of Experimental Education, XX, 1951-1952, S. 79-85)*.
 32 N. Limbosch, *Analyse de la fonction d'instituteur par la methode des incidents critiques (Revue Belge de Psychologie et de Pédagogie, XVIII, Nr. 75, Sept. 1956, S. 69-87)*.

1. „Denken Sie an den besten unter ihren Lehrern, deren Arbeitsweise Sie seit mindestens drei Monaten überwachen konnten, insbesondere an die letzte Tätigkeit, die Ihre Aufmerksamkeit angezogen hat, weil Sie in beruflicher Hinsicht wirklich beachtlich war.“ „Geben Sie nur genau an, was er zu jenem Zeitpunkt gemacht hat.“
2. Ähnliche Fragen über den „schlechtesten“ Lehrer.
3. „Denken Sie an den Lehrer, dessen Arbeitsweise Sie seit mindestens drei Monaten überwachen konnten, auf den Sie ein besonderes Augenmerk hatten, weil Sie spürten, daß er es in beruflicher Hinsicht zu ausgezeichneten Ergebnissen bringen kann.“ „Geben Sie genau an, was er gemacht hat oder was er mit Ihrer Unterstützung zu jenem Zeitpunkt bereits imstande war zu leisten.“
4. Ähnliche Frage über den Lehrer, der im Begriff war „einen großen Fehler“ zu begehen.
5. „Es ist wahrscheinlich, daß sich unter denjenigen Ihrer Lehrer, die Sie seit mindestens drei Monaten überprüfen konnten, der eine oder andere befindet, der diese oder jene Sache gemacht hat, die in beruflicher Hinsicht einen schönen Erfolg darstellt.“ „Geben Sie mir genau das letzte oder einige der letzten Beispiele dieser ausgezeichneten Leistungen an, die Sie beobachtet haben. Es ist nicht nötig, daß diese von einem allgemein als sehr gut beurteilten Mitarbeiter vorgebracht werden.“
6. Ähnliche Fragen über die „Mißerfolge“ oder „Schnitzer“.

Im Verlauf des Interviews von fünfzehn Volksschulrektoren im Großraum Brüssel notierte N. Limbosch:

- 88 positive kritische Ereignisse (charakteristisch für den Erfolg).
- 51 negative kritische Ereignisse (charakteristisch für den Mißerfolg).

Die Analyse und die Strukturierung der gesammelten Angaben zeugen davon, welche technische Hilfestellung diese Methode bietet.

Der amerikanische „Council on Education“ hat gleichfalls die Technik der kritischen Ereignisse benutzt, um „das gesamte beobachtungsfähige Verhalten der Erzieher, das den Erfolg oder Mißerfolg im Unterrichten erklären könnte“³³, zu erkennen.

Die kritischen Anforderungen wurden schließlich in drei Gruppen klassifiziert:

1. Personelle Qualifikationen:
 - a) Optimismus
 - b) Gerechtigkeit
 - c) Selbstkontrolle
2. Professionelle Qualifikationen:
 - a) Kenntnis von Unterrichtsgegenstand und -methode
 - b) Fähigkeit, den Schüler zum selbständigen Lernen anzuleiten
 - c) Loslösen von persönlichen Nebenbeschäftigungen (business-like approach).

3. Soziale Qualifikationen:

- a) Sympathie
- b) Demokratische Einstellung
- c) Liebenswürdige, ermutigende Haltung
- d) Einfühlungsvermögen in die Reaktionen anderer.

Schließlich zeigt D. G. Ryans^{33a}, wie die Technik der kritischen Ereignisse angewendet wurde, um eine objektive Ausarbeitung des „Classroom Observation Record“ sicherzustellen, einer Serie von Skalen zur Untersuchung der charakteristischen Merkmale von Erziehern^{33b}.

III. Kritik an der Methode

1. Man stützt sich auf mitunter wenig übliche und daher schwierig systematisch zu beobachtende Verhaltensweisen.
2. Oftmals erinnern sich die befragten Personen nur an einen Vorfall, weil er außergewöhnlich war. Es kann gefährlich sein, diese Ausnahmefälle als repräsentativ zu betrachten.
3. Die gesammelten Vorfälle sind oftmals unzusammenhängend und daher schwer auszuwerten.

6. Andere Techniken

I. Protokollierung der Beobachtungen

A. Kodierung

Die beobachteten Verhaltensweisen sind häufig derart fließend, daß die Untersucher gezwungen sind, eine Kodierung für die Protokollierung der Beobachtungen auszuarbeiten. Es existieren keine allgemein anerkannten Ver-

³³ K. M. Evans, Research on Teaching Ability (Educational Research Bd. I, Nr. 3, Juni 1959).

^{33a} D. G. Ryans, Characteristics of Teachers, Washington, American Council on Education, 1962, 2. Aufl., S. 79–92.

^{33b} In Deutschland hat Hans Meister die Methode angewendet, um das Schulklassenklima zu erkennen. Man siehe: J. P. Ruppert, Die seelischen Grundlagen der sozialen Erziehung, Band III, Weinheim, Julius Beltz Verlag, 1965, S. 305–344.

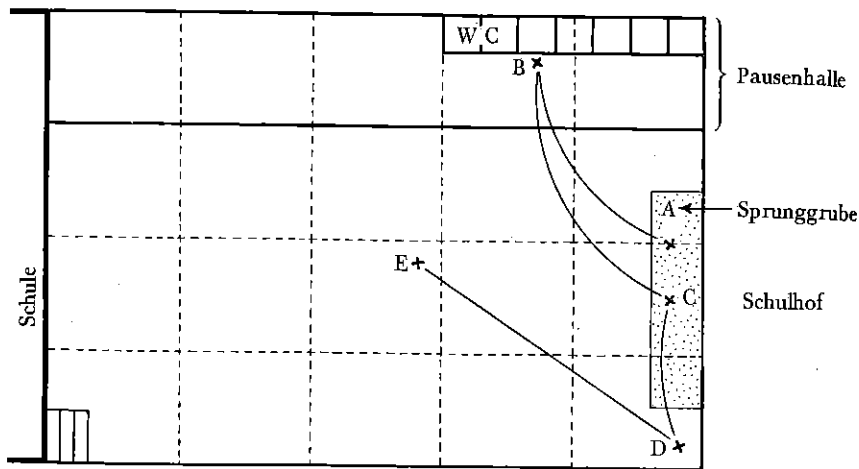
fahrenswesen auf diesem Gebiet. Jeder schmiedet sich demnach seine Instrumente selbst, die ihm für die Erfordernisse seiner Arbeit am geeignetsten erscheinen.

Zeichenbeispiele, die sich verallgemeinern lassen:

- ↔ Gerät in Konflikt;
- ➔ macht einen konstruktiven Vorschlag;
- ? verlangt eine Erklärung;
- ?? versteht nicht;
- ↓ gibt die Arbeit, die Diskussion auf.

B. Skizzen und graphische Darstellungen: zwei Beispiele

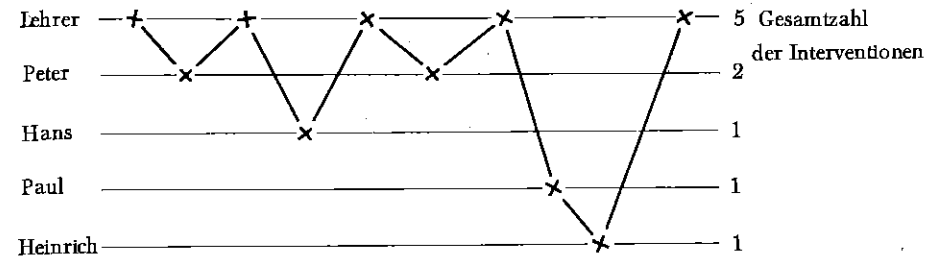
1. Beobachtung der Platzveränderung eines Kindes während der Pause³⁴
Den Untersuchern steht zu Beginn der Beobachtung ein hinreichend großer Schulhofplan zur Verfügung. Aus Gründen der besseren Übersicht ist dieser Plan in Quadrate aufgeteilt, und die Quadrierung wird mitunter so unauffällig wie möglich auf dem Boden nachgezeichnet.



- A. spielt mit Paul.
- C. kommt von der Toilette zurück, zerstört die Sandburg von Peter. Peter schlägt ihn.
- D. weint für sich allein.
- E. gesellt sich zu Leo und Ludwig, die mit Murmeln spielen.

³⁴ Nach D. Thomas et al., Some New Techniques for Studying Social Behavior, New York, Teachers College, Columbia Univ., 1929.

2. Graphische Darstellung einer Klassendiskussion



Jede Beteiligung an der Diskussion wird durch ein Kreuz angezeigt. Jedes neue Kreuz wird mit dem vorhergehenden verbunden.
Am Schluß der Beobachtung zählt man, wie oft jeder Schüler sich an der Diskussion beteiligt hat. Das Profil läßt die allgemeine Entwicklung der Diskussion erkennen. Das obige kurz gefaßte Beispiel zeigt die entscheidende Rolle des Lehrers, der außer einem Mal (Paul—Heinrich) nach jeder Intervention eines Schülers das Wort ergreift.

II. Die Zeitanalyse (time schedule)

Der Erzieher weiß im allgemeinen sehr wenig über die Zeiteinteilung der Schüler außerhalb des Schulunterrichts.

Es ist jedoch von größter Bedeutung zu wissen, wie sich ein Tag, und wenn möglich, eine ganze Woche eines Kindes abwickelt — das Wochenende einbezogen — (Zeit, die mit Spielen oder Lesen zugebracht wird, Freizeitgestaltung ganz allgemein, Zahl der Arbeitsstunden für Hausaufgaben usw.).

Wenn man die Mitarbeit der Eltern oder der Prüflinge selbst erwirkt, scheint das folgende Verfahren am meisten zu befriedigen. Für jeden Tag stellt man einen Zeitplan auf und macht alle 15 bzw. 20 Minuten Notizen, vom Aufstehen bis zum Zubettgehen. Bei längeren Zeitabständen wird die Beschreibung zu allgemein.

Die Beobachtung wird je nach Bedarf wiederholt.

III. Time Sample und Zufallsbeobachtungen

Um sicherzustellen, daß die Beobachtungszeit für jeden Prüfling einer Gruppe ausreichend und gleich lang ist, hält man die Beobachtungsdauer konstant, wobei man möglichst viele, kurzfristige Beobachtungen durchführen sollte. Ihre Reihenfolge überläßt man dem Zufall. Man erhält somit eine Reihe von Stichproben kleiner Zeiteinheiten (*time samples*).

In Anlehnung an Arrington schätzt Anastasi³⁵, daß man, — von 24 Beobachtungen von je 5 Minuten ausgehend — ein relativ konstantes Ergebnis erhält.

Neben dieser systematischen Arbeitsmethode bringt die *zufällige Beobachtung* wertvolle Informationen, weil es sich dabei im allgemeinen um ausgesprochen spontane Verhaltensweisen handelt, bei denen der Erzieher unvorhergesehener Zeuge ist. Derartige Beobachtungen werden in anekdotischer Form (*Zufallsbeobachtung, anecdotal record*) festgehalten.

IV. Instrumentelle Ausrüstung

Es ist unmöglich, eine komplette Aufstellung über die in der Pädagogischen Forschung verwendeten Instrumente zu geben. Wir beschränken uns auf einige allgemeine Indikationen.

A. Photographie und Film

- Versteckte Kameras;
- Chronophotographie; Man berechnet die Zeitdauer der Bewegungen oder Handlungen in Relation zu der Geschwindigkeit, mit der das Filmband abrollt.
- Mit Hilfe einer Infrarotlampe und Spezialfilmen ist es möglich, im Dunkeln zu photographieren. Dieses Verfahren wird vor allem angewendet, um die Spontanreaktionen der Kinder während einer Filmvorführung zu beobachten.
- Ophthalmographie: Kamera (Spezialkonstruktion) zur Registrierung der Augenbewegungen während des Lesens: Stillstehen, Hin- und Hergehen, kontinuierlicher Rhythmus, Lesegeschwindigkeit und Augenkoordinierungen.

35 Anastasi, *Psychological Testing*. New York, Macmillan, 1961, 2. Aufl. S. 533.

B. Tonbandgeräte und Mikrophone

Die Installierung von hochwertigen Mikrofonen an Wänden und Decken der Untersuchungsräume und Experimentalschulen gestattet eine Registrierung durch versteckte Beobachter.

C. Spiegelfenster (*One-way mirrors*)

Spiegel, die auf der einen Seite reflektieren und auf der anderen durchsichtig sind. Man setzt sie in zunehmendem Maße zur Einrichtung von Beobachtungszellen ein, die z. B. auf die ganze Länge einer Klasse verteilt sind. Auf diese Weise können etwa ein Dutzend Studierende beobachten, ohne die Arbeit und die Atmosphäre der Gruppe zu stören.

D. Lehrmaschinen

Hiervon gibt es bereits eine beachtliche Anzahl. Die meisten dieser Maschinen eignen sich für Untersuchungen des Lernvorgangs und des Lehrverfahrens³⁶.

E. Lesebeschleuniger (*Science Research Association, Chicago*)

Dieser Apparat erlaubt die Messung der Maximalgeschwindigkeit, mit der ein Proband einen vorgelegten Text lesen kann. Eine Schablone schiebt sich von oben nach unten über eine Seite und verdeckt jeweils eine Zeile; die Geschwindigkeit kann zwischen 500 und 3 000 Wörtern pro Minute variieren. Der Lesebeschleuniger wird in zunehmendem Maße angewendet, um Schüler im schnellen Lesen zu schulen.

F. Psychotechnische Apparate³⁷

Impulszähler (bei zahlreichen Tests angewendet zur Registrierung von Fehlern;
Chronometer;

36 Wilbur Schramm, *Programmierter Unterricht, heute und morgen*, Berlin-Bielefeld, Cornelsen Verlag, 1963.

Anm. d. dtsh. Bearbeiters: Vgl. auch die vom Pädag. Zentrum Berlin veröffentlichte Dokumentation zum *Programmierten Unterricht*.

37 Viele der Apparate stellen ganz spezielle Tests dar.

Chronoskop (Messen der Reaktionszeit auf visuelle oder akustische Reize);
Spirometer;
Dynamometer;
Registriergerät für komplexe Reaktionen von Bonnardel;
Elektronisches Tachistoskop;
Projektionstachistoskop;
usw.

Abschnitt II:

Die Tests und die Spezialtechniken

1. Die Tests

Es gibt zur Zeit etwa 9–10 000 standardisierte Tests auf der ganzen Welt. Ein Werk wie das vorliegende kann in einem Kapitel diesen umfangreichen Stoff selbstverständlich nicht hinreichend behandeln^{37a}.

Wir müssen uns daher auf eine Definition des Tests und seines Aufbaus beschränken. Eine allgemeine Klassifikation vermittelt sodann eine erste Orientierung^{37b}.

I. Definition

Ein standardisierter Test ist ein Untersuchungsverfahren, das folgenden Anforderungen Rechnung trägt:

1. Der Gegenstand und der Schwierigkeitsgrad der Fragen werden systematisch geprüft (Aufgabenanalyse).
2. Die Testdurchführung und die Korrektur sind so einheitlich wie möglich durchzuführen (Standardisierung im eigentlichen Sinne).
3. Die Klassifizierung wird im Hinblick auf Normen vorgenommen, die aus der Vorprüfung einer mehr oder minder großen Zahl von Probanden hervorgegangen sind, was erlaubt, die Antworten (ganz oder teilweise) in eine statistische Verteilung einzuordnen (Normierung).
4. Die Antworten auf die gestellten Fragen sollen ein genaues Maß für die in der Untersuchung angegangenen Aspekte oder Phänomene geben (Validität = Gültigkeit).
5. Unter gleichbleibenden Bedingungen muß die Wiederholung der Prüfung stets zu den gleichen Ergebnissen führen (Zuverlässigkeit = Reliabilität).

37a Die vollständigste Referenzquelle ist: O. Buro, Ed., Mental Measurements Yearbook, Highland Park, N. J. Gryphon Press 1938–1941–1949–1953–1959–1965.

37b Anm. d. dtsh. Bearbeiters: Für eine gründliche Behandlung vgl. auch: G. A. Lienert, Testaufbau und Testanalyse, Weinheim und Berlin, J. Beltz, 1967, 2. Auflage.

II. Testaufbau

A. Allgemeines Vorgehen

1. Voruntersuchung

a) Theoretische Grundlage

R. Zazzo schreibt: „Ein Test ist ganz einfach die Formgebung einer ihm vorausgegangenen Erkenntnis. Er ist ein „Modell“, das eine „Feststellung“ erlaubt. Ein Punkt, das ist alles. Wenn die Feststellung nichts aussagt, bedeutet dies, daß das Modell falsch oder inadäquat ist, daß der zugrunde liegende Gedanke irrig oder illusorisch ist. Wenn man von einer falschen Hypothese ausgeht, entbehrt die Schlußfolgerung selbstverständlich jeglichen Wertes³⁸“.

Man sollte nicht allzu sehr auf diesem Punkt bestehen. In der Vergangenheit hat man oftmals kunstvoll ausgeklügelte Konstruktionen auf schlecht definierten Konzeptionen aufgebaut. So kann man, falls man sich damit begnügt, dem Wort „Trägheit“ seine eingebürgerte Bedeutung beizumessen, einen „Trägheitstest“ konstruieren, und, von diesem ausgehend, großartige Häufigkeitsverteilungen ausarbeiten, Indizes errechnen und Quotienten bilden. Aber alle diese Bemühungen laufen letztlich auf die strenge Quantifizierung eines falschen oder jedenfalls derartig ungenauen Urteils hinaus, daß es nichts aussagt³⁹.

b) Definition der allgemeinen Ziele

Will man einen Leistungstest z. B. für das Ende der höheren Schule konstruieren, so lassen sich die allgemeinen Ziele auf zweierlei Arten definieren:

38 R. Zazzo, *L'examen psychologique de l'enfant*, Neuchâtel, Delachaux und Niestlé, 1960, S. 9 (R. Zazzo stellt hier implizit die Hypothese auf, daß der Konstrukteur keinen technischen Fehler begangen hat).

39 Vor Anwendung eines Tests ist es auch wichtig, seine Grundbegriffe zu untersuchen. Ohne diese grundsätzliche Prüfung ist das Vorgehen der Prüfer ebenso unwissenschaftlich wie z. B. das der „Wunderdoktoren“ von früher, für die der Erfolg nichts weiter als ein glücklicher Zufall war. Man übernimmt kein Schulbuch, ohne es nicht vorher kritisch geprüft zu haben. Eine ähnliche Vorgehensweise ist noch wichtiger bei den Tests, die vielfach theoretisch nicht hinreichend gerechtfertigt sind. Denn grundsätzliche Fehler verbergen sich leicht hinter der „Mechanik“ der Prüfungen. Die ungenauen Statistiken (nicht streng durchgeführte Eichungen, unzureichende Validierung, die sich hinter Korrelationen verbergen, deren Bedeutung zu früh verallgemeinert werden usw.) runden das irreführende Bild ab.

1. Man versucht, unabhängig vom Lehrstoff die Fähigkeiten und Hauptkenntnisse zu präzisieren, die ein herangewachsener Jugendlicher entwickelt bzw. sich angeeignet haben sollte, der gerade im Begriff ist, einen Beruf zu erlernen oder sich dem Hochschulstudium zuzuwenden.
2. Man geht von dem Lehrstoff aus und versucht, die wichtigsten Ziele freizulegen^{39a} und beschließt, den Test im Hinblick auf eine bestimmte Anzahl dieser Ziele zu konstruieren.

Beispiel: Die allgemeinen Ziele des naturwissenschaftlichen Unterrichts sind:

- Den Beobachtungssinn und die Kritikfähigkeit durch das Studium der Phänomene und durch das Experiment zu entwickeln (I);
- Einen gewissen Grundstock an faktuellen Kenntnissen zu erwerben (II).

c) Definition der speziellen Ziele

Für jeden allgemeinen Punkt des naturwissenschaftlichen Unterrichts, auf den wir gerade angespielt haben, kann man überprüfen, ob das Lernen der einzelnen Programmpunkte erreicht worden ist: Diese sind die speziellen Testziele.

In der folgenden Tabelle hat jedes Einzelkreuz als Koordinaten das erste allgemeine Ziel und einen Unterrichtsprogrammpunkt; jedes Doppelkreuz entspricht dem allgemeinen Ziel II und einem Unterrichtsprogrammpunkt. Das erste Spezialziel wird demnach sein, zu verifizieren, ob der Beobachtungssinn und die Kritikfähigkeit beim Studium der mechanischen Energie (Ziel I/1) entwickelt worden sind.

Naturwissenschaft: Physik

Unterrichtsprogrammpunkte	Generelle Ziele		
	I	II	
1. Mechanische Energie	x	xx	
2. Gravitation	x	xx	
3. Elektrische Energie	x	xx	
4. Statische Elektrizität	x	xx	
5. Magnetismus	x	xx	Spezielle Ziele
6. Chemische Energie	x	xx	
7. Säuren	x	xx	
8. Basen	x	xx	
usw.			

39a Zur Feststellung der allgemeinen Ziele sollte man auch nachforschen bei: B. S. Bloom, *Taxonomy of Educational Objectives*, o. c.

Natürlich ist es weder immer notwendig noch möglich (die Zahl der Test-items (Aufgaben) sollte nicht zu hoch sein, und die Prüfung darf nicht zu lange dauern), für jeden Unterrichtsprogramm-punkt zu verifizieren, ob alle generellen Ziele erreicht worden sind.

d) Bestimmung der relativen Bedeutung der verschiedenen Ziele

Man kann der Meinung sein, daß Ziel I/1 wichtiger ist als Ziel II/1 und sich z. B. dazu entschließen, dem ersten doppelt soviel an Gewicht beizumessen wie dem zweiten, eine Vorkehrung, die sich entweder durch eine doppelte Anzahl von relevanten Items von I/1 äußert oder durch die Zuordnung eines Punktwertes, der für Item I/1 zweimal höher ist (Gewichtung der Items).

e) Konsultieren von Experten

Im allgemeinen wurden schon Experten zu Rate gezogen, bevor das Endstadium (d) erreicht war. Ist das nicht geschehen, so ist für den Autor der Zeitpunkt gekommen, seine Schlußfolgerungen seinen Kollegen bzw. den Experten zu unterbreiten. Die Beweggründe für diese Vorkehrung sind allzu evident, als daß wir noch darauf eingehen müßten.

2. Vorbereitung und Korrektur der ersten Itemform

a) Erstmalige Abfassung der Items

Es handelt sich darum, eine ausreichende Anzahl Fragen bzw. Items in die geeignetste Form zu bringen, und zwar nicht nur, um den zu erforschenden Unterrichtsgegenstand zu erfassen, sondern auch, um in der Folge eine Auswahl treffen zu können, da sich gewisse Items als geeigneter, klarer, bedeutungsvoller und leichter klassifizierbar erweisen als andere.

Die Items können mannigfaltige Formen annehmen:

Die wichtigsten sind:

1. Bei den Leistungs- oder Handlungstests: Die Zahl der möglichen Items ist praktisch ebenso groß wie die Zahl menschlicher Verhaltensweisen;

2. Bei den Papier-Bleistift-Tests:

● Nach der Art der Antwort:

- Frage mit nur einer einzig möglichen Antwort (häufig ist ein Wort einzusetzen);
- Mehrfachwahl-Antwort: Zwei Möglichkeiten: richtig — falsch; ja — nein; der Schönste — der Häßlichste; Wahl zwischen zwei vorge-

schlagenen grammatikalischen Formen usw.; Die Möglichkeit zweier Antworten räumt jedoch eine zu große Chance ein, die richtige Antwort zu treffen, selbst wenn man sie rein zufällig wählt.

Fünf mögliche Antworten: Dies ist die zur Zeit am meisten angewandte Lösungsform, denn sie reduziert sehr stark den Zufallseinfluß⁴⁰.

● Nach der auszuführenden Aufgabe^{40a}.

Unter mehreren vorgeschlagenen Wörtern das Synonym bzw. Antonym eines Reizwortes wählen; Analogien zwischen Begriffen oder Bildern entdecken; logische Schlüsse ziehen; ein Zeichen oder eine Form unter anderen ankreuzen; sortieren, klassifizieren; entschlüsseln bzw. verschlüsseln eines Codesystems; vervollständigen (Liste, Satz, Skizze, Figur, Konstruktion ...) eine Ordnung wiederherstellen (verworrene Sätze, Puzzlespiele) die fehlenden Teile erkennen; eine in einer anderen versteckten Form herausfinden; Reihen fortsetzen, interpretieren (Zahlen, Texte, Bilder, Zeichnungen ...); Zusammenstellen von Elementen, zwischen denen eine Beziehung besteht; anpassen, aus dem Gedächtnis reproduzieren, usw.

Die erste Abfassung der Items ist eine zeitraubende und schwierige Aufgabe, die immer häufiger von einer Forschungsgruppe und nicht mehr von einem einzelnen übernommen wird⁴¹.

b) Die Überprüfung der Items

Sie wird an einer möglichst repräsentativen Stichprobe aus der Population durchgeführt, für die der Test bestimmt ist.

Man sammelt so Hinweise:

- über den Schwierigkeitsgrad der Items (Tabelle der Häufigkeit der richtigen Antworten, Histogramm pro Frage): die zu leichten bzw. zu schwierigen Items werden gestrichen, und zwar nach genauen Kriterien (prozentualer Anteil der richtigen Antworten: z. B. mehr als 85 % richtige Antworten = zu leicht; weniger als 25 % richtige Antworten = zu schwierig);

40 Es gibt selbstverständlich noch andere mathematische Verfahren, die es erlauben, den Zufallseinfluß abzuschätzen bzw. zu reduzieren.

40a Siehe auch: A. Rey, *Connaissance de l'individu par les tests*. Brüssel, Dessart, 1963, S. 136ff.

41 Den Lehrern, die Tests für ihren eigenen Bedarf konstruieren wollen, wird empfohlen, während ihrer üblichen Tätigkeit mögliche Items zu entwerfen und zu notieren. Zum Zeitpunkt der eigentlichen Testkonstruktion wird der Lehrer somit in der Lage sein, auf diesen ersten und wertvollen Grundstock aufzubauen.

- über die Validität der Items: verworrene, zweideutige Items;
- über die Richtigkeit der gegebenen Instruktionen;
- über den Zeitpunkt der Durchführung, die Dauer der Korrektur und die möglichen Schwierigkeiten beim Notieren.

3. Erste experimentelle Testform

a) Erstellung

Der Test wird jetzt schon in eine strenge Form gebracht. Die Zahl der Items wird reduziert. Diese werden entweder nach dem zunehmenden Schwierigkeitsgrad oder in zyklischer Reihenfolge der Schwierigkeitsgrade⁴² geordnet.

b) Testdurchführung

c) Analyse

- Tabelle der Häufigkeit der richtigen Antworten – Histogramm.
- Schwierigkeitsgrad und Trennschärfe der Items^{42a}.
- Korrelationsberechnung zwischen den Resultaten der einzelnen Items und des Gesamttests. Wenn diese Korrelation zu niedrig ist, wird das Item gestrichen. Wenn das ausgeschaltete Item jedoch die Gesamtheit oder einen wichtigen Aspekt eines dem Test zugrundeliegenden Problems darstellt, muß man ein Ersatzitem finden, um nicht Gefahr zu laufen, daß der Test dem ursprünglich verfolgten Ziel nicht mehr Rechnung trägt.

d) Erste Zuverlässigkeitskontrolle, z. B. mit Hilfe der Testhalbierungsmethode.

e) Erste Globalkontrolle der Gültigkeit (Validität): Scheint der Test dasjenige, wozu man ihn konstruiert hat, hinreichend gut zu messen? Vergleich der Resultate mit der Beurteilung der Lehrer, mit den Resultaten anderer Tests ...

⁴² In diesem Fall bildet man eine Reihe von Untergruppen, wobei die Items in der Reihenfolge vom leichtesten zum schwierigsten klassifiziert werden. Man erhält somit einen allgemeinen Zyklus leicht, schwierig, leicht, schwierig, usw. Diese Anordnung ermutigt den Prüfling, zu versuchen, alle Items zu lösen, während er bei der Reihenfolge einfach, leicht, schwierig, dazu neigt, sofort aufzuhören, wenn er nach seiner Meinung auf eine echte Schwierigkeit stößt.

^{42a} Für eine gründliche Behandlung dieses Problems siehe: F. B. Davis, *Item Analysis Data*, Cambridge (Mass.) Harvard Univ., 1949.
Anm. d. dtsh. Bearbeiters: Vgl. auch G.A. Lienert: *Testaufbau und Testanalyse*, Weinheim, 2. Aufl. 1967.

4. Zweite experimentelle Form

Die für die erste experimentelle Form durchgeführte Analyse wird noch genauer präzisiert. Im allgemeinen erhält der Test jetzt ungefähr seine definitive Form. In besonderen Fällen zieht man erneut die Experten in diesem Fach und die Psychotechniker zu Rate. Man geht schließlich zum dritten Versuch über (der nicht unbedingt der letzte sein muß). Man läßt den Test drucken, jedoch sollte er noch so lange als Versuch betrachtet werden, als die eingehenden Gültigkeitsprüfungen noch nicht zufriedenstellend verlaufen sind.

B. Standardisierung

Der Hauptfaktor bei der Standardisierung besteht darin, daß alle Schüler einer gleichen Schulbildungsstufe derselben Prüfung unterzogen werden.

Darüber hinaus werden die Bedingungen der Durchführung und der Auswertung in eine einheitliche Form gebracht. Im Idealfall müßten die Tests, um vergleichbare Resultate zu erzielen, stets unter identischen Bedingungen durchgeführt werden.

- a) Gleicher Tag, gleiche Stunde, gleicher Grad der Ermüdung, gleiche Temperatur, gleiche materielle Voraussetzungen (Sitzgelegenheit usw.), Ruhe usw.
- b) Gleicher Grad der Vertrautheit der Probanden mit den Tests ganz allgemein und der verwendeten Technik im besonderen: Ein Schüler, dem seit langem Mehrfachwahlantworten oder die Verwendung von Antwortblättern des IBM-Typs geläufig sind, wird z. B. gegenüber dem Kind, das diese Erfahrungen nicht besitzt, im Vorteil sein.
- c) Gleiche Motivation⁴³;
- d) Gleiche Ausgangsinstruktionen und gleiches Training;
- e) Gleiche Dauer, wenn diese begrenzt ist.

Eine solche Übereinstimmung ist selbstverständlich kaum zu verwirklichen; es ist jedoch wichtig, daß man sie so weit als möglich zu erreichen sucht.

⁴³ Es ist bekannt, daß einige Prüflinge durch die während der Testdurchführung herrschende Prüfungsatmosphäre erschreckt und gehemmt werden; andere hingegen nehmen die Prüfung auf die leichte Schulter. Flanagan hat einen Index zur Messung der Motivation entwickelt. Siehe: J. Flanagan, *The Development of an Index of Examinee Motivation* (Educ. Psychol. Measurement, 1955, Nr. 15, S. 144–151).

C. Eichung

Die Normen sind statistische Verteilungen, die erstellt werden, um die Testergebnisse der Probanden, die sich nach Alter, Bildungsgrad usw. so weit als möglich entsprechen, miteinander vergleichbar zu machen.

Entsprechend dem Umfang des Untersuchungsprogramms haben die Normen für mehr oder minder große Gruppen (eine Klasse, Schule, Stadt, Nation) bzw. mehr oder minder gut definierte Gruppen (ländliches oder städtisches Milieu, sozio-ökonomischer Status...: Differentielle Normen) Gültigkeit.

Die nach Prozenten bzw. nach Dezilen vorgenommene Eichung ist lange Zeit das weitverbreiteste Verfahren gewesen. Die Prozentrangskala erlaubt es, den Rangplatz zu bestimmen, den ein Schüler unter hundert anderen, klassifiziert in der Reihenfolge ihrer Ergebnisse einnimmt.

Die Normalisierung⁴⁴, die im Gegensatz zur Prozentrangskala den Häufigkeiten der Normalverteilung Rechnung trägt, bietet ein weitaus rationelleres Klassifizierungssystem.

Man unterscheidet zwei Hauptgruppen:

- a) Die Eichung in ganzen Standardabweichungseinheiten, die direkt nach den Häufigkeiten der Normalverteilung berechnet werden; Beispiel: Z-Werte (*Z-scores*);
- b) Die Eichung in gebrochenen Standardabweichungseinheiten, die es erlauben, eine ungerade Zahl von Intervallen aufzustellen, was zwei wesentliche Vorteile bietet:
 - Man kann nur über ein zentrales Intervall verfügen, in dessen Mittelpunkt sich der Mittelwert befindet;
 - Die Notenwerte werden in verfeinerter Form klassifiziert.

Es gibt mehrere Skalentypen dieser Art (zu 5, 7 und 9 Intervallen). Die Skala zu 9 Intervallen einer halben Standardabweichung (*stanines*) scheint in Zukunft die Rolle spielen zu müssen, die die Prozentrangskala in der Vergangenheit gespielt hat.

Die Aufstellung von Normen durch Vergleich der Testresultate der Probanden ist nicht die einzige Eichungsmethode. Man kann zwei andere Kriterien übernehmen:

1. Interpretation der Resultate im Hinblick auf ihren Voraussagewert, z.B. im Hinblick auf die Erfolgswahrscheinlichkeit, die sie während der Studien erkennen lassen;

⁴⁴ Siehe Kapitel Statistik.

2. Interpretation der Resultate im Hinblick auf den Inhalt. Auf diese Weise können die Resultate eines Wortschatztests bewertet werden, indem man als Bezugspunkt die Wörter nimmt, die beim Sprechen am häufigsten verwendet werden.

Die Interpretation, die sich auf die Normen gründet, wie sie in den Test-Anleitungen vorgeschlagen werden, muß äußerst behutsam erfolgen. Schrader⁴⁵ bemerkt hierzu, daß in dem Augenblick, in dem man den Rangplatz eines Prüflings für zwei verschiedene Tests vergleicht, dieser durch die Art der Gruppe, die zur Eichung gedient hat, bestimmt wird; und die Tatsache, daß dieselbe Gruppe als Vergleichspunkt für zwei Tests gedient hat, bietet keine hinreichende Garantie: „... Wenn die Normen auf einer Eichgruppe basieren, die sich zu gleichen Teilen aus Jungen und Mädchen zusammensetzt, dann würde ein Junge, der hier für seine sprachlichen und seine rechnerischen Fähigkeiten den gleichen Rangplatz — ausgedrückt in Percentilen — erhalten hat, einen höheren Rangplatz in seinen verbalen Fähigkeiten als in seinen numerischen einnehmen, wenn die Eichung an einer Gruppe, die sich ausschließlich aus Jungen zusammensetzt, vorgenommen worden wäre“.

Demnach ist es wichtig, den Ursprung der vorgeschlagenen Normen sorgfältig zu prüfen. Wenn der Autor keine präzisen Aussagen hierüber macht, ergibt sich die Notwendigkeit, *lokale Normen* aufzustellen, *eine Lösung, die übrigens fast immer am meisten befriedigt*.

D. Gültigkeit (Validität)

Einen Test validieren, heißt zu prüfen, ob er das wirklich mißt, wozu man ihn einsetzen will.

Je nach ihrer Zweckbestimmung müssen die Tests drei Funktionen erfüllen: sie sollen prognostisch, diagnostisch und deskriptiv sein; alle drei Aspekte verlangen unterschiedliche Validierungsmethoden.

1. Prognostischer Aspekt

Die sicherste Art, eine Voraussage zu verifizieren, ist offensichtlich die, zu beobachten, ob sie sich bewahrheitet. Man beobachtet deshalb während einer angemessenen Zeitdauer die Verhaltensweisen bzw. die Leistungen der geprüften Schüler, um die Bestätigung bzw. die Widerlegung der Prognose (Methode des *follow-up*) zu finden, und man errechnet gegebenenfalls einen Validitätskoeffizienten (Korrelation).

⁴⁵ W. Schrader, Norms (Enc. of Ed. R., op. cit. S. 925).

2. Diagnostischer Aspekt

Wenn ein Test dazu dienen soll, diejenigen Unterrichtsprogrammunkte freizulegen, die, schlecht verstanden bzw. übergangen, die normale Entwicklung des Schülers behindern, erlaubt die Tatsache, daß die aufgrund der Diagnose „verabreichten Heilmittel“ die „Heilung“ bewirken, gewisse Rückschlüsse hinsichtlich des Wertes der Diagnose. Aber man muß auch den Wert der „Heilmittel“ in Betracht ziehen.

3. Deskriptiver Aspekt

Man versteht unter deskriptiven Tests solche, die eingesetzt werden, um einen Zustand zu beschreiben, eine Situation (Kenntnisniveau, Analyse der Charaktereigenschaften). Diese Klassifikation ist rein didaktisch, denn es kommt selten vor, daß die Deskription nicht auch für prognostische bzw. diagnostische Zwecke verwendet wird.

Man unterscheidet zwei Hauptvaliditätstypen dieser Prüfungen:

a) *Inhaltsvalidität (Content validity)*. Ein Leistungstest zur Überprüfung der im Rahmen eines festgesetzten Volksschulprogramms erworbenen Kenntnisse am Ende der Schulzeit muß die wichtigsten Aspekte dieses Programms effektiv erfassen. Wir wiederholen noch einmal, daß die Bestimmung der Bedeutung oft ein Werturteil enthält und daher einen axiomatischen Akzent darstellt. Auch kann man kaum nur von relativer Gültigkeit sprechen: Sind die vom Verfasser fixierten Ziele erreicht? Z. B. so wie man die Geometrie als Instrument für Denksportübungen ansieht oder als ein technisches Hilfsmittel, das der Lösung von praktischen Problemen dienen soll, wird man verschiedene Geometrietests konstruieren. Auch sollte der Benutzer sowohl ein klares Bild von seinen eigenen Vorstellungen haben, als auch von denen, die der Ausarbeitung des Instruments, das er einsetzen möchte, zugrundeliegen.

Darüber hinaus sollte die Analyse es erlauben, festzustellen, bis zu welchem Punkt der Test einen reinen Faktor mißt (Beispiel: bei einem Mathematiktest sind die Ergebnisse z. T. davon abhängig, inwieweit die Texte verständlich in der Muttersprache ausgedrückt werden, usw.).

b) *Konstruktvalidität (Construct validity)*. Der Pädagoge wie auch der Psychologe erklärt bzw. beschreibt Verhaltensweisen mit Hilfe von theoretischen oder hypothetischen Begriffen bzw. Modellen (*constructs*): Intelligenz, schöpferische Fähigkeiten, Aufrichtigkeit... Diese Begriffe sind nur aufgrund von Verhaltensäußerungen bekannt. Zur Beurteilung eines Tests, der sich auf derartige Konstrukte erstreckt, überprüft man, inwieweit die

Prüfung die ihr zugeordneten Verhaltensweisen erfaßt. Daher hat das IPAR (Institute of Personality Assessment and Research, Universität Kalifornien) für die Konstruktion eines Test zur Erfassung schöpferischer Begabung damit begonnen, schöpferisch besonders qualifizierte Personen (Architekten, Erfinder, Künstler, usw.) zu beschreiben und ihre Verhaltensweisen mit denen anderer weniger schöpferisch tätiger Personen zu vergleichen. Die dabei beobachteten Differenzen werden hypothetisch als kennzeichnend für schöpferische Fähigkeiten betrachtet. Zur Validierung des Tests prüft man, ob er den somit definierten Eigenschaften Rechnung trägt. Darüber hinaus erbringt ein *follow-up* eventuell die Bestätigung des Prüfungswertes.

E. Zuverlässigkeit (Reliability)

Um sicherzugehen, daß das Messen einer Straßenlänge mit Hilfe des Meßbandes eines Feldmessers zuverlässig ist, wiederholt man den Meßvorgang mehrere Male, um die durch die unterschiedliche Spannung des Meßbandes, die Dehnung oder die Kontraktion des Metalles, die falsche Verpfählung usw. bedingten Abweichungen einzukalkulieren. Auf diese Weise kann man sogar einen Zuverlässigkeitskoeffizienten errechnen, den man bei späteren Messungen verwendet.

Um die Zuverlässigkeit eines Tests zu ermessen, sollte man ihn auch bei der gleichen Probandengruppe mehrere Male hintereinander und unter gleichen Bedingungen durchführen können. Es wird deutlich, daß hier die Abweichungen wesentlich größer und komplexer sind als im physikalischen Bereich.

Tatsächlich ist jeder Augenblick im Leben eines Menschen einmalig (Veränderung des physiologischen und psychischen Gleichgewichts, erzielte Lernerfolge im Laufe vorangegangener Erfahrungspunkte, usw.).

Es wird daher wahrscheinlich niemals möglich sein, die Testzuverlässigkeit absolut zu bestimmen. Man beruft sich zur Zeit auf verschiedene Beurteilungsmittel, die von Anastasi zusammengestellt wurden⁴⁶.

In der Mehrzahl der Test-Anleitungshäfte geben die Verfasser die von ihnen errechneten Koeffizienten an. Man darf jedoch nicht übersehen, daß diese Koeffizienten, strenggenommen, nur für die Stichproben von Wert sind, mit denen sie errechnet wurden. Wenn man den gleichen Test bei ganz anderen Populationsschichten anwendet, wird es demnach nützlich sein, die Zuver-

46 A. Anastasi, *Psychological Testing*, New York, The Macmillan Company, 1961, S. 123.

Verfahren	Art des Zuverlässigkeitskoeffizienten	Erfasste Variationen
Wiederholungstest mit Hilfe der gleichen Testform zu verschiedenen Gelegenheiten.	Stabilitätskoeffizient	Temporalfluktuationen
Wiederholungstest mit Hilfe von Parallelformen zu verschiedenen Gelegenheiten.	Stabilitäts- und Äquivalenzkoeffizient	Temporalfluktuationen und Spezifität der Items
Wiederholungstest mit Hilfe von Parallelformen zu der gleichen Gelegenheit.	Äquivalenzkoeffizient	Spezifität der Items
Trennung eines Tests in zwei als äquivalent angesehene Teile (split-half-method): Vergleich der Gesamtergebnisse für gerade und ungerade Items, oder andere Unterteilung.	Interner Konsistenzkoeffizient (oder Index)	
Kuder-Richardson-Formel (oder andere)	Interner Konsistenzkoeffizient (oder Index)	Spezifität und Heterogenität der Items.

lässigkeitskoeffizienten gleichzeitig mit der Aufstellung spezifischer Normen noch einmal zu errechnen.

III. Testklassifizierung

Zwar vereinfacht es die Klassifizierung der Tests, wenn man sich auf die Verfahrensweisen bezieht, dagegen stößt man auf große Schwierigkeiten, wenn man auf den Zielen aufbaut. Ist ein Intelligenztest nicht fast immer gleichzeitig ein Leistungs- und Persönlichkeitstest? Es genügt, die verschiedenen Nutzenanwendungen zu betrachten, die man z. B. aus dem Labyrinth-Test von Porteus zieht, um sich von der Unzweckmäßigkeit starrer Abgren-

zungen zu überzeugen. Man wird auch feststellen, daß die Leistungstests eine sehr flexible Anwendung erlauben. R. C. Hall schreibt: „Ein in der 6. Volksschulklasse durchgeführter Rechentest kann als Begabungstest eingesetzt werden oder um die Leistungen im Rechnen oder in den Naturwissenschaften für die folgenden Jahre vorausszusagen, oder auch als Rechenleistungstest im 6. Schuljahr⁴⁷“. T. L. Kelley geht sogar so weit zu behaupten, daß es unmöglich ist, eine echte Unterscheidung zwischen den Begabungs- und den Leistungstests zu treffen. „Man würde sich“, so schreibt er, „ganz einfach der Meinung hingeben, daß zwei verschiedene Bezeichnungen für ein- und dieselbe Sache notwendigerweise eine echte Unterscheidung darstellen⁴⁸“.

Die von uns übernommene Klassifizierung besitzt daher nur einen didaktischen Wert.

A. Klassifizierung aufgrund der Verfahrensweise

1. Leistungstest; Papier-Bleistift-Test; Verbaltest

Der Leistungstest erfordert das Manipulieren von Objekten, die Konstruktion nach einem Modell, das Zusammentragen von Teilen usw. Der Papier-Bleistift-Test verlangt von dem Probanden eine schriftliche Antwort. Bei dem mündlichen Test wird die Antwort lediglich ausgesprochen.

2. Objektiver Test — Subjektiver Test

a) hinsichtlich des Prüfers

Der objektive Test ist derjenige, bei dem die Persönlichkeit des Auswerters prinzipiell ausgeschaltet und bei dem nicht nach persönlichen Kriterien bewertet wird. Man begnügt sich im allgemeinen bei dieser Anforderung mit Modellantworten und Auswertungsschlüsseln. Die einfachsten Mittel, auf die man sich beruft, damit die Auswertung einheitlich ist — gleichgültig, wer auswertet —, sind die Fragen, auf die eine Antwort aus einer Auswahl von mehreren möglichen erfolgt.

Der subjektive Test appelliert weitgehend an die Wertbestimmung und das persönliche Urteil des Auswerters. Wir werden feststellen, daß dies noch häufig der Fall ist bei Persönlichkeitstests, wo das klinische Verständnis des Prüfers eine wichtige Rolle spielt.

47 R. C. Hall, *Understanding Testing*, Washington, Office of Education, 1960, S. 10.

48 T. L. Kelley, *Interpretation of Educational Measurement*, Yonkers, World Book, 1927.

b) hinsichtlich des Probanden

Der objektive Test appelliert an die effektive Leistung. Häufig weiß der Proband überhaupt nicht, auf welches präzise Ziel hin der Prüfer diese Leistung fordert. Man versucht somit, die absichtlichen Verfälschungen in einem bestimmten Sinn auszuschalten. Beispiel: Geschicklichkeitsspiel, das nicht dazu dient, die Sache selbst, als vielmehr die durch es hervorgerufenen charakterlichen Reaktionen zu studieren.

Der subjektive Test fordert den Probanden auf, das Ergebnis seiner Selbstbeobachtung auszudrücken.

3. Standardisierter Test — nicht-standardisierter Test

Der standardisierte Test wurde zu Beginn dieses Kapitels definiert.

Nicht-standardisierter Test. Vornehmlich die Amerikaner bezeichnen hiermit die Kontrollübungen, die Prüfungsfragen, die von den Lehrern ausgearbeitet werden, um die Kenntnisse zu überprüfen. Diese Aufgabenstellung haben auch die Fragebogen, die zum Zwecke von begrenzten Umfragen erstellt sind, ohne strenge statistische Kontrolle.

4. Individualtest — Gruppentest

Individualtest. Der Prüfer führt den Test nur bei einem einzigen Probanden durch, dessen Antworten sowie häufig auch die Verhaltensweisen während der Befragung er notiert.

Gruppentest. Man greift auf Gruppentests zurück, sowohl aus Gründen der Zeitersparnis, als auch, um zu einem gleichen Zeitpunkt und unter den gleichen äußeren Voraussetzungen einen Meßwert für verschiedene Prüflinge zu erhalten.

5. Schnelligkeitstest — Niveautest

Schnelligkeitstest (Speed Test) (Zeitlich gemessener oder zeitlich begrenzter Test). Die Dauer dieser Prüfung ist streng begrenzt; die Probanden müssen während einer vorgeschriebenen Zeit so viele Fragen wie möglich beantworten (oder so viele Operationen wie möglich ausführen).

Niveautest (Power Test) (Zeitlich unbegrenzter Test). Der Prüfer beschäftigt sich hier weitaus mehr mit der Fundierung der Kenntnisse, mit dem Verständnis oder der reflektierten Ausdrucksform als mit der Schnelligkeit, mit der der Proband antwortet, mit seiner Widerstandskraft in Belastungssituationen oder seinen Spontanreaktionen.

B. Klassifizierung aufgrund des Gegenstandes

I. Die Intelligenztests

Die Intelligenztests messen einmal die allgemeine Begabung, d. h. die Potentiale, ohne sie — zumindest theoretisch — in Relation zu einer eng definierten Tätigkeit zu betrachten, zum anderen spezifische Begabungen, Faktoren oder kombinierte Faktoren, in Relation zu dem Lernerfolg auf bestimmten Gebieten. In der Umgangssprache unterscheidet man ähnlich: Man bestätigt, daß ein Schüler „intelligent“ ist — und versteht darunter eine polyvalente Kapazität der Person — oder daß er auf einem bestimmten Gebiet z. B. in Mathematik „begabt“ ist.

Da die uns interessierenden Tests bei der zweiten Gruppe rein pädagogisch orientiert sind und sich fast ausschließlich auf Schulleistungen beziehen, klassifizieren wir diese Prüfungen in die Leistungstests (diagnostische Tests). Wir werden übrigens feststellen, daß selbst die Tests zur Messung der allgemeinen Begabung, vor allem der verbalen, schwer zu trennen sind von den Schulleistungen. Auch hat man allzu häufig die Intelligenz mit der Fähigkeit verwechselt, Aufgaben innerhalb eines bestimmten Bildungssystems mit gutem Ergebnis auszuführen. Diese offenbar zu eng gefaßte Konzeption erklärt insbesondere, warum die Intelligenztests bis heute den divergierenden Fähigkeiten so wenig Platz eingeräumt haben^{48a}.

1. Die Tests zur Messung des allgemeinen Intelligenzniveaus

Das allgemeine Intelligenzniveau wird global gemessen, mit Hilfe von zahlreichen Prüfungen, die dank ihrer Vielseitigkeit in empirischer Form eine repräsentative Intelligenzstichprobe liefern. Diese Lösung wurde von den Begründern des Testverfahrens, angefangen bei Binet und Simon, übernommen.

Man unterscheidet:

a) *Die Entwicklungstests.* Man wendet sie vorwiegend bei Kleinkindern an, angefangen bei der Geburt, d. h. während jenes Lebensabschnittes, wo ein Altersunterschied von einigen Monaten bzw. einigen Wochen eine mittlere Leistungsdifferenz zur Folge hat, die höher ist als diejenige, die normalerweise zwischen gleichaltrigen Probanden besteht. Z. B. werden die Differenzen zwischen einem 3monatigen und einem 6monatigen Kind deutlicher

^{48a} Siehe: C. de Landsheere, Pour une pédagogie de la divergence (Synthèses, 1963, Nr. 204, S. 1–12).

hervortreten als jene zwischen zwei 8jährigen Kindern, von denen das eine begabt und das andere unbegabt ist.

Beispiel: Skala der psychomotorischen Entwicklung in der frühesten Kindheit, von O. Brunet und I. Lézine⁴⁹.

Anm. d. dtsh. Bearbeiters, Vgl. Hetzer-Test.

Hervorgegangen aus der Vergleichsstudie verschiedener Testbatterien, vornehmlich aus der von Bühler-Hetzer und A. Gesell, ist diese Skala das Ergebnis langjähriger Untersuchungen; sie ist auf eine repräsentative Stichprobe der französischen Bevölkerung geeicht.

Sie erstreckt sich auf 19 Niveaus und enthält je 10 Tests für die Altersgruppen: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 18, 21, 24 und 30 Monate und je 6 Tests für 3, 4 und 5 Jahre.

Das Kind wird unter dem Aspekt seiner Verhaltensentwicklung, seiner visuell-manuellen Koordinationsfähigkeit, seiner Sprachentwicklung, seines Verhaltens in der Gemeinschaft und beim Spiel untersucht.

Das Material umfaßt: Puzzle-Spiele, Ball, Klingel, Würfel, Flasche, Spiegel, Kinderklapper, Bilderbücher, Einlegespiele, Bleistift, Ring usw.

Neben den eigentlichen Testprüfungen sieht die Untersuchung auch Fragen vor, die an die Mutter zu richten sind (Beispiel: 18 Monate: 6 Prüfungen, 4 Fragen).

Das Notieren der Ergebnisse erfolgt auf Spezialblättern, die es erlauben, Profile zu erstellen. Man errechnet einen Entwicklungsquotienten, beginnend beim 4. Monat.

b) *Allgemeine Begabungstests.* Man unterscheidet Tests für Kinder und für Erwachsene. Jede Gruppe umfaßt Einzel- und Kollektivuntersuchungen, verbale, nicht-verbale bzw. gemischte Untersuchungen. Wir geben lediglich drei typische Beispiele:

*Wechsler-Intelligenz-Test für Kinder (HAWIK)*⁵⁰.

Dieser Test, der sich aus Verbal- und Handlungsteil zusammensetzt wird allgemein als einer der besten Intelligenztests für Kinder bezeichnet.

Im Gegensatz zu dem Binet-Simon-Test entspricht der Prüfung von Wechsler keine Altersskala, sondern eine Punktskala (gleiche Items für alle Probanden; sie werden in der Reihenfolge zunehmender Schwierigkeitsgrade dargeboten).

Die Punktwerte erlauben die Berechnung von drei Abweichungs-Intelligenzquotienten: Verbal, nicht-verbal und komplette Skala. Man prüft auch die Streuung der partiellen Punktwerte.

Das Zentrum für Angewandte Psychologie (Centre de Psychologie Appliquée) in Paris schlägt eine Eichung auf einer französischen Population vor; eine belgische

49 Editions Scientifiques et psychotechniques, Paris. Siehe auch: O. Brunet, u. I. Lézine, *Le développement psychologique de la première enfance*, Paris, P.U.F., 1951.

50 F. B. Hardesty, H. J. Priester, *Handbuch für den Hamburg-Wechsler-Intelligenztest für Kinder*, Bern-Stuttgart, 1956.

Eichung wurde von R. Berte⁵¹ vorgenommen. Deutsche Bearbeitung: Hamburg-Wechsler-Intelligenz-Test für Kinder, HAWIK, von P. Hardesty und H. Priester.

Testaufbau ^{51a)}

A. Verbalteil:

- a) Allgemeines Wissen: 30 Items. Beispiele: Wieviel Stück hat ein Dutzend? Warum schwimmt Öl auf Wasser? Was ist ein Barometer?
- b) Allgemeines Verständnis: 16 Items. Beispiele: Was sollst Du tun, wenn du dich in den Finger geschnitten hast? Welche Aufgaben hat die Polizei?
- c) Rechnerisches Denken: 16 leichte Rechenaufgaben. Beispiel: Wie viele Stücke habe ich, wenn ich einen Apfel einmal durchschneide?
- d) Gemeinsamkeitenfinden: 16 Items. Beispiele: Was ist das Gemeinsame bei einer Pflaume und einem Pfirsich? Was ist das Gemeinsame bei einer Katze und einer Maus?
- e) Wortschatztest: 40 Items. Beispiele: Was ist ein Hund? Was ist Hopfen?
- f) Zahlennachsprechen: (Zusätzlicher Test nach Belieben). Es werden im Abstand von einer Sekunde folgende Zahlen genannt, die zu wiederholen sind: 5-3-8-7-1-2-4-6-9.

B. Handlungsteil:

- a) Bilderergänzen: 20 Items. Beispiele: Bei einem Hahn fehlt der Sporn. Bei einem Tisch fehlt das vierte Bein.
- b) Bilderordnen: 7 Items. Beispiele: Karten, die Teile eines Hundes, einer Bildergeschichte darstellen, sollen in der richtigen Reihenfolge hingelegt werden.
- c) Mosaik-Test: Der Versuchsleiter legt mit auf allen sechs Seiten verschieden bemalten Klötzen ein Muster. Die Versuchsperson wird aufgefordert, das Muster nachzulegen.
- d) Figurenlegen: Der Versuchsperson werden Einzelteile einiger Figuren (Junge, Pferd, Gesicht, Auto) nacheinander vorgelegt mit der Aufforderung, diese Teile so zusammensetzen, daß ein richtiges Bild daraus entsteht.

Zusätzlich zu dem Test-Handbuch und dem Test-Protokoll-Bogen wird zur Durchführung des HAWIK ein Testkasten benutzt, der folgende Teile enthält:

1. Ein Heft mit den Aufgaben des Rechnerischen Denkens (13-16), den Aufgaben des Bilderergänzens und den Mustern des Mosaik-Tests.
2. Ein Kasten mit Karten für das Bilderordnen.
3. Neun bunte Klötze für den Mosaik-Test, die gleichzeitig für die ersten drei Aufgaben des Rechnerischen Denkens benutzt werden.
4. Vier Schachteln mit den Teilen für das Figurenlegen.
5. Eine Karte, auf der die Anordnungen angegeben sind, nach denen die einzelnen Teilstücke beim Figurenlegen vom Versuchsleiter auszulegen sind.

51 R. Berte, *Essai d'adaptation du WISC à des écoliers belges d'expression française*, Louvain, C. N. R. P. S., 1961.

51a Anm. d. dtsh. Bearbeiters: Statt der im Original aufgeführten französischen Bearbeitung wurden hier Beispiele aus der deutschen Bearbeitung des HAWIK aufgeführt, die von C. Bondy herausgegeben wurde bei Huber: Bern-Stuttgart, 2. Aufl. 1963.

6. Ein Bewertungsschlüssel für den Zahlen-Symbol-Test.

Bei diesem Test, den De Landsheere nicht anführt, geht es darum, bestimmten Figuren (Stern, Ball, Dreieck, Kreuz) eine bestimmte Zahl oder ein bestimmtes Zeichen zuzuordnen.

Test zur Messung des Verständnisses von Texten, von P. Rennes⁵²

Dieser verbale Gruppentest umfaßt 25 leichte Texte: Sprichwörter, kurze Erzählungen oder kleine Exposés. In Frankreich geübt, wendet er sich an Jugendliche und Erwachsene.

Für jedes Item muß man unter vier Auslegungen die nach seiner Meinung zutreffende Antwort wählen. Beispiel: „Der Skandal ist fast immer Einbildung und entspricht fast nie den Tatsachen“.

- A. Der Skandal ist abhängig von der Art, wie wir die Tatbestände beurteilen.
- B. Nicht jeder empfindet einen bestimmten Sachverhalt als skandalös.
- C. Es sind die Tatbestände, die einen Skandal bewirken.
- D. Es kann irgendein beliebiger Tatbestand Anlaß geben, einen Skandal auszulösen.

Performance Scale von R. Pintner⁵³

Nicht-verbale Kollektivprüfung, entworfen für etwa 9- bis 15jährige. Alle Fragen werden in Form von Bildern vorgelegt, und die Antworten sind aus mehreren möglichen auszuwählen. Diese Fragen sind in der Tat kleine Denkaufgaben: Es sind Figuren zu teilen, unvollständige Zeichnungen zu ergänzen, Muster übereinanderzulegen, Serien zu bilden, zusammengehörige Figuren paarweise zusammenzustellen, Papier zu falten.

Die Normen beruhen auf mehr als 6000 Prüfungen. Man errechnet einen Abweichungs-Intelligenzquotienten.

2. Das differentielle Testen von Begabungen

Es befaßt sich mit dem Studium der Intelligenzkomponenten, die entweder empirisch oder durch die Faktorenanalyse bestimmt werden.

a) Die Prüfungen, die sich auf einen Einzel-Faktor erstrecken:

1. Der Generalfaktor (g), der als eine Art allgemein üblicher Intelligenzgradmesser angesehen wird. Beispiel: Domino-Test (D 48), von Anstey⁵⁴.

52 Institut für Angew. Psychologie (Centre de Psych. Appliquée, Paris – Editest, Brüssel). Anm. d. dtsh. Bearbeiters: Vgl. auch: Test VL, Dtsch. Schultests, Weinheim 1967.

53 World Book Company. Anm. d. dtsh. Bearbeiters: Vgl. auch: FR-Test (Figure Reasoning von Daniels) oder Bildertest, Dtsch. Schultests, Weinheim 1967.

54 Franz. Bearbeitung von R. Pichot: Institut für Angewandte Psychologie (Centre de Psychologie Appliquée) Editest.

Anwendbar vom 12. Lebensjahr an, besteht dieser Test aus Reihen von Dominosteinen. Man muß die Gesetzmäßigkeit klären, die jeder Reihe zugrunde liegt, so daß man diese zu vervollständigen imstande ist. Die Prüfung leistet in besonders hohem Maße der Messung von Faktor (g) Genüge: $r = .90$.

Die Eichung wurde bei französischen Probanden aufgrund des Bildungsniveaus und nach Altersklassen vorgenommen.

Als besonders in Europa gut erprobt läßt sich der D 48 Test auch anwenden, um das Nachlassen der Konzentrationsfähigkeit zu messen^{54a}.

2. Die Gruppenfaktoren, die sich nach der geistigen Leistungsfähigkeit zusammenstellen⁵⁵.

b) Die Prüfungen, die sich auf mehrere Faktoren erstrecken:

Beispiel: Faktorielle Testbatterie der geistigen Grundfunktionen, Primary Mental Abilities, P. M. A. von L. und T. Thurstone, Revision 1949⁵⁶.

Diese Tests messen acht von Thurstone isoliert herausgestellte geistige Grundfunktionen: Sprachliches Verständnis (V), Raumvorstellung (S), induktives Denken (R), Umgang mit Zahlen (N), Wortflüssigkeit (W), Gedächtnis (M), Schnelligkeit der Wahrnehmung (P) und Motorik (Mo).

Es gibt drei Testbatterien, deren faktorielle Zusammensetzung unterschiedlich ist.

5–7 Jahre:	Mo	P	N	+	R	V	S
7–11 „	–	P	N	R	V	S	
11–17 „	–	–	N	R	V	S	W

Testbatterie 5–7 Jahre

Sprachliches Verständnis: In vier Bildreihen einen angegebenen Gegenstand kennzeichnen. Beispiel: Gemälde – Maler – Farbtopf – spielendes Kind. Anweisung: „Kennzeichne den Künstler“.

Schnelligkeit der Wahrnehmung: Aus einer Reihe von geringfügig abweichenden Mustern dasjenige herausuchen, das einem vorgegebenen Muster gleicht.

Quantitative Tests. Beispiel: In einer Reihe drei Flugzeuge ankreuzen – In einer Reihe den ersten und den letzten Fisch ankreuzen.

Motorischer Test: Besteht aus vier Linien von Punktpaaren, die man durch eine vertikale Linie verbinden muß. Tests zur Prüfung der Raumvorstellung: Zeichnungen nach Vorlagen vervollständigen usw.

54a Anm. d. dtsh. Bearbeiters: Vgl. hierzu auch: Untertests vom IST – Amthauer, Göttingen, 2. Aufl. 1955, und Kettwiger Schulreifetest, Dtsch. Schultests, Weinheim 1967.

55 Die modernen Forschungen weisen darauf hin, daß die spezifischen Faktoren (s) keine einfachen Faktoren sind. Darum erscheint es richtiger, sie „Gruppenfaktoren“ zu nennen. Wir können hier keine Analyse der faktoriellen Begriffe geben.

56 Science Research Associates – Centre de Psychologie Appliquée (Institut für Angewandte Psychologie).

Testbatterie 11–17 Jahre

Sprachliches Verständnis: Synonyme. Beispiel: Aus einer Anzahl von Wörtern dasjenige herausuchen, das die gleiche Bedeutung hat wie das Wort „verheimlichen“: schwindeln – verschweigen – sorgfältig abwägen – bemänteln – zusammenfügen. Raumvorstellung: Besteht darin, sich kleine Zeichen – z. B. ein F – vorzustellen, die man rotieren läßt.

Induktives Denken: Es sind Reihen zu vervollständigen. Beispiel: a b m c d m e f m . . .

Umgang mit Zahlen: Es sind Additionen zu überprüfen.

Wortflüssigkeit: Es ist die größtmögliche Anzahl von Wörtern anzugeben, die mit dem gleichen, beliebig zu bestimmenden Buchstaben beginnen.

Als typisches Meßinstrument einer bestimmten faktoriellen Konzeption hat der PMA seit seinem Aufkommen (1941) zu einem erheblichen Fortschritt geführt und leistet auch heute noch gute Dienste. Wir haben ihn wegen seiner klaren Struktur als Beispiel gewählt. Er beginnt jedoch zu veralten. Außerdem wirft man ihm den zu großen Einfluß der Geschwindigkeiten auf die Punktzahlen und Unzulänglichkeiten in der Validierung vor.

II. Schulleistungstests (Achievement Test)

Zwischen den Schulleistungstests und den Intelligenztests besteht eher ein Unterschied im Grad als im Charakter. Die ersteren messen im Prinzip die unter bekannten, nämlich den schulischen Bedingungen, erzielten Lernergebnisse, die uns interessieren, während die Intelligenztests angeborene Möglichkeiten und zahlreiche Erfahrungen des täglichen Lebens messen.

Die Leistungstests sollen einerseits eine Erfolgsprognose erlauben, andererseits ein klares Bild von der Situation bzw. dem Kenntnisstand vermitteln und schließlich eine Diagnose erstellen, d. h. eine Schwierigkeit lokalisieren und gegebenenfalls die Ursache aufzeigen.

1. Die prognostischen Tests

a) *Die Tests zur Messung des allgemeinen Reifegrades.* Diese Tests unterscheiden sich von den allgemeinen Begabungstests tatsächlich nur im Hinblick auf ihr ausdrücklich definiertes Ziel: Voraussage aufgrund der Fähigkeit des Schülers, sich den im Unterrichtsprogramm vorgesehenen Stoff anzueignen (Schulreifetests).

Beispiel: Test „6 Jahre“ von A. Van Wayenberghe⁵⁷.

Dieser Test verfolgt im wesentlichen folgende Ziele:

Er soll einen ersten Eindruck über die Schulreife vermitteln (Verständnis und Aus-

⁵⁷ Clerehaut, Brüssel. Anm. d. dtsh. Bearbeiters: Vgl. hierzu im deutschen: Kettwiger Schulreifetest, Dtsch. Schultests, Weinheim 1967.

führung von mündlichen Anweisungen, Zahlenbegriff, kleine Denkaufgaben usw.). Vom Eintritt in die erste Volksschulklasse an soll er Fälle aufzeigen, die einer eingehenden individuellen Prüfung unterzogen werden sollten (psychischer Entwicklungsstand, Affektivität, Antriebskraft, Linkshändigkeit, Orientierungs- und Wahrnehmungsfähigkeit usw.)

Testaufbau:

Reduzierte Form des „*Test d'intelligence B. D.*“ von Buyse-Decroly: Definition durch Angabe des Verwendungszweckes^{57a}; Aufträge die beidhändig auszuführen sind, numerische Übungen (die zeitliche Begrenzung wurde ausgeschaltet und die Zahl der Items reduziert);

Zehn Reihen von Zeichen zum Erkennen einer vorliegenden Raumlage-Labilität, die der Skala von Borel-Maisonny (Entdeckung künftiger Legastheniker) entnommen sind;

Komplexe Figur, die zu kopieren ist;

Test von Goodenough (Mann-Zeichen-Test) oder von Fay (eine Frau geht spazieren – es regnet).

Der Voraussagewert dieser Testbatterie liegt hoch. Die Untersuchungen über die Validität werden fortgeführt.

b) *Die spezifischen Reifetests*

Beispiel: Lee-Clerk Reading Readiness Test, rev. Aufl. 1962, von J. M. Lee und W. W. Clerk⁵⁸.

Test zum Messen der Reife für das Erlernen des Lesens, anwendbar gegen Ende des Kindergartenalters und zu Beginn der Volksschule. Er besteht aus vier Teilen, für die selbstverständlich die Anweisungen mündlich gegeben werden.

Aus zwei parallellaufenden Kolonnen gleiche Buchstaben herausuchen (12 Items);

Beispiel:

s	g
m	u
u	s
g	m

Reihen von vier Buchstaben, wobei ein einziger anderer Buchstabe vorkommt: Beispiel: S s S E: der nicht zu der Reihe zugehörige ist anzukreuzen.

Zwanzig Bilderreihen. Dem Kind wird die Aufgabe gestellt, in jeder Reihe ein bestimmtes Bild anzukreuzen (Überprüfung von Wortschatz, Textverständnis sowie der Auffassungsgabe von gegebenen Anweisungen). Durch Vergleichen ein bestimmtes Wort in einer Wortreihe erkennen.

(20 Items) Beispiel: Ball: Ball – Wall – Fall – Knall.

Diese sehr gern verwendete Methode bietet einen hohen Zuverlässigkeitsgrad ($r = .96$ Spearman-Brown). Der Voraussagewert ist gut (ungefähr $r = .60$).

^{57a} Anm. d. dtsh. Bearbeiters: Kleinere Kinder können noch keine abstrakten Definitionen von einem Gegenstand geben. Sie beschränken sich darauf, seinen Verwendungszweck anzugeben, z. B.: Eine Gabel ist etwas, womit ich essen kann.

⁵⁸ California Test Bureau.

Man bestimmt den Meßwert der Lesereife, den man im allgemeinen folgendermaßen interpretiert:

Meßwert der „readiness“	Klassifizierung	Prognose	Vorgeschlagene Wartezeit
1,5–1,9	hoch	ausgezeichnet	keine
0,7–1,4	mittelhoch	gut	keine
0,4–0,6	mäßig hoch	mäßig	1 bis 6 Monate
0 –0,3	niedrig	schwach	7 Monate u. mehr

2. Die Tests zur Messung des Leistungsstandes

Einige Autoren unterscheiden innerhalb dieser Rubrik die *Survey-Tests* (zur Messung des Lernerfolgsgrades, den ein Schüler in einem bestimmten Unterrichtsstoff zu irgendeinem Zeitpunkt im Laufe seiner Studien erzielt) und die *Kenntnis-Inventar-Tests* [zur Überprüfung der Mindestkenntnisse, die man besitzen muß, um in die nächste Stufe zu gelangen. In Deutschland z. B. Göppinger-Leistungstest oder den Allgem. Schulleistungstest (AST 4+) der Dtsch. Schultests, Weinheim 1967, Anmerkung des Bearbeiters], die der Erstellung einer objektiven Bilanz der erworbenen Kenntnisse dienen sollen, bevor man mit neuem Studienabschnitt beginnt. In Wirklichkeit unterscheiden sich die gängigen Tests nur nach dem Gebrauch, den man von ihnen macht.

ICF 6/5 Inventar des Kenntnisstandes im Französisch, Ende des 6. Schuljahres, Beginn des 5. Schuljahres⁵⁹.

Gruppenprüfung von 50 Minuten Dauer:

Grammatikalische Analyse

Logische Analyse

Flexion (Beugung)

Gegensätze, z. B.: Ein Wort finden, das das Gegenteil von „stark“ ausdrückt und sich mit „ach“ reimt (z. B. schwach);

Synonyma

Textverständnis (Wortschatz, allgemeines Verständnis).

Die kombinierten Notenwerte erlauben eine Auswertung des Kenntnisstandes in der Grammatik, des Umfangs und der Präzision des Wortschatzes sowie ein Urteil über den Stand des verbalen Verständnisses. Der Globalnotenwert wird ebenfalls geeicht.

59 INOP.

Rechentest, 1. und 2. Schuljahr, von L. Cleempoel und F. Hotyat⁶⁰.

Diese Prüfung dient der Messung des Leistungsstandes im Rechnen am Ende des ersten / zu Beginn des zweiten Volksschuljahres. Sie gibt einen Gesamtleistungsstand an und erlaubt auch die Durchführung bestimmter Leistungsanalysen.

Es gibt folgende Arten von Übungen:

Zahlenkenntnis: Mengengleichheit, Zweierreihe, Zählen vorwärts und rückwärts, Umgehen mit Geld;

Zahlen nach ihrer Größe ordnen;

Rechnen: Im Zahlenraum bis 10 und zwischen 10 und 20; Ergänzungsaufgaben, Zahlen zerlegen.

Einfache und zusammengesetzte Aufgaben.

Dezilnormen und Skalierung in Standardabweichungen zu fünf Intervallen basierend auf der Untersuchung von 1471 Schülerarbeiten.

3. Die analytischen Tests

Die analytischen Tests haben zum Ziel, die Schwächen und falschen Lerngewohnheiten in allen Bereichen des schulischen Lernens aufzudecken.

Die *Kontrolltests* sollen allgemeine Schwierigkeitszonen messen, (Beispiel: Mangelhafte Orthographie infolge schwachen visuellen Gedächtnisses), während die *diagnostischen Tests* der Lokalisierung eines speziellen Problems dienen (Beispiel: Unkenntnis einer bestimmten Beugungsregel).

Auch hier gibt es wenig französischsprachige Autoren, die zwischen den beiden Prüfungsarten, für die allgemein die Bezeichnung *diagnostische Tests* üblich ist, unterscheiden. Es wird hier übrigens deutlich, daß die Analyse der eigentlichen diagnostischen Prüfungen es erlaubt, die allgemeineren Fehlerquellen zu identifizieren.

Beispiel: Diagnostische Prüfungen der Orthographie, von S. Borel-Maisonny⁶¹.

Wie alle anderen von S. Borel-Maisonny vorgeschlagenen Tests zeigt auch dieser einen ausgesprochenen klinischen Charakter. Er leistet große Dienste bei der Prüfung von Kindern und Erwachsenen, die erhebliche Schwierigkeiten in der Orthographie haben.

Eine Reihe leichter Diktate zeigt die mögliche Fehlerquelle an:

Visuelles Versagen oder einfache Unkenntnis ohne besondere Bedeutung;

Schlechtes visuelles Gedächtnis;

Mangel an Konzentration beim Zuhören;

60 Institut Supérieur de Pédagogie du Hainaut, Morlanwelz. Anm. d. dtsh. Bearbeiters: Vgl. hierzu auch: Rechentest (RT4+), Dtsch. Schultests, Weinheim 1967.

61 S. Borel-Maisonny, Langage oral et langage écrit, I. Neuchâtel, Delachaux u. Niestlé, 1962, 2. Aufl. S. 94 ff. Anm. d. dtsh. Bearbeiters: Vgl. hierzu auch: Diagnostischer Rechtschreibtest (DRT 2, DRT 3), Dtsch. Schultests, Weinheim 1967.

Fehler, die auf eine falsche Aussprache deuten;
 Grundsätzliche Sprachschwierigkeiten;
 Übereinstimmungsfehler durch unzureichendes Sprechen;
 Unaufmerksamkeit.

Nach S. Borel-Maissonny haben jedoch die orthographischen Mängel häufig einen viel allgemeineren Grund: Unzulänglichkeiten im sprachlichen und gedanklichen Erfassen des Gegenstandes.

Rechenanalysetests (Ende der Volksschule, Beginn des höheren Schulunterrichts), von Bongrain, Burion, Durviaux, Hotyat und Manouvrier, 1961⁶².

Die Untersuchung erstreckt sich auf folgende im belgischen Lehrplan vorgesehene Unterrichtsstoffe:

Zählen von ganzen Zahlen und Dezimalzahlen;
 Schriftliches Rechnen (ganze und Dezimalzahlen);
 Kopfrechnen: Das Einmaleins, Multiplizieren mit 125, 0,25, 0,75;
 Dividieren mit 25, 0,5, 0,25;
 Bruchrechnen: Kürzen, Verwandeln, die 4 Grundrechenarten;
 Metrisches System: Längenmaße, Körpermaße usw.;
 Geometrische Formen: Nachgezeichnete geometrische Linien;
 Kenntnis von geometrischen Figuren, Berechnung des Umfangs, der Flächen, des Rauminhalts, Abwicklungen usw.;
 Aufgaben: Dreisatz, Prozentrechnung, umgekehrtes Verhältnis und Zinsrechnung, Durchschnittberechnung.

Für jeden Unterrichtsstoff gibt man die prozentualen Anteile der richtigen Übungen sowie den Prozentsatz der Schüler an, die 2/3 der Übungen erfolgreich durchgeführt haben.

Für jedes der drei Items jeder Serie trägt man die Ergebnisse aller Schüler einer Klasse in eine Tabelle ein. Man erkennt auf diese Weise sehr schnell, wo noch Lücken in der Beherrschung des Unterrichtspensums bestehen.

Darüber hinaus enthält das Handbuch die Analyse der häufigsten Fehler⁶³.

III. Die Persönlichkeitstests

Die Begabungen, auch wenn sie noch so hervorragend sind, genügen noch nicht, um den Schulerfolg sicherzustellen. Die persönliche Ausgeglichenheit und das Interesse sind dabei auch eine wichtige Voraussetzung.

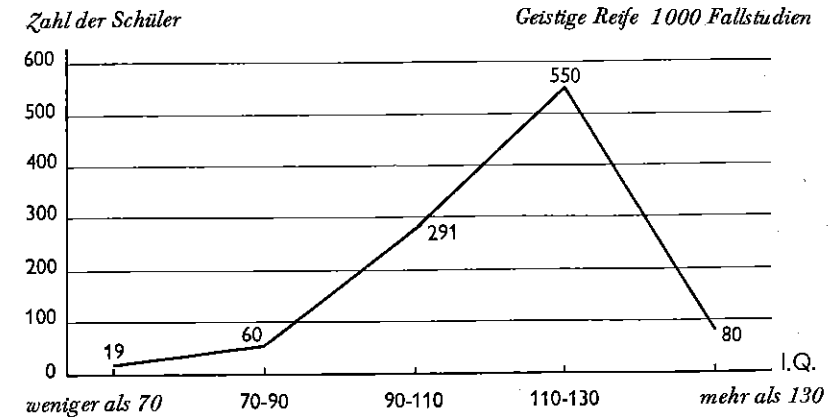
Man macht immer wieder die Beobachtung, daß viele Schüler mit schlechten Schulleistungen einen hohen Intelligenzquotienten besitzen. Namentlich G. Mauco gibt an, daß 80% der Schüler, die die Psychologisch-pädagogische

⁶² Institut Supérieur de Pédagogie du Hainaut, Morlanwelz, Belgien.

⁶³ Siehe auch die sehr bedeutende analytische Untersuchung der Commission Consultative Universitaire de Pédagogie, CCUP, L'arithmétique au niveau de la sixième primaire, Brüssel, Ministerium für Erziehungs- und Bildungswesen, s. d.

Zentren konsultieren, einen durchschnittlichen (IQ = 100) oder überdurchschnittlichen Intelligenzquotienten haben. Bei 1000 Fallstudien ergibt sich folgende Verteilung⁶⁴ (siehe unten).

Das Studium der Persönlichkeit ist eine Domäne der dafür besonders ausgebildeten Psychologen. Dennoch befindet sich der Erzieher infolge des täglichen Umgangs mit dem Kind in der bevorzugten Situation, ständig zu beobachten; er kann aber das Kind nicht erziehen, ohne seinem Charakter Rechnung zu tragen. Außerdem ist das Gespräch zwischen dem Psychologen und dem Pädagogen nur dann möglich, wenn der letztere eine gute psychologische Vorbildung hat.



1. Die subjektiven Tests

Sie stützen sich auf die Autoanalyse, auf die Fähigkeit der Selbstbeobachtung der Probanden, und sie sind somit allen Gefahren der Verfälschung ausgesetzt, die diese Methode mit sich bringt. Es sind vor allem Fragebogen.

Beispiel: SRA Youth Inventory, von H. Remmers und B. Shimberg, 1956⁶⁵.

Feststellung der Sorgen und Probleme von Jugendlichen zwischen 15 und 18 Jahren. Der Problemfragebogen enthält in der amerikanischen Fassung 296 Items, die sich auf 8 Bereiche verteilen (^{65a}).

⁶⁴ Siehe *L'inadaptation scolaire et sociale et ses remèdes*, Paris, Bourrelier, 1959, S. 185.

⁶⁵ Science Research Associates.

^{65a} Anm. d. dtsh. Bearbeiters: Hier wurde die deutsche Fassung des SRA von Süllwold u. a.: Problemfragebogen für Jugendliche. Hogrefe, Göttingen 1967 statt der amerikanischen Fassung von uns eingefügt.

Wir bringen hier die deutsche Fassung des SRA, der vom Deutschen Institut für Internationale Pädagogische Forschung von H. Belsler unter Mitarbeit von W. Alts und I. Winter ins Deutsche übertragen wurde. Die statistische Auswertung, Analyse und Interpretation der erhobenen Daten wurden 1959 von F. Stüllwold geplant und durchgeführt.

Im Jahre 1959 wurde von M. Berg eine Erhebung an 1803 männlichen und 1965 von B. M. Homberg eine solche an 1892 weiblichen Berufsschülern vorgenommen. Der deutsche Problemfragebogen enthält, da er auf spezielle kulturelle Unterschiede eingehen mußte, 306 Items. Er hält sich weitgehend an das amerikanische Original und ist wie dieses in folgende 8 Problembereiche unterteilt:

1. *Meine Schule*: In diesem Bereich werden neben allgemeinen Einstellungen gegenüber der Schule speziell schulische Probleme, Fragen über die Einschätzung der persönlichen Leistungsfähigkeit in einzelnen Fächern und bei Prüfungen, sowie Fragen zum Schüler-Lehrer-Verhältnis angesprochen.

Beispiel: Ich möchte gerne in der Schule etwas mehr praktische Ausbildung haben.

2. *Nach der Schulzeit*: Fragen der beruflichen Erziehung, Fortsetzung der schulischen Ausbildung und Berufswahl.

Beispiel: Für welchen Beruf eigne ich mich am besten?

3. *Über mich selbst*: Wie der Prüfling sich sieht, welches seine Anpassungsprobleme sind.

Beispiel: Ich fühle mich leicht verletzt. Ich habe Angst, Fehler zu machen.

4. *Ich und die anderen*: Verhältnis zu Klassenkameraden und Erwachsenen.

Beispiel: Ich muß mehr Selbstvertrauen entwickeln. Ich muß noch lernen ein guter Zuhörer zu sein.

5. *Zu Hause*: Fragen nach Eltern, Geschwistern, dem Erziehungsstil, den Wohnverhältnissen und dem familiären Leben.

Beispiel: Meine Eltern haben kein Vertrauen zu mir.

6. *Jungen und Mädchen*: Sexualprobleme, Beziehungen zum anderen Geschlecht.

Ich weiß nicht, wie man sich die Zuneigung eines Mädchens erhalten kann.

Beispiel: Ich frage mich warum ich so gehemmt bin, wenn ein Junge mich anspricht.

7. *Gesundheit*: Körperliches Wohlbefinden als Voraussetzung für das seelische und soziale Wohlbefinden.

Beispiel: Ich möchte gerne eine bessere Figur haben. Ich habe keinen Appetit.

8. *Allgemeines*: Anteilnahme an Fragen des gesellschaftlichen und kulturellen Lebens.

Beispiel: Ich grübele über den Sinn des Lebens nach.

Ich bin unsicher in Glaubensdingen.

Das Antwortsystem erlaubt es dem Probanden, anzugeben, ob ihm das Problem schwerwiegend erscheint, von durchschnittlicher Wichtigkeit ist, oder ob es ihn nur gelegentlich oder gar nicht berührt.

2. Die objektiven Tests

Hier wird die Autoanalyse durch eine Leistung ersetzt, aufgrund welcher die Psychologen gewisse Aspekte der Persönlichkeit beobachten. Die objektiven

Tests erlauben eine strenge Quantifizierung, der Prüfling kann nur schwerlich seine Antworten in einem ihm angemessen erscheinenden Sinn verfälschen.

Beispiel: Perseverations-Test, von R. Zazzo und M. Stambak⁶⁶.

Überprüfung der geistigen Unbeweglichkeit, d. h. der mehr oder minder starken Opposition gegen eine Modifizierung der angenommenen Lerngewohnheiten. Diese Testbatterie enthält fünf Tests:

1. Test mit abwechselnden Groß- und Kleinbuchstaben (abc, ABC, a A b B c E).

Jede Reihe wird 15 Sekunden lang geschrieben. Aus dem Nachlassen der Leistungsspannkraft ergibt sich der Meßwert für die Perseveration.

2. Farbttest: Blau als rot bezeichnen und umgekehrt.

3. Zahlentest: Zahlen sind normal und dann umgekehrt zu schreiben.

4. Test „Satz I“: Einen Satz mit Großbuchstaben, dann mit Kleinbuchstaben schreiben, dann abwechselnd Groß- und Kleinbuchstaben.

5. Test „Satz II“: Verdoppeln der Großbuchstaben.

R. Zazzo hat nachdrücklicher als sein Vorgänger die Unabhängigkeit des Perseverationsfaktors herausgestellt und eine sorgfältig ausgearbeitete Validität der Prüfung dargelegt. Normen vom 9. Lebensjahr an aufwärts.

3. Projektive Verfahren

Der Proband wird aufgefordert, ein undeutliches bzw. unvollständiges Schriftzeichen oder eine Zeichnung zu interpretieren bzw. zu vervollständigen (Tintenklecks, ungenaues Bild, unvollständiger Satz usw.) oder sich mit gewissen schöpferischen Arbeiten zu befassen (Zeichenentwurf, Spiel, Konstruktion). Hier liegt die Hypothese zugrunde, daß der Proband, der praktisch vom Nichts ausgeht oder nur ganz vage orientiert ist, bei der Lösung der ihm gestellten Aufgaben nur das bringt, was er aus sich selbst schöpft und auf diese Weise enthüllt, und daß er somit das Bild seiner Persönlichkeit „projiziert“.

Es ist bekannt, daß die projektiven Verfahren ihre Anhänger wie auch ihre erbitterten Gegner haben. Sie erfordern in jedem Fall große Umsicht und klinische Erfahrung und fallen daher in das Gebiet eines hochqualifizierten Psychologen.

Beispiel: Thematic Apperceptions Test (Thematischer Apperzeptions-Test) TAT, von H. Murray⁶⁷.

⁶⁶ Institut für Psychologie, Paris (Laboratoire de Psychologie).

⁶⁷ Harvard Univ. Press — Centre de Psychologie Appliquée. Siehe: H. Murray, Exploration de la personnalité, Paris, P.U.F., 2 Bd., u. D. Anzieu, Les méthodes projectives, Paris, P.U.F., 1960, S. 94 ff. Anm. d. dtsh. Bearbeiters: Vgl. Revers, W.J.: Der Thematische Apperzeptionstest (TAT). Handbuch zur Verwendung des TAT. Bern u. Stuttgart 1958.

Der im Jahre 1935 von Murray und seinen Mitarbeitern entwickelte TAT hat sich auf der ganzen Welt sehr schnell verbreitet. Er wird auf drei bedeutenden Forschungsgebieten verwendet, und zwar für Persönlichkeitsstudien, zu anthropologischen Untersuchungen und auf dem Gebiet der Sozialpsychologie.

Die Bildtafeln des TAT stellen andeutungsweise entweder Einzelpersonen dar oder Gruppen in verschiedenen Zusammenstellungen sowie Landschaften. Eine Testbildtafel ist leer und weiß.

Der Proband wird aufgefordert — oftmals unter dem Vorwand, daß seine Phantasie getestet werden solle — eine möglichst zusammenhängende Geschichte zu erzählen und die Gefühle der dabei auftretenden Personen zu charakterisieren.

Dabei geht Murray von der Hypothese aus, daß der Proband in seinen Bildern Begebenheiten aus seinem eigenen Leben wiedergibt, daß er sich selbst mit einer der Personen identifiziert und in den anderen den ihn unmittelbar umgebenden Personenkreis sieht: Eltern, Vorgesetzte, Untergebene usw.

Analyse:

— Der Form: Erschöpfende Darstellung des Gesamtbildes oder nur Teildarstellung, zusammenhängende Erzählung, Klarheit und Reichtum im Ausdruck usw.

Diese Anfangsanalyse liefert Hinweise auf die Fähigkeiten, die Persönlichkeit und das geistige Niveau des Prüflings.

— Des Inhalts:

Strebungen und Gefühle des Helden: Herrschsucht, Unterwerfung, Unabhängigkeit, Freiheitsstreben, Liebebedürfnis, Angst, Schuldgefühl, extravertierte bzw. introvertierte Haltung usw.

Umwelteinfluß auf den Helden: Herrschsucht, Protektion, Zurücksetzung.

Verlauf- und Höhepunkt der Geschichte: Logisch oder unlogisch, zum Siege oder zur Niederlage des Helden führend; Konfliktlösung oder Weiterbestehen des Konflikts usw.

— Allgemeine Themen.

Der Untersucher synthetisiert sodann die Ergebnisse und prüft, bis zu welchem Punkt die Diagnose durch andere Tests bzw. Prüfungen validiert wird.

Unter den kritischen Stellungnahmen zum TAT wirft man dem TAT u. a. folgendes vor: Einfluß der Intelligenz, der Information, Doppelprojektion: des Untersuchers und des Prüflings.

2. Das Kind und die Gemeinschaft

Die Soziometrie

I. Einführung

Die Schule empfängt das Kind in dem Augenblick, in dem es die existentielle Wirklichkeit anderer entdeckt und entläßt es erst wieder im Erwachsenenalter oder kurz zuvor. Der Schüler verbringt dort mehr als die Hälfte seiner Zeit, und die Beziehungen, die er dort anknüpft, bestimmen weitgehend seine Einstellungen und seine sozialen Verhaltensweisen.

Die Tätigkeit des Erziehers ist aus psychologischen und pädagogischen Gründen notwendig.

Seine Einschaltung in den psychologischen Bildungsplan ist zunächst einmal vorbeugender, heilender und konstruktiver Natur:

— Vorbeugend insofern, als der Erzieher z. B. verhindert, daß ein Kind einer Gruppe zum Opfer fällt, die es abweist, aus Gründen, deren Diagnose übrigens von Wichtigkeit ist;

— Heilend insofern, als er wenig erfreuliche Verhaltensweisen bekämpft, bevor sie sich im Charakter fest eingepreßt haben;

— Konstruktiv vor allem insofern, als er die maximale Entfaltung des einzelnen und der Gruppe fördert.

Die Gruppenuntersuchung ist für die Pädagogik in mancherlei Hinsicht interessant:

— Welche Relation besteht zwischen der Stellung eines Schülers in seiner Gruppe und seiner Verhaltensweise sowie seinen Schulleistungen?

— Inwieweit kann die soziale Stellung jedes einzelnen dem Wohle anderer dienen? R. Beezer und H. Hjelm⁶⁸ kommen in einer neueren Untersuchung zu folgendem Schluß: „Die soziometrischen Verfahren könnten dazu dienen, die beliebtesten unter den besten Schülern zu identifizieren. Diese Schüler zu ermutigen, den Weg der Hochschulbildung einzuschlagen, wäre vermutlich eine indirekte Art, ihre Kameraden zur Weiterentwicklung ihres Bildungsniveaus anzuregen.“

— Es besteht die Hoffnung, daß das starre System der Klasseneinteilung, die nach dem einzigen Kriterium des Alters vorgenommen wird, mit allen ihren Abweichungen, die sie mit sich bringt, bald durch ein flexibles

68 R. Beezer u. H. Hjelm, Factors Related to College Attendance, Washington, U. S. Office of Education, Coop. Research Program, 1961, S. 40.

Gruppensystem ersetzt wird, das nicht nur nach dem Gesichtspunkt der Begabungen, sondern auch nach dem der Persönlichkeit konstruiert ist. Die Soziometrie wird dazu beitragen müssen, die optimalen Bedingungen der Gruppenarbeit zu erschließen.

Es drängt sich jedoch eine Vorbemerkung auf. In dem Maße, wie das Kind heranwächst, gehört es gleichzeitig mehreren, manchmal sehr unterschiedlichen Gruppen an und kann sich je nach dem Milieu, in dem es sich befindet, sehr verschieden verhalten. Der Unscheinbarste in der Klasse entpuppt sich auf dem Spielplatz als ein dynamischer Anführer; ein anderer, der sich bei Schulwettkämpfen abseits hält, weil er sich seiner Kräfte noch nicht ganz sicher ist, um sich mit rücksichtslosen Kameraden zu messen, wird dagegen ein Führer bei den Pfadfindern und paßt sich einem homogenen Milieu leicht an.

Es ist demnach gefährlich, aufgrund von einigen wenigen Beobachtungen oder Messungen allgemeine Rückschlüsse auf den sozialen Status des Kindes zu ziehen. Wir beobachten besonders, daß die soziometrischen Daten nur einen begrenzten Wert besitzen, es sei denn, daß dieser aufgrund einer direkten, längeren Beobachtung verifiziert wird.

II. Definition der Soziometrie

Die Soziometrie ist die quantitative Behandlung aller Arten von menschlichen Beziehungen, insbesondere derjenigen, die eine Sympathie oder Antipathie zu anderen Gruppenmitgliedern im Rahmen einer Wahlsituation⁶⁹ ausdrücken.

Diese Quantifizierung geschieht entweder aufgrund direkter Beobachtungen oder mit Hilfe standardisierter Untersuchungsverfahren.

Die Quantifizierung des ersten Typs wird häufig vernachlässigt. Wenn man sich zu ausschließlich auf die soziometrischen Tests beruft, läuft man Gefahr, zu übersehen, daß sich die menschliche Natur in ihrer Vielgestaltigkeit nicht aufgrund einiger flüchtigen Berechnungen erfassen läßt.

A. Die direkte Beobachtung

Der Erzieher beobachtet ständig die Einstellung seiner Schüler gegenüber ihren Mitschülern und ihm selbst.

69 Ake Bjerstedt, Interpretation of Sociometric Choice Status, Munksgaard, 1956, zitiert in Enc. of Ed. Res., op. cit., S. 1319.

Wenn der Lehrer vor einer neuen Klasse steht, stellt er schnell einige markante Persönlichkeiten fest, einige Anführer, nach denen sich die Gruppe zu strukturieren beginnt.

Diese Eindrücke müssen jedoch einer Prüfung unterzogen werden. Man weiß, wie die Lehrer oftmals sehr zu Unrecht auf ihrem einmal gefaßten Urteil über ihre Schüler beharren. Derjenige, der in dem Ruf eines gefährlichen Anführers steht, sieht sich ständig von der Strafe des Lehrers bedroht, und bevor er zwischen dem Verlust seiner Persönlichkeit und dem Aufruhr wählen kann, entscheidet er sich notgedrungen für das letztere und verhält sich genau so, wie man es verhindern wollte.

G. Bastin hat darauf hingewiesen, daß die Beurteilungen der Lehrer nur begrenzten Wert haben. Er bestätigt Gronlunds Theorie und zeigt, daß die meisten Lehrer nach drei Monaten Unterricht noch nicht in der Lage sind, mit einiger Sicherheit die beliebten, isolierten oder aus ihrer Klassengemeinschaft ausgestoßenen Kinder zu identifizieren⁷⁰.

Der Beobachter darf jedoch nicht übersehen, daß eine Gruppe, an der sich ein Erwachsener beteiligt, oder die sich — bewußt oder unbewußt — beobachtet fühlt, sich von einer Gruppe, in der diese erwachsene Person fehlt, unterscheidet⁷¹.

I. Wann und wie beobachtet man?

1. Außerhalb der Klasse

Es scheint, daß die ersten Beobachtungen des sozialen Verhaltens des Kindes sich leichter während der Pausen bzw. während des Spielens vor oder nach dem Unterricht machen lassen.

Diese Beobachtungen werden vorzugsweise auf besondere Blätter notiert.

Hans hält sich immer abseits.

Peter fügt sich leicht in alle Gruppen ein.

Sobald Ludwig einer Gruppe beitrifft, ist das Spielen gestört.

Beim Aufstellen von Fußballmannschaften reißen sich beide Parteien um Heinrich.

Diese vorläufigen Feststellungen sind nicht uninteressant, jedoch erstrecken sie sich im allgemeinen nur auf sehr vordergründige Verhaltensweisen.

70 G. Bastin, Le statut social des adolescents, les observations des professeurs et le test sociométrique (Cahiers de Pédagogie et d'Orientation professionnelle, Lüttich, Nr. 12, 1953, S. 15–19).

71 Es ist eine allgemein bekannte Tatsache, daß der eine Lehrer überall Undiszipliniertheit hervorruft, während ein anderer dort, wo seine Kollegen nur auf Opposition gestoßen sind, die Zusammenarbeit und Aufgeschlossenheit weckt.

Nach und nach sollten sich die Bemerkungen aufgrund verschärfter Beobachtungen vertiefen und es erlauben, z. B. das schüchterne Kind, das sich beherrscht, oder das empfindliche Kind, das — ohne es sich anmerken zu lassen — durch irgendeine Streiterei oder Spöttelei zutiefst gekränkt ist, zu identifizieren.

Bei Gemeinschaftsspielen ist es interessant, nicht nur die Gruppierung der Schüler zu beobachten, sondern auch die Entwicklung des Spiels.

Wer beginnt? Ist es häufig das gleiche Kind? Wer macht leicht Fehler? Wer tadelt streng, selbst bei kleineren Fehlern? Wer schlägt das Spiel vor? Wer organisiert es? Wer spricht für die kleinsten den Auszählreim?

Auf dem Schulweg ist der Erzieher mitunter Zeuge kleiner aufschlußreicher Szenen, die Sympathien und Charaktereigenschaften enthüllen. Einige Bemerkungen hierüber auf dem Blatt des Schülers beschreiben sie:

Beispiel: Heinrich D... — 2. Juni 1966

In dem Augenblick, als ich den Weiher im Botanischen Garten erreiche, flüchten Heinrich, Peter und Hans. Sie haben mit Steinen nach den Goldfischen geworfen, als plötzlich der Gärtner auftaucht. Heinrich ist der flinkste. Hans fällt. Peter läuft weiter, Heinrich jedoch hält an, um seinem Freund zu helfen.

In den Schulen, in denen eine strenge Disziplin wenig Gelegenheit zu Spontanreaktionen bietet, ergibt manchmal ein Schulausflug wertvolle Aufschlüsse. Ein Eisenbahnabteil oder die Sitzreihen eines Omnibusses laden ebenso sehr zur Cliquenbildung nach besonderer Zuneigung ein. Die szenische Veränderung, z. B. der Aufenthalt in der Natur oder in einer fremden Stadt, läßt oftmals Verhaltensweisen zutage treten, die der Lehrer schwerlich unter anderen Umständen beobachten kann.

Die Gespräche mit den Eltern ergeben ebenfalls wertvolle Hinweise über den Umgang und die Charakteranlagen der Kinder.

Die Beziehungen zur Nachbarschaft üben auf das Kind im allgemeinen bis zur Adoleszenz einen gewissen Einfluß aus.

Etlliche Forscher, u. a. R. Cunningham, haben auf einem Stadt- oder Dorfplan die Beziehungen der Kinder eingetragen, so wie sie sie identifizieren konnten. Festgelegte Farben zeigen die Art der Beziehungen an: Spiel, gemeinsames Arbeiten, familiäre Freunde, Bandenbildung usw.

Diese Untersuchungen haben vornehmlich zu der Feststellung geführt, daß die geographische Verteilung der Freundschaften entsprechend dem sozioökonomischen Status der Familien⁷² signifikant variiert.

⁷² Siehe R. Cunningham, *Understanding Group Behavior of Boys and Girls*, New York, Columbia Univ. Press, 1951, S. 182.

2. In der Klasse

Auch in der Klasse treten die Wahlen, die Ablehnungen und die Gleichgültigkeit bei zahlreichen Gelegenheiten zutage. So beobachtet man z. B., mit welchen Klassenkameraden sich die Schüler zusammensetzen, wenn sie ihren Platz frei wählen dürfen.

Der Einfluß eines Klassenführers kann so groß sein, daß die Gruppe, die sich um ihn bildet, sich seiner Meinung vorbehaltlos anschließt, ohne sie vorher kritisch zu prüfen.

Mit zunehmendem Alter der Schüler rationalisieren sich auch die Wahlen, was die Identifizierung der echten Gefühle noch erschwert; so versuchen die Schüler bei der Bildung einer Arbeitsgruppe, sich dem höchst qualifizierten Kameraden anzuschließen, ohne Rücksicht auf echte Sympathien.

Gewisse Tätigkeiten eignen sich besonders gut für die Beobachtung.

Es kommt beispielsweise nicht selten vor, daß ein Schüler dazu aufgefordert wird, seine Kameraden abzuhören (Fremdsprachenlektionen, Vortragsübungen usw.). Die offene oder versteckte Ablehnung, das teilnahmevolle Lächeln, die Wahl der „Opfer“, wenn das Abhören befürchtet wird, sowie gewisse Hartnäckigkeiten können Beweise der Sympathie, der Spannungen oder Antipathie sein.

Die Durchführung eines Vorhabens in Gruppenarbeit ist auf allen Schulstufen ebenso aufschlußreich. Sei es, daß es sich um die Einstudierung eines Theaterstückes handelt oder um die Teilnahme an einer Wohlfahrtsveranstaltung, sei es, daß es gilt ein Fest zu organisieren oder eine Feldstudie zu unternehmen; der von der Gruppe bereitwillig akzeptierte Beobachter hat Gelegenheit, den „geborenen“ Organisator zu diagnostizieren, den Anführer, der das Vertrauen aller Kameraden genießt, oder denjenigen, den niemand als Mitarbeiter haben will, und schließlich denjenigen, der von seinen Kameraden ignoriert wird, als sei er nicht anwesend.

In manchen Schulen pflegen die Kinder einen Freundschaftspreis zu verteilen. Wir sind keine Anhänger dieser Gepflogenheit, jedoch da, wo es Sitte ist, ist es sinnvoll, die Verteilung der Stimmen zu notieren.

Diese unmittelbar aus dem Schulleben gegriffenen Beobachtungen sind wichtig. Wir werden sehen, daß man bei den soziometrischen Tests zur Erzielung guter Reaktionen im allgemeinen versucht, die Schüler zu motivieren, indem man sie gewisse Gemeinschaftsarbeiten und hypothetische Aufgaben vorstellungsweise in Angriff nehmen läßt (zu häufig ohne praktische Folgen). Viele Untersucher haben aus diesen künstlichen Experimenten ungenaue Rückschlüsse gezogen, weil sie es versäumt haben, ihre Resultate mit dem wirklichen Gruppenleben zu vergleichen.

II. Quantifizierungsversuch

Die direkten Beobachtungen können auf eine Soziomatrix übertragen werden, in die man die Wahlen und die Ablehnungen nach einem einfachen System einträgt, das im Zusammenhang mit den soziometrischen Tests näher erläutert wird.

Diese Tabelle wird im Verlauf der Beobachtungen vervollständigt. Je nachdem sich die Felder mit Eintragungen füllen, richtet sich die Aufmerksamkeit jeweils auf die Schüler, für die noch keine Notierungen gemacht wurden. Hierbei ist es wichtig festzustellen, ob das Fehlen der Beobachtungen im Untersucher selbst begründet oder auf den Mangel an Äußerungen von seiten des Kindes zurückzuführen ist.

Die Daten der Soziomatrix können schließlich in der Form eines Soziogramms ausgedrückt werden, wobei es interessant ist, dies mit den Resultaten des eigentlichen Soziometrietests zu vergleichen.

III. Kritik

Die direkte Beobachtung weist mehrere schwache Punkte auf.

1. *Man kann nicht alle Kinder gleichzeitig beobachten.* Die in Form einer Soziomatrix oder eines Soziogramms durchgeführte Synthese beruht auf der Hypothese, daß die Beziehungen zwischen den Gruppen und innerhalb der Gruppen hinreichend konstant sind.

Wenn auch einige Verhaltensweisen tiefere Gefühle widerspiegeln, so sind andere doch sehr unbeständig. Ein Kind, das gewöhnlich nicht sehr beliebt ist, kann nach einem aufsehenerregenden Ereignis, oder weil es ein ungewöhnliches und neiderregendes Spielzeug mitbringt, zum Mittelpunkt des Interesses werden.

Das Kind ist in ständiger Entwicklung begriffen. Einige Wochen, vor allem in Krisenzeiten, können wesentliche Veränderungen hervorrufen. Die Beobachtungen müssen daher noch einmal überprüft werden.

2. *Wurde bei jedem Probanden die gleiche Anzahl von Beobachtungen angestellt?* Manche Autoren schlagen vor, daß bei gut definierten Voraussetzungen der Beobachtung eines jeden Kindes systematisch die gleiche Zeit gewidmet werden soll. Aber diese Vorkehrung allein ist noch nicht ausreichend. Es kann durchaus der Fall eintreten, daß sich zu der festgelegten

Zeit nichts ereignet, während die darauf folgende Zeit mit dramatischen Ereignissen angefüllt ist.

Theoretisch ist es möglich, eine statistisch gültige, chronologische Stichprobe zu definieren. Aber diese komplizierte Technik steht kaum mit der täglichen Praxis in Einklang. Darüber hinaus zieht sich die Untersuchung lange hin, es sei denn, daß sich eine große Zahl von Beobachtern gleichzeitig an der Untersuchung beteiligt, wobei nicht selten die primäre Kritik ihren ganzen Wert zurückerhält.

3. *Die direkte Beobachtung ist zeitraubend,* und es müssen oftmals Entscheidungen getroffen werden, bevor sie Ergebnisse zeitigt. In manchen Fällen bleibt der Untersuchungsleiter nicht lange genug in Kontakt mit der Gruppe, um sich mit jedem Kind vertraut zu machen.

4. Ohne einen festen Rahmen wird *die Beobachtung aufgrund subjektiver Faktoren des Untersuchers leicht verfälscht.*

5. Schließlich kann das kritische Ereignis, das dem Zeugen auffällt, weitaus weniger bedeutsam sein als es ihm zunächst erscheint: ein Wutausbruch oder ein schroffes Ablehnen ist mitunter etwas ganz und gar Außergewöhnliches.

Dennoch wäre es sinnlos und gefährlich, die Statistik als erhabenstes Kriterium der Verhaltensforschung anzusehen. Der Boden, auf dem sich der Untersucher bewegt, ist äußerst ungesichert. Die diagnostische Begabung soll die statistischen Verfahren kompensieren und umgekehrt.

Die Techniken, die wir im Folgenden näher untersuchen wollen, dienen der Objektivierung der Beobachtung und führen zu einer strengeren Quantifizierung.

B. Die soziometrischen Verfahren

Man unterscheidet fünf Verfahren:

1. Den soziometrischen Fragebogen,
2. den eigentlichen soziometrischen Test,
3. den Test zur Messung der soziometrischen Wahrnehmung (*socio-empathy*),
4. den Personenratetest (*Guess-Who-Tests*),
5. die objektiven Tests zur Messung der sozialen Beziehungen.

I. Die soziometrischen Fragebogen

Die soziometrischen Fragebogen differieren erheblich untereinander. In manchen Fällen möchte der Untersucher wissen, wie die Mitglieder einer Gruppe aufeinander reagieren, was sie empfinden und wie fundiert ihre Gefühle sind, ohne die Reaktionen in einen bestimmten Rahmen zu bringen. In anderen Fällen ist der Fragebogen im Hinblick auf eine ganz bestimmte Tätigkeit gedacht: er nähert sich dem soziometrischen Test, ist jedoch in seinem Charakter weniger starr und weniger objektiv als dieser. In seiner einfachsten Form kann sich der Fragebogen auf eine einzige Frage beschränken:

Wer ist Dein bester Freund?

Oder: Dies ist die Liste Deiner Klassenkameraden. Kennzeichne die Namen Deiner Freunde mit einem Kreuz.

Andere Fragebogen sind weiter ausgearbeitet.

Beispiel I: Heimweg von der Schule.

1. Wo wohnst Du? Ort: Straße:
2. Wie gelangst Du nach der Schule gewöhnlich nach Hause?
 - a) Alleine — mit anderen Schulkameraden — mit einem Bruder oder einer Schwester — mit Deinem Vater oder Deiner Mutter?
 - b) Zu Fuß — mit der Straßenbahn — mit dem Fahrrad — mit dem Auto?
3. Wenn Du ohne Deine Eltern nach Hause gehst:
 - a) Setzen Deine Eltern einen Zeitpunkt fest, zu dem Du unter allen Umständen zu Hause sein mußt?
 - b) Bestimmen sie den Rückweg?
Wenn ja, welchen?
4. Wenn Du mit Deinen Freunden nach Hause gehst:
 - a) Mit welchen Klassenkameraden?
 - b) Mit welchen Schülern aus anderen Klassen?
(Gib bitte den Namen und die Klasse an: z. B. P. Lor, Klasse 5 A)
 - c) Spielt Ihr auf dem Heimweg?
 - d) Welches ist Euer Lieblingsspiel?
 - e) Gibt es manchmal Streitigkeiten?
Wenn ja, an welche erinnerst Du Dich am besten?
 - f) Erinnerst Du Dich an einen komischen Zwischenfall, der sich auf Deinem Heimweg zugegetragen hat?
Wenn ja, erzähle ihn kurz:

Beispiel II: Fragebogen, der die erfolgreiche Zusammenarbeit einer Arbeitsgruppe bestimmen soll.

1. Habt Ihr seit Beginn der Arbeit denselben Gruppenleiter behalten? Wer ist es oder wer sind sie seit Arbeitsbeginn gewesen?

2. Seid Ihr der Meinung, daß der gegenwärtige Gruppenleiter sich dazu eignet? Sich nicht dazu eignet? Warum?
3. Verläuft die Zusammenarbeit der Gruppe zu Eurer Zufriedenheit?
4. Beschreibt das Verhältnis der Gruppenmitglieder untereinander.
Ausgezeichnet — sehr gut — gut — ziemlich gut — nicht gut?
5. Seid Ihr der Meinung, daß eine andere Gruppenzusammensetzung ein schnelleres Arbeiten gewährleisten würde?
6. Welches ist Eure Aufgabe in der Gruppe?
7. Entspricht sie Euren Vorstellungen?
8. Welche würdet Ihr eventuell lieber ausüben?
9. Wer hat bisher nach Eurer Meinung die meisten eigenen Ideen gehabt?
10. Seid Ihr der Meinung, daß sich die Mitglieder Eurer Gruppe besser verstehen, seitdem Ihr an diesen Vorhaben arbeitet? usw.

II. Der eigentliche soziometrische Test

Der soziometrische Test hat sich so sehr verbreitet, daß man manchmal angenommen hat, er stelle die ganze Soziometrie dar. Bevor wir die Vor- und Nachteile dieses Verfahrens näher betrachten, wollen wir ihre Anwendung kurz erläutern.

Ziel dabei ist es, zu bestimmen, welche Wahlen und welche Ablehnungen die Mitglieder einer Gruppe während der Teilnahme an einer bestimmten Tätigkeit formulieren, und somit die Struktur der Gruppe zu erfassen.

Die wichtigsten Faktoren sind, daß die Schüler wirklich motiviert sind und daß sie sich in aller Aufrichtigkeit äußern.

Man stellt z. B. die folgende Frage: „Für die praktischen Arbeiten der naturwissenschaftlichen Fächer müssen wir Gruppen von jeweils sechs Schülern aufstellen. Gib in der Reihenfolge ihrer Beliebtheit die fünf Schüler an, mit denen Du gerne zusammenarbeiten würdest. Wir übernehmen natürlich keine Garantie dafür, daß Deine Gruppe auch tatsächlich nach Deinem Wunsch aufgestellt wird, weil man auch den Vorschlägen Deiner Kameraden Rechnung tragen muß. Dennoch wollen wir uns bemühen, Deine Wünsche zu berücksichtigen“.

Inwieweit die Antworten freimütig gegeben werden, hängt von der Klassenatmosphäre und insbesondere von dem Vertrauensverhältnis zwischen Schülern und Lehrer ab. Wenn ein Schüler sich dessen bewußt ist, daß seine Meinung wirklich berücksichtigt wird, reagiert er im allgemeinen aufgeschlossen. Aber viele Untersuchungsleiter begehen den Fehler, eine derartige Frage vollkommen gekünstelt zu stellen. Wenn jedoch die Schüler nach einem ersten Experiment die Erfahrung machen, daß ihr Wunsch nicht erfüllt wird, antworten sie auf spätere Fragen nur notgedrungen, und die ge-

gebenen Antworten sind ziemlich aussagearm. Dieser Punkt ist wichtig genug, um darauf zu bestehen: Wenn der Untersucher selbst nicht der Leiter der Klasse ist, in der er die Untersuchung durchführt, muß er sich mit diesem vorher abstimmen, damit er von den Resultaten der soziometrischen Prüfung auch wirklich Gebrauch machen kann.

In der Praxis versucht man gewöhnlich Antworten für zwei verschiedene Situationen gleichzeitig zu erhalten: z. B. Arbeitsgruppen und Sportmannschaften.

Man wertet die Antworten mit Hilfe einer *Soziomatrix* aus, die in ihrer einfachsten Form angewendet, dazu dient, die Anzahl der Wahlen zu notieren. Eine etwas weiter entwickelte Form läßt darüber hinaus die Rangordnung der Wahlen erkennen.

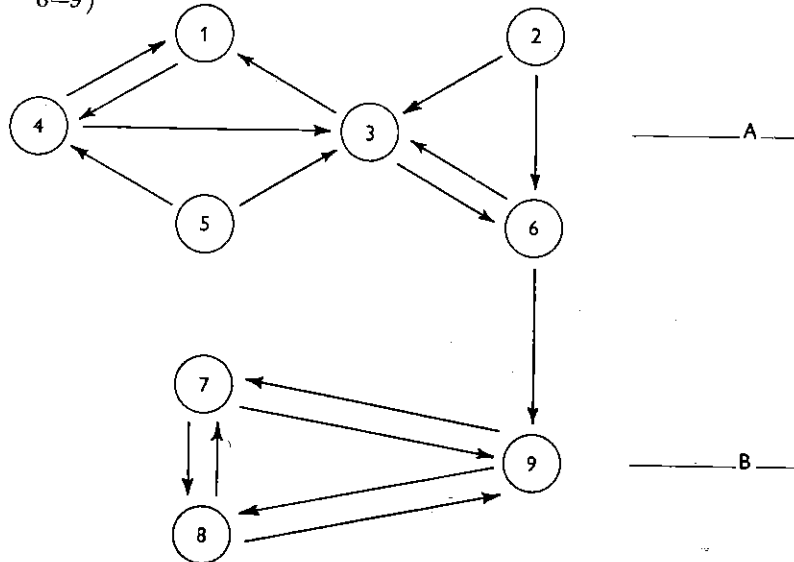
Schließlich können auch, um aussagekräftigere Punktwerte zu erhalten, die Antworten gewichtet werden (siehe Beispiel weiter unten).

Die Soziomatrix wird durch eine graphische Darstellung veranschaulicht: Das *Soziogramm*.

Das Soziogramm zeigt in seiner ursprünglichen Form lediglich die Wechselbeziehungen der Zellen an.

Obwohl noch wenig entwickelt, läßt das folgende Soziogramm doch klar erkennen:

1. Das Bestehen von zwei Untergruppen: A (1-2-3-4-5-6) und B (7-8-9)



Soziogramm einer Pfadfindergruppe

2. Innerhalb der Untergruppe A die sehr unterschiedliche Position der Probanden 2 (isoliert) und 3 (beliebt);

3. Den starken Zusammenhalt der Untergruppe B.

Dennoch erfahren wir hier nichts über die Tiefe bzw. den Grad der Empfindungen. Jedoch besteht ein außerordentlicher Unterschied zwischen einer oberflächlich verbundenen und einer aufgrund von gemeinsamen Erlebnissen eng verbundenen Gruppe.

Das Scheibensoziogramm, das sich aus drei oder vier konzentrischen Kreisen zusammensetzt, vermittelt ein feiner abgestuftes Bild der Realität. Die zur Zeit am weitesten verbreitete Darstellung mit drei Kreisen läßt im Zentrum die „Beliebten“, diejenigen, die von den Amerikanern als „Stars“ bezeichnet werden – und in den Außenbezirken die Isolierten klar erkennen.

Die Definition von „beliebt“ variiert je nach Autor: die einen schlagen die vor, die eine Standardabweichung oder mehr über dem Gruppendurchschnitt liegen; andere halten für „Stars“ die 20% der Besten der soziometrischen Verteilung.

Wenn fünf nicht gewichtete Wahlen zulässig sind, betrachtet Bronfenbrenner diejenigen mit neun Wahlen oder mehr als beliebt, und als abgelehnt diejenigen, die nur eine oder keine Wahl erhält; diese Abgrenzungen sind signifikant auf dem 2% Niveau.

Für den täglichen Gebrauch und bei weniger umfassenden Untersuchungen kann man sich an die folgende Regel halten:

1. Die Zahl der Wahlen nicht zu begrenzen.

2. Zur Bestimmung der Beliebten und der Isolierten:

a) Wenn die Gruppe 20 oder mehr Probanden zählt:

Man betrachtet als beliebt diejenigen, die fünf oder mehr als fünf erste oder zweite Wahlen erhalten;

Man betrachtet als isoliert diejenigen, die nicht mehr als eine der ersten oder zweiten Wahl erhalten;

b) Wenn die Gruppe weniger als 20 Probanden zählt:

Beliebte: Vier oder mehr als vier erste oder zweite Wahl;

Isolierte: Keine erste oder zweite Wahl.

Zur Erläuterung führen wir im Folgenden einen soziometrischen Test auf, der in einer Klasse von 17 Schülern durchgeführt wurde, sowie die Soziomatrix, in der die erhaltenen Antworten zusammengefaßt sind, und ein Soziogramm, welches sie strukturiert.

Es wird die einfachste Auswertungstechnik angewendet, und zwar diejenige, die wir an letzter Stelle zitiert haben.

Fragebogen

Name	Vorname	Schuljahr	Zweig	Datum
------	---------	-----------	-------	-------

Erste Frage:

Mit welchen Klassenkameraden würden Sie⁷³ am liebsten zu einer Arbeitsgruppe zusammengefaßt werden?

Geben Sie beliebig viele Kameraden in der Reihenfolge Ihrer Beliebtheit an und beginnen Sie mit demjenigen, mit dem Sie am liebsten in einer Gruppe zusammen wären; (nur Familienname, sofern es sich nicht um einen mehrfach auftretenden Namen handelt).

Antwort:

Zweite Frage:

Mit welchen Klassenkameraden würden Sie am liebsten zu einer Sportmannschaft zusammengefaßt werden?

Antwort:

Dritte Frage:

Versuchen Sie zu erraten, von welchen Klassenkameraden Sie bei Frage 1 und 2 gewählt worden sind.

Antwort:

Wir stellen fest, daß dieser Fragebogen nur Wahlen erfaßt. Die dritte Frage erstreckt sich auf die soziometrische Wahrnehmung (siehe weiter unten).

Anstelle von nur drei Fragen hätte man auch sechs Fragen bringen können:

1. Die Kameraden, mit denen man zusammenarbeiten möchte (Wahl);
2. Die Kameraden, mit denen man nicht zusammenarbeiten möchte (Ablehnung);
3. Die Kameraden, mit denen man gemeinsam Sport treiben möchte (Wahl);
4. Die Kameraden, mit denen man nicht gemeinsam Sport treiben möchte (Ablehnung).
5. Versuchen Sie zu erraten, von welchen Klassenkameraden Sie gewählt wurden (Arbeitsgruppe und Sportmannschaft).
6. Versuchen Sie zu erraten, von welchen Klassenkameraden Sie abgelehnt wurden (Arbeitsgruppe und Sportmannschaft).

⁷³ Anm. d. dtsh. Bearbeiters: Man schreibe „Du“, wenn es sich um eine Befragung jüngerer Schüler handelt, ab Obersekunda „Sie“.

Die Auswertung der Antworten geschieht folgendermaßen:

1. Gewichtung:

- 5 Punkte für eine erste Wahl;
- 4 Punkte für eine zweite Wahl;
- 3 Punkte für eine dritte Wahl;
- 2 Punkte für eine vierte Wahl;
- 1 Punkt für eine fünfte Wahl;
- x für eine Wahl unterhalb der fünften.

2. Aufbau der Soziomatrix

- a) Man fertigt eine Tabelle für alle Schüler der Gruppen an (Meistens gibt man die Anfangsilbe des Familiennamens an).
- b) Jedes Kästchen der Tabelle wird in drei Zellen geteilt:

1	2
3	

In Zelle (1) gibt man die Wahl für die Arbeit an.
 In Zelle (2) gibt man die Wahl für das Spiel an.
 In Zelle (3) gibt man die erratene Wahl an.

Beispiel:

	Lutz	Paul
Markus	5 - (4)	- 1 -

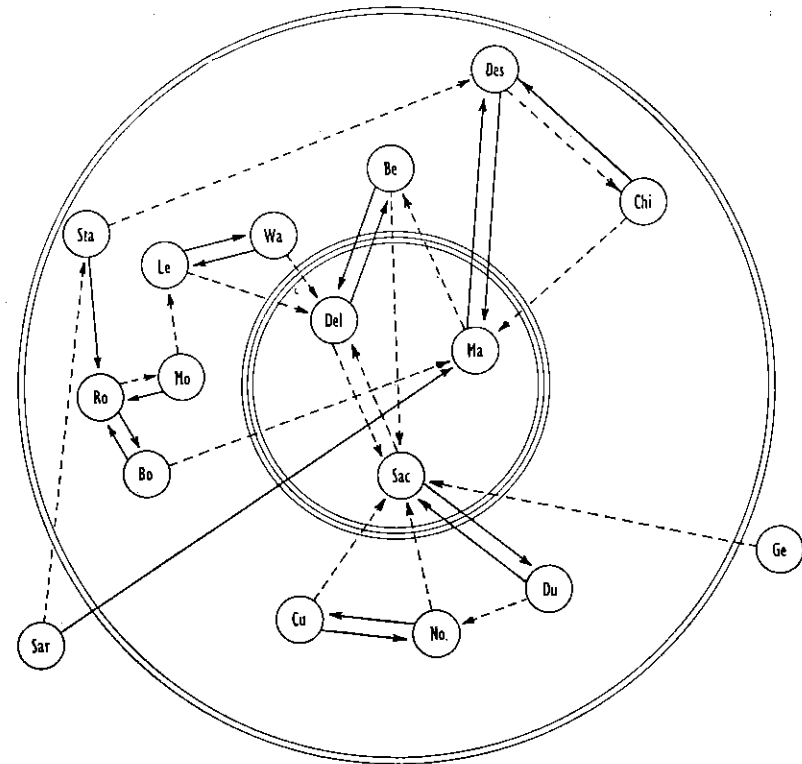
Erklärung:

Markus hat Lutz als Arbeitskameraden an erster Stelle gewählt.
 Markus hat Lutz nicht als Sportkameraden gewählt.
 Markus hat Lutz an zweiter Stelle angegeben bei dem Versuch, zu erraten, von wem er für die Arbeit und den Sport gewählt wurde.
 Markus hat Paul nicht als Arbeitskameraden gewählt.
 Markus wählte Paul an fünfter Stelle als Sportkameraden.
 Markus ist nicht der Meinung, daß Paul ihn gewählt hat.

	Be	Bo	Chi	Cu	Del	Des	Du	Ge	Le	Ma	Mo	No	Ro	Sac	Sar	Sta	Wa
Be	/	x	x	3	5	x	x	x	x	1	x	2	x	4	3	x	
Bo	2	/		3	1	3	4	x	4	2		5	5				
Chi	3		/	3	5	2	5		4	1	4						
Cu	3			/	2	1	3	2				5		4	5		
Del	5	5	x	x	/	x	3	3	x	x	x	x	1	1	x	4	4
Des	3	3	4	4		/				5	5						
Du	x		x	3	3	4	/		2	x	1	2	4	5	5		
Ge				2				/			3			4	4		
Le				4					/		3			3	4		
Ma	4	4	1	3	3	5	5			/		2	2	2	x		
Mo		(4)	(3)					4	4		/		5	5			
No	x	x		5	5	2	3	3	1	2		/		4	4		
Ro		5	5						2	4	4		/	3			
Sac	3	x	x	x	4	4	2	1	5	5	x	2	x				
Sar	1	1		2	x	x	2	3	x	5	4	5	3	/	4	4	
Sta	5				4		4					5	3		/		
Wa	3	x	x	x	5	4	x	x	x	3	5	x	x	x	x	/	

Diese Soziomatrix erlaubt eine hohe Anzahl von verschiedenen praktischen Anwendungen. Das folgende Soziogramm zeigt die Struktur der wichtigsten Wahlen für das Einzelkriterium Arbeit (Frage 1):

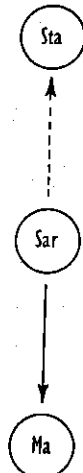
- Die vollen Linien stellen eine erste Wahl dar.
- Die punktierten Linien stellen eine zweite Wahl dar.
- Die Pfeile geben die Richtung der Wahl an.
- Im Zentrum die Beliebten; in den Randzonen die Isolierten.



Dieses Soziogramm läßt die Existenz von drei beliebten Schülern und zwei Isolierten erkennen, sowie zwei Untergruppen, die ziemlich deutlich gekennzeichnet sind: Die eine zählt 11 Schüler, die andere 4.

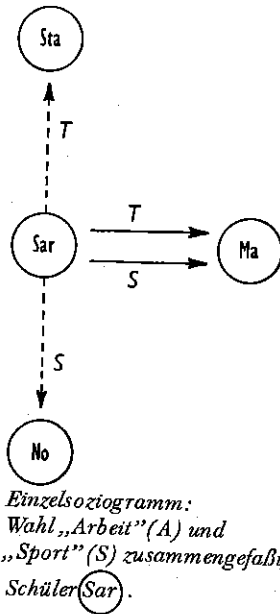
Aus diesem allgemeinen Soziogramm lassen sich leicht Einzelsoziogramme extrahieren, die die Situation der einzelnen Gruppenmitglieder stärker beleuchten.

Das Einzelsoziogramm zeigt, daß keiner Sar an erster oder zweiter Stelle gewählt hat. Die Zeichnung der soziometrischen Wahrnehmung gibt an, daß Sar der Meinung ist, er sei von vier Kameraden gewählt worden, während ihn niemand gewählt hat (man erkennt aus der Soziomatrix, daß Ma, der von Sar an erster Stelle für die Arbeit und den Sport gewählt wurde, nur eine einzige Wahl für Sar im vierten Rang abgegeben hat).

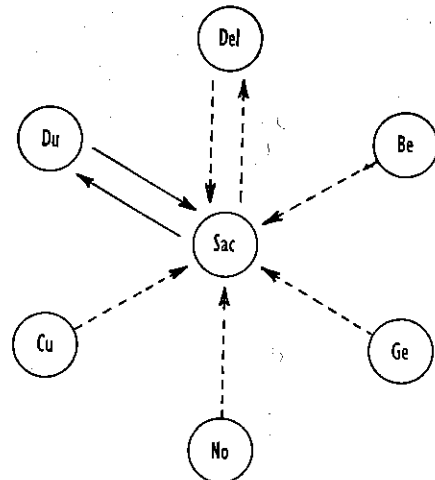


Einzelsoziogramm: Wahl „Arbeit“ — ein Isolierter, ein Beliebter

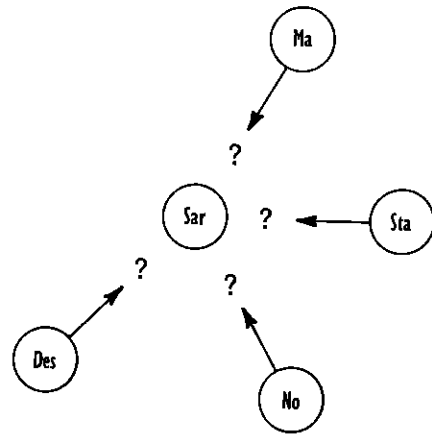
Darüber hinaus geben wir eine aufschlußreiche Situation wieder, die sich dank der Prüfung der soziometrischen Wahrnehmung (Frage 3) identifizieren läßt.



Einzelsoziogramm:
Wahl „Arbeit“ (A) und
„Sport“ (S) zusammengefaßt,
Schüler (Sar).



Darstellung der soziometrischen
Wahrnehmung, Schüler (Sar).



Eine primäre Analyse dieser Art rückt Situationen ins Blickfeld, bei denen eine Bestätigung wichtig erscheint. Die direkte Beobachtung erlaubt es in der Mehrzahl der Fälle, die aus dem Soziogramm gezogenen Rückschlüsse zu validieren. Die Situation muß sodann erklärt werden, und eventuell sucht man nach Mitteln und Wegen zur Abhilfe. Hier beginnt die eigentliche Arbeit.

Es gibt weitaus strengere Auswertungs- und Interpretationsmethoden. Sie gründen sich auf die Errechnung zahlreicher soziometrischer Meßwerte und allgemein auf eine gut entwickelte statistische Behandlung. Diese Methoden werden in dem Werk von G. Bastin. *Les techniques sociométriques*, Paris, P.U.F., 1961, eindeutig dargelegt.

Kritik:

1. Die aufgrund des soziometrischen Tests gewonnenen Informationen besitzen nur für die jeweils untersuchte Gruppe Validität, und zwar für den Zeitpunkt und die Situation der Testdurchführung.

Ein häufig gewählter Jugendlicher tritt beispielsweise als Führer einer Jugendbewegung in Erscheinung; es ist hingegen möglich, daß er sich in einem anderen Milieu nicht bewährt, ja, daß er dort sogar abgelehnt wird.

H... ist nicht sehr intelligent. Sein Aussehen ist nicht ansprechend. Seine Ausdrucksweise ist schwerfällig, und er wird offensichtlich von seinen Klassenkameraden gemieden. Als ausgleichende Entschädigung hierfür übernimmt H. die Führung einer kleinen Pfadfindergruppe, die sich aus Kindern aus einem niedrigen sozio-kulturellen Familienniveau zusammensetzt. H. erfreut sich in seiner Gruppe besonderer Beliebtheit und fühlt sich damit teilweise bestätigt — obwohl er sich der leichten Selbsttäuschung bewußt ist. Im Grunde findet er nur in diesem Milieu Anerkennung als Führer.

Der Fall von H. tritt häufig auf, und zwar nicht nur bei Jugendlichen, sondern auch bei Erwachsenen. Er erklärt weitgehend das Zustandekommen und die Existenz von tausenderlei kleinen gesellschaftlichen Zusammenschlüssen und Vereinen, die nur dank der unermüdlichen Aufopferung eines Vorsitzenden oder eines Sekretärs, der in ihnen seine Bestätigung findet, aufrecht erhalten werden.

2. Die in dem Test geäußerten Wahlen sagen nichts darüber aus, ob eine echte Beziehung besteht zwischen demjenigen, der wählt und demjenigen, der gewählt wird. Es kann sich lediglich um einen oberflächlich geäußerten Vorschlag handeln, und etliche Autoren heben die schwache Korrelation zwischen den geäußerten und den erhaltenen Wahlen hervor: $r = -.30$ bis $+.35$ (Cunningham u. a.; Jennings).

So erklärt sich auch zum Teil der oftmals erhebliche Unterschied zwischen den Ergebnissen der direkten Beobachtung und denjenigen der Tests.

3. Wir haben bereits darauf angespielt: Eine Wahl läßt nicht unbedingt auf echte Sympathie schließen. Man kann den Wunsch haben, mit einem Klassenkameraden zusammenzuarbeiten, weil man ihn als guten Organisator schätzt oder für die gestellte Aufgabe als besonders qualifiziert hält, obgleich keinerlei freundschaftliche Beziehung besteht.

4. Ein Soziogramm stellt flüchtige Wahlen, die sich vielleicht schon am folgenden Tag nicht mehr wiederholen würden, und dauerhafte Freundschaften auf die gleiche Ebene.

5. Schließlich können die soziometrischen Tests in drastischer Form eine Ablehnung zum Bewußtsein bringen und somit einen traumatischen Effekt verursachen.

Abschließend kann festgestellt werden, daß der soziometrische Test Symptome erkennen läßt und Meßwerte erstellt, die Gegenstand einer strengen psychologischen Überprüfung sein sollen.

III. Der Test der soziometrischen Wahrnehmung (socio-empathy)⁷⁴

Bei dieser Untersuchungsart fordert man die Probanden auf, diejenigen zu benennen, denen sie eine vorgegebene Eigenschaft zuerkennen oder die sie entweder für eine bestimmte Tätigkeit gewählt oder abgelehnt hätten.

Auf diese Weise sieht man, inwieweit der Prüfling in echter Übereinstimmung mit seiner Gruppe steht.

Eine ausgeprägte Divergenz zwischen den mutmaßlichen Annahmen und den tatsächlichen Gegebenheiten erlaubt es, eine bestimmte Anzahl Konflikte im Innern der Gruppe zu erklären, Konflikte, die bei einer direkten Beobachtung kaum in Erscheinung treten und insofern schwieriger zu bekämpfen sind. Man stelle sich beispielsweise die seelische Verfassung eines Kindes vor, das sich von einem Mitschüler abgelehnt glaubt, während dies nicht zutrifft.

Wir haben hinsichtlich des soziometrischen Tests Fragenbeispiele gegeben, die sich auf die soziale Wahrnehmung erstrecken.

IV. Der Personenratetest (Guess Who Test)

Der *Guess Who Test* ist eine indirekte Form der Nominierungstechnik. Diese besteht aus einem einfachen Feststellungsakt, bei dem der Proband

⁷⁴ Wir verwenden diese Klassifikation nach Merl Bonney, op. cit., der sich auf Ausubel, Borgatta und auf Norman bezieht. Millie Almy, op. cit., nennt die Tests, die wir im folgenden mit Personenratetest bezeichnen, „Soziale Wahrnehmungstests“.

diejenigen Gruppenmitglieder angibt, die nach seiner Meinung in erheblichen Maße eine vorgegebene Eigenschaft besitzen.

Beispiel: Führeigenschaft

Schreibe^{74a} in die unten dafür vorgesehenen Zwischenräume die Namen von fünf Mitgliedern Deiner Gruppe, die nach Deiner Meinung die Eigenschaft eindeutig besitzen.

1. 2. 3. 4. 5.

In dem *Guess Who Test* wird diese Frage in konkreter Form dargelegt. Man beschreibt kurz eine Eigenschaft, und zwar so charakteristisch wie möglich, und fordert dann das Kind auf, zu erraten, wer in seiner Gruppe der vorgeschlagenen Beschreibung entspricht.

Beispiel: Es folgt eine kurze Beschreibung eines oder mehrerer Deiner Klassenkameraden. Errate ihren Namen und schreibe ihn auf.

1. Sei es, daß man sich verletzt hat, daß man ein Buch vergessen hat, daß man sonstige Schwierigkeiten hat, er ist (sie sind) stets zum Helfen bereit:

2. Er will (sie wollen) stets, daß man das tut, was er (sie) will (wollen), und ist beleidigt, wenn man seine (ihre) Wünsche nicht erfüllt:

3. Sobald man auf den Schulhof kommt, schlägt er (schlagen sie) ein amüsantes Spiel vor und führt (führen) es an:

4. Er ist (sie sind) immer dabei, seine (ihre) Mitschüler zu ärgern:

Man zählt, wie oft ein Kind für jede Beschreibung genannt wurde und stellt danach eine Häufigkeitstabelle auf.

Es ist bekannt, wie richtig und unverfälscht oftmals die Beurteilungen der Kinder, dieser großen Psychologen, sind. So bemerkt L. Cronbach⁷⁵, daß der Erzieher hier die Gelegenheit findet, seine Meinung mit der von zwanzig oder dreißig Mitarbeitern zu vergleichen, deren sicheres Urteil einen Zuverlässigkeitswert von .90 erreichen kann, wenn die Gruppe sich gut kennt und die Beschreibungen eindeutig sind.

^{74a} Anmerkung des deutschen Bearbeiters: Bei älteren Schülern schreibe man „Sie“.

⁷⁵ L. Cronbach, *Essentials of Psychological Testing*, New York, Harper, 1960, 2. Aufl. S. 518.

Der *Guess Who Test* stellt hingegen die Mängel und Schwächen der Soziometrie ganz allgemein dar:

1. Sofern es sich nicht um eine eng verbundene Gruppe handelt, die bereits eine große Toleranz gegenüber der Kritik entwickelt hat, kann die offene Identifizierung der Mängel anstößig sein.
2. Die Kinder geben vermutlich nur solche Klassenkameraden an, die die vorgegebene Eigenschaft in erhöhtem Maße besitzen. Es wäre daher falsch, wenn man die Meinung vertreten würde, daß die Kinder, die nicht aufgeführt sind, keiner aufmerksamen Prüfung bedürften.
3. Eine vorübergehende Situation kann die Ergebnisse verfälschen. Die Prüfung soll daher validiert werden.
4. Schließlich darf man nicht vergessen, daß eine Anerkennung die Verhaltensweise beeinflusst. Das Kind neigt dazu, sich nach der Vorstellung, die andere sich von ihm machen, zu verhalten.

V. Objektive Tests der sozialen Beziehungen

Zur Verdeutlichung erwähnen wir einerseits zwei kürzlich durchgeführte objektive Tests, die wertvolle Hinweise auf die Dynamik einer Gruppe von Kindern zu geben scheinen und andererseits zwei ältere Verfahren, die wegen ihrer Bemühung um Objektivität sehr interessant sind.

1. *Russel Sage Social Relations Test* von D. Damrin⁷⁶.

Dieser Test, der für den Volksschulunterricht abgefaßt ist, soll die Fähigkeit zur Zusammenarbeit messen.

Mit Hilfe von 36 verschiedenfarbigen Formblöcken soll eine Gruppe oder eine Klasse drei Konstruktionen nach einem vorgegebenen Modell durchführen.

Jedes Kind erhält zu Beginn einen oder zwei Formblöcke. Die Gruppe wird aufgefordert, einen Arbeitsplan zu entwerfen. Sobald der Plan fertig ist, muß er innerhalb von 15 Minuten ausgeführt werden.

Die Notierung erfolgt zu zwei verschiedenen Zeitpunkten:

- Während der Arbeitsplanung: Teilnahme, Kommunikation, bekundetes Interesse, Autonomie, Organisationstalent, endgültiger Plan;
- Während der Arbeitsausführung: Interesse, Atmosphäre, Aktivität, Erfolg.

Während der Arbeitsplanung unterscheidet man sieben Gruppentypen, ausgehend von der reifen bis zur anarchistischen Gruppe.

Während der eigentlichen Arbeit unterscheidet man neun Typen, die sich in drei

⁷⁶ D. A. Damrin, *The Russel Sage Social Relations Test* (Journal of Experimental Education, Bd. 28, Sept. 1959, S. 85–100).

Gruppen aufteilen: konzentriert auf die gestellte Aufgabe, nicht in Opposition, Opposition.

Die Eichung war zum Zeitpunkt der Veröffentlichung noch nicht beendet; der Autor legte den Test als ein Forschungsinstrument vor.

2. *The Syracuse Scales of Social Relations* von E. Gardner und G. Thompson⁷⁷.

Standardisierungsversuch der Durchführung und der Notierung der soziometrischen Verfahren.

Der Test umfaßt Skalen für drei pädagogische Bildungsstufen: Volksschulstufe, Mittlerer Bildungsweg-Unterstufe (bis zur mittleren Reife), Mittlerer Bildungsweg-Oberstufe (bis zum Abitur)^{77a}.

Das Originelle an der Prüfung ist, daß der Proband bei der Wahl der Kameraden als Vergleichsobjekt „eine Person nimmt, die er seit langem kennt“.

Prozentranguormen aufgrund einer umfangreichen Eichstichprobe erlauben die Einstufung jedes einzelnen Probanden. Darüber hinaus können mehrere Meßtypen für die in Betracht kommenden Einzelpersonen und Gruppen erstellt werden.

Man stellt der von Damrin vorgeschlagenen Prüfung vergleichsweise zwei ältere Tests gegenüber:

3. *Zwei-Personentest* von H. Henning.

Dieser Test, der erstmals im Jahre 1927⁷⁸ beschrieben wird, umfaßt eine Reihe von Geräten, die die paarweise Intervention von zwei Personen erfordern (z. B. Scheren, die von zwei Personen gehandhabt werden), die sich bald in Situationen der Mitarbeit, bald in solchen der Konkurrenz befinden. Die verschiedenen Bedingungen erlauben es, den Sinn für die Gemeinschaftsarbeit, die Liebenswürdigkeit, die Gefälligkeit, die Grobheit, die Aufrichtigkeit usw. zu bewerten.

4. *Zwei-Personentest, Sozialtest, Ehetest* von H. Hanselmann

Zwei Personen sollen gemeinsam eine Zeichnung entwerfen. Die eine arbeitet, während die andere zuschaut. Man vertauscht die Rollen nach jeder Minute, und die eine Person muß dort fortfahren, wo die andere aufgehört hat.

⁷⁷ E. Gardner u. G. Thompson, *Syracuse Scales of Social Relations*, New York, World Book Co., 1959.

^{77a} Anm. d. dtsh. Bearbeiters: Bis zum 6. Schuljahr geht in Belgien die Volksschulstufe, bis zur mittleren Reife das „enseignement moyen-inférieur“ (in verschiedenen Schulen, z. B. Gymnasium, Handelsschule etc.), bis zum Abitur das „enseignement moyen-supérieur“. Das Hochschulstudium wird „enseignement supérieur“ genannt.

⁷⁸ H. Henning, *Test de caractère* (IV. Internationale Psychotechnische Konferenz, 1927).

Vor allem den zwei letzten Untersuchungsverfahren begegnet man mit zwei wichtigen Einwänden; die Phänomene, die sie messen sollen, werden nicht exakt definiert, und die Validitätsuntersuchungen sind praktisch nicht vorhanden.

Eins der Hauptprobleme besteht darin, festzustellen, ob die unter experimentellen Bedingungen beobachteten Verhaltensweisen für diese Situation spezifisch sind oder ob sie im Gegensatz dazu eine konstante Orientierung der Persönlichkeit anzeigen.

C. Andere Aspekte bei der Untersuchung der sozialen Entwicklung

Bis heute hat man vor allem die Anpassung des Kindes an entsprechende Altersgruppen in Betracht gezogen.

Aber das Problem des Akzeptierens, der Ablehnung, des Sozialstatus des Kindes zeigt sich auch in seinen Beziehungen zur Welt der Erwachsenen und insbesondere zu seiner Familie.

I. Das Kind in seiner Familie

Die folgenden Untersuchungsinstrumente werden eingesetzt, um auf schnelle Art und Weise einen allgemeinen Überblick über die intrafamiliären Beziehungen zu gewinnen.

1. Fragebogen der Kind-Eltern-Beziehungen von G. Bastin und H. Delrez (Editions Scientifiques et Psychotechniques).

63 Fragen erlauben eine erste Auswertung der Beziehungen zwischen den Eltern und ihren Kindern und die Art, wie die letzteren diese Beziehungen wahrnehmen.

Es wurde nach folgenden Gesichtspunkten untersucht: Atmosphäre der Zuneigung oder der Gefühlskälte, des Akzeptierens oder der Ablehnung, des Vertrauens oder Mißtrauens, der Herrschsucht oder des „laissez faire“, des Lobes oder der Geringschätzung, Atmosphäre der „over protection“ oder das Gegenteil, der Ruhe oder der Nervosität usw.

Eichung: Normalisierte Punktwerte in 9 Intervallen (allgemeine Norm, Norm Vater, Norm Mutter) aufgrund der Untersuchung von 365 Jungen im Alter von 13 bis 15 Jahren.

Itemsbeispiele:

- Mein Vater sagt mir, daß ich es niemals zu etwas bringen werde.
- Mein Vater spricht zu anderen Leuten mit Stolz über mich.
- Meine Mutter regt sich schnell auf, wenn ich nicht sofort gehorche.
- Meine Mutter ist so wechselhaft in ihren Stimmungen, daß ich nie weiß, was ich tun soll, um es gut zu machen.

2. The Family Relations Indicator (FRI) von J. Howells und J. Lickorish (National Foundation for Educational Research in England and Wales).

Dieses klinische Untersuchungsverfahren ist nur für die Hand eines spezialisierten Psychologen bestimmt. Es wird angewendet bei Kindern im Alter von 3–16 Jahren; die besten Antworten werden im allgemeinen von 7- bis 12jährigen Kindern erzielt.

Der Test besteht aus 33 Bildtafeln; auf jedem Bild sind zwei oder drei Personen in einer einfachen familiären Situation dargestellt.

Der Proband wird aufgefordert, zu sagen, was die Personen tun oder sagen; man versucht nicht, das Kind eine zusammenhängende Geschichte erzählen zu lassen.

Die gleiche Situation wird in verschiedener Form dreimal wiederholt, und die charakteristischen Items der Verhaltensweisen werden auf einem besonderen Blatt notiert.

Bei einer Validitätsuntersuchung, die sich auf 50 Familien erstreckte, bestätigten sich die bei der Untersuchung gewonnenen Angaben zu 80–90% als richtig.

3. The Family Relations Test von E. Bene und J. Anthony (National Foundation for Educational Research in England and Wales).

Auch diese Prüfung soll wie die vorhergehende nur von geschulten Psychologen vorgenommen werden.

Man legt dem Kind verschiedene Darstellungen von Personen vor und fordert es auf, seine Familie zusammenzustellen. Man teilt sodann Karten aus, auf die man jeweils eine Gefühlsäußerung oder eine Einstellung eingetragen hat.

Das Kind legt die Karten unter die entsprechenden Personen. Die Durchschnittspunktzahl wird durch Zählen der Karten erstellt.

4. Check Lists of Intra-familial Attitudes von R. Cattell⁷⁹.

Wir haben bereits in dem Abschnitt über die Kontrollisten (checklists) auf diese Inventarien hingewiesen und bei dieser Gelegenheit ein detailliertes Beispiel gegeben.

II. Die soziale Entwicklung, wie sie in den Fragebogen zur Messung der Anpassung Berücksichtigung findet

Die Mehrzahl der Fragebogen zur Persönlichkeitsforschung, die sich mit der Messung der Anpassung befassen, lassen eine Vielzahl von Items hinsichtlich der sozialen Anpassung zu.

Dies trifft auch für den Test zur Messung der persönlichen Anpassung von C. Rogers, für die *Mooney Problem Check Lists* von R. Mooney und L. Gor-

⁷⁹ R. Cattell, *A Guide to Mental Testing*, op. cit., S. 361–379. Anm. d. dtsh. Bearbeiters: Vgl. auch den Test „Familie in Tieren“ von Luitgard Gräser (1956).

don, für den *SRA Youth Inventory* von H. Remmers und B. Shimberg, den *Pupil Adjustment Inventory*, den Fragebogen von H. Bell usw.

III. Vineland Sozial Maturity Scale, von E. Doll

Diese Skala der sozialen Entwicklung untersucht vor allem die zunehmende Erlangung von Unabhängigkeit gegenüber der Familie und den Gruppen.

IV. Die Soziale Intelligenz

Im Jahre 1920 unterscheidet E.L. Thorndike erstmalig eine „soziale Intelligenz“ neben der „theoretischen Intelligenz“ und der „praktischen Intelligenz“. Er stützt sich auf die Beobachtung an solchen Menschen, die geeignet erscheinen, soziale Beziehungen anzuknüpfen und die geistige Verfassung ihrer Mitmenschen zu verstehen.

Da die Beziehung zwischen Lehrer und Schüler wesentlich eine soziale ist, ist das Interesse an dieser Frage sowohl für die Auswahl als auch die Ausbildung der Lehrer von Bedeutung.

Der im Jahre 1927 veröffentlichte und 1949 in revidierter Form erschienene *Social Intelligence Test* von F. Moss u. a. (Center for Psychological Service) spiegelt direkt die Theorien Thorndikes wider.

Er umfaßt fünf Gruppen von Items:

- Beurteilung sozialer Situationen (Beispiel: Welcher Angestellte soll entlassen werden?).
- Untersuchung des Gedächtnisses eines Menschen: Namen- und Gesichtsgedächtnis.
- Allgemeine Menschenkenntnis: Sinn für Humor (Den besten Scherz auswählen).

Schon 1936 hat R.L. Thorndike durch eine Faktorenanalyse festgestellt, daß keiner der in der Untersuchung von Moss behandelten Faktoren von der sozialen Intelligenz abhängig war. Es dauerte bis 1965, bis Guilford und sein Forschungsteam feststellten, daß die soziale Intelligenz eine Gruppe von Fähigkeiten umfaßt, die von den vorhin erwähnten Intelligenzfaktoren unabhängig sind⁸⁰.

Sich auf das von Guilford vorgeschlagene dreidimensionale Modell des Intellekts (vgl. Anhang) stützend, haben die Forscher die Hypothese auf-

80 M. O'Sullivan, J.P. Guilford, R. De Mille; The Measurement of Social Intelligence, Los Angeles, Universität of Southern Calif., Report 34, 1965.

gestellt, daß die soziale Intelligenz 30 verhaltensbezogene Faktoren umfasse^{80a}.

Nur die sechs ersten Faktoren wurden untersucht und bei der ersten Untersuchung bestätigt, es handelt sich um die Serie, die sich mit den verhaltensbezogenen Erkenntnissen, mit der Fähigkeit, das Denken, die Gefühle und die Absichten anderer zu verstehen, befaßt.

Die vier anderen großen Bereiche wurden nachfolgend untersucht, dabei handelt es sich um:

1. Das Gedächtnis (Behalten) der Verhaltensweisen, die Fähigkeit, sich der sozialen Charakteristika der Individuen zu erinnern.
2. Das divergierende Denken bei sozialen Verhaltensweisen, sozial schöpferisches Tun, die Fähigkeit, sich einer (neuen) Situation anzupassen, indem man unterschiedliche oder ungewöhnliche Verhaltensweisen zeigt.
3. Das konvergierende Denken bei sozialen Verhaltensweisen, die Fähigkeit, sozial nach allgemeiner Gewohnheit, ohne eigene Gestaltungsweise zu handeln.
4. Die Beurteilung sozialer Verhaltensweisen, die Fähigkeit, die sozialen Verhaltensweisen nach rechtlichen oder anderen Kriterien zu bewerten.

Testbeispiele (im ganzen 23 bei 240 Studenten angewandte Tests)

- a) Erkennen von Verhaltenseinheiten, die Fähigkeit, körperliche Ausdrucksweisen zu verstehen.
Ausdruckstest: Wählen Sie von den vier Fotos, die menschliche Gesichter darstellen, dasjenige aus, das das gleiche Gefühl ausdrückt, wie die Fotografie der hier zu sehenden Frau.
- b) Erkennen von Klassen, von Verhaltensweisen, die Fähigkeit, verschiedene Möglichkeiten des Ausdrucks zu unterscheiden, die die gleichen Dispositionen eines Gegenstandes ausdrücken.
Ausschlußtest: Wählen Sie eines dieser Fotos aus, das nicht zu den drei anderen paßt.
- c) Erkennen verhaltensbezogener Beziehungen, die Fähigkeit, die zwischen zwei Individuen bestehenden Beziehungen zu verstehen.
Beziehungstest: Wählen Sie von drei Sätzen den aus, der einem Gesichtsausdruck entspricht, indem Sie die von einem anderen Gesicht ausgedrückten Gefühle beachten.
- d) Erkennen von Systemen des Verhaltens, die Fähigkeit, eine soziale Situation oder eine Reihe von Ereignissen zu verstehen.
- e) Erkennen der Verhaltensänderungen, die Fähigkeit, eine Geste, eine ganze soziale Situation zu reinterpreten.
- f) Erkennen der verhaltensbezogenen Implikationen, die Fähigkeit, die Folgen einer sozialen Situation voraussagen zu können.

Nachdem die Konstruktvalidität feststeht, bleibt noch zu entscheiden, welchen Voraussagewert diese Tests haben werden.

80a Anm. d. dtsh. Bearbeiters: Vgl. die Darstellung der drei Aspekte der intell. Begabung von J.P. Guilford, in: Weinert, F.: Pädag. Psychologie, Köln—Berlin 1967, S. 118ff.

3. Die Einstellungen und die Meinungen

Einführung

Ursprünglich nur von der Sozialpsychologie⁸¹ untersucht, sind die Einstellungen und Meinungen auch für die Pädagogische Forschung interessant. Thurstone definiert die Einstellung als „den Grad der positiven (lieben, zugehen) oder negativen Gefühle, die mit einem psychologischen Objekt (Anmerkung des deutschen Bearbeiters: Das können sein: Personen, Symbole, Aussagen, Ideen) verbunden sind“.

Die Einstellung ist dynamisch. Es handelt sich hierbei um eine Veranlagung zum Agieren oder Reagieren unter dem Impuls der Zuneigung oder des Hasses, der Furcht oder des Ressentiments. Sie trägt daher einen mehr oder minder fundierten und dauerhaften effektiven Charakter.

Die Einstellung ist spezifisch („Ich liebe jenes Volk nicht“) oder generell („Man muß den Ausländern mißtrauen“).

Was die Meinung anbetrifft, so ist sie der verbale oder gestikuläre Ausdruck der Einstellung.

Klineberg⁸², der sich auf Newcomb, Hartley, Krech und Crutchfield beruft, behält sich fünf Dimensionen der Einstellung vor:

1. *Die Richtung*: Man ist für oder gegen die Aufnahme eines Ausländers in eine Gruppe, oder: für oder gegen die Koedukation von Jungen und Mädchen.

2. *Der Grad*: Einstellungsbreite.

Sind wir der Meinung, daß die Studenten mit *allen* Problemen der Hochschulverwaltung assoziiert sein sollen oder nur mit einigen davon?

3. *Die Intensität*: Sie zeigt an, inwieweit die Einstellung für den Probanden von Wichtigkeit ist.

Ein Lehrer kann der Meinung sein und sich offen dafür aussprechen, daß die Erziehung von behinderten Kindern eine begeisternde Aufgabe ist und dennoch weder die Neigung verspüren noch den Mut aufbringen, sich den armen kleinen Geschöpfen zu widmen.

⁸¹ Siehe insbesondere die folgenden allgemeinen Untersuchungen: D. Krech u. R. Crutchfield, *Théorie et problèmes de psychologie sociale*, Paris, P.U.F., 1952, 2. Bd.

M. Jahoda, M. Deutsch u. S. Cook, *Publ., Research Methods in Social Relations*, New York, Dryden Press, 1951.

L. Festinger u. D. Katz, *Publ., Les méthodes de recherche dans les sciences sociales*, Paris, P.U.F., 1959, 2. Bd.

⁸² O. Klineberg, *Psychologie sociale*, Paris, P.U.F., S. 551.

4. *Die Kohärenz*: Sie zeigt den Zusammenhang zwischen den Einstellungen und den Meinungen einerseits und dem tatsächlichen Verhalten andererseits an.

Man kann die Vorzüge einer demokratischen Schumatmosphäre übermäßig herausstreichen oder gleichzeitig in seiner eigenen Klasse autoritär sein.

5. *Die Prägnanz (salience)*: Es gibt an, in welchem Maße die Einstellung das eigentliche Verhalten des Einzelnen beeinflusst.

I. Methoden zur Messung der Einstellungen

A. Der Fragebogen und das Interview

Jemanden befragen, was er mag, was er nicht mag, was ihm gleichgültig ist, scheint die direkteste Methode zu sein, seine Einstellung und seine Meinungen zu erfahren.

Oftmals jedoch geschieht es, daß die befragte Person ihre wahren Gefühle bewußt verbirgt und ihre Persönlichkeit vor Indiskretion schützt, indem sie neutral reagiert („weder ja noch nein“). Eine solche Antwort fällt im allgemeinen in die Kategorie „zweifelhaft, unentschieden, gleichgültig“, was aber nicht der wirklichen Einstellung des Probanden entspricht.

Schließlich sei noch bemerkt, daß der Fragebogen vor allem dazu dient, die *Richtung* der Einstellungen zu definieren.

B. Die direkte Beobachtung der Verhaltensweisen

Auch die direkte Beobachtung der Verhaltensweisen ist bei weitem kein unfehlbares Mittel, die Einstellung kennenzulernen.

Ein Schüler kann sich für ein Wohltätigkeitswerk einsetzen, doch keineswegs aus reiner Nächstenliebe, sondern in der Hoffnung die Wertschätzung seiner Lehrer zu gewinnen. Man kann sich aus Gründen des Eigennutzes für eine Idee einsetzen, die man mißachtet. Es hat Erzieher gegeben, die — nur um ihren Irrtum nicht eingestehen zu müssen — mit einer Unterrichtsmethode fortfuhren, die erfahrungsgemäß erhebliche Schwächen aufwies.

Ein von außen auferlegter Zwang ruft oftmals Einstellungen hervor, die ganz und gar nicht dem freien Willen des Probanden entsprechen. Manch einer liebt die Sonne Afrikas und ist gezwungen, sein Leben in Europa zu verbringen. Man kann — so sagt Edwards — den Geschmack von Rinder-

filet besonders schätzen und dennoch nur billiges Fleisch kaufen, wenn man aufgrund des knappen Haushaltsgeldes dazu gezwungen ist⁸³. Es besteht demnach nicht unbedingt eine Übereinstimmung zwischen dem Verhalten und der Einstellung.

Nichtsdestoweniger sind viele Verhaltensweisen aufschlußreich, vor allem dann, wenn die Probanden sich unbeobachtet fühlen oder wenn sie nicht genau wissen, worauf sich die Beobachtung bezieht. Darüber hinaus erreichen manche Einstellungen einen derartig hohen Intensitätsgrad, daß sie unbewußt richtunggebend für das ganze Verhalten des Probanden sind. So kann sich der Haß im Gesichtsausdruck oder im Tonfall der Stimme offenbaren.

Die direkte Beobachtung der Verhaltensweisen kann mit Hilfe von Beobachtungsbogen oder sorgfältig entwickelten check lists gelenkt werden.

C. Die Erforschung der Einstellung mit Hilfe von projektiven Verfahren

Stark ausgeprägte Einstellungen sind tief im Innern des Wesens verwurzelt. Es ist daher nicht verwunderlich, daß man sich überlegt hat, zu ihrer Messung projektive Verfahren einzusetzen.

Prohansky hat beispielsweise die Einstellung gegenüber der Arbeiterschicht untersucht, indem er dem TAT analoge Bilder verwendet, die allerdings Arbeiter im Konflikt⁸⁴ darstellen. Diese Methode hat jedoch noch keine brauchbaren Resultate erbracht.

D. Die Einstellungsskalen

Die Einstellungsskalen sind mehr oder minder direkte Fragebogen, die den Vorteil bieten, sowohl über die Art der Einstellung als auch über ihren Intensitätsgrad systematisch zu informieren.

*Konstruktion*⁸⁵: Gleich vielen anderen Forschungsinstrumenten stellen die Einstellungsskalen zunächst willkürliche Konstruktionen dar, die sich auf den gesunden Menschenverstand und die Logik des Untersuchers gründen. Tatsächlich handelt es sich dabei um einfachere Fragengruppen, die sich alle auf das gleiche Problem beziehen, z. B. die Einstellung gegenüber

83 A.L. Edwards, *Techniques of Attitude Scale Construction*, New York, Appleton-Century-Croft, 1957, S. 6.

84 H.M. Prohansky, A projective Method for the Study of Attitudes (*Journal for Abnormal and Social Psychology*, Nr. 38, 1943, S. 393–395).

85 Für gründlichere Behandlung, siehe: P. Debaty: *La mesure des Attitudes*, Paris, P.U.F., 1967.

dem Fortschritt, die durch die beiden Extreme Radikalismus — Konservatismus gekennzeichnet ist.

Die Auswertung erfolgte ohne Gewichtung der Items durch einfaches Addieren der für die eine oder andere Einstellung⁸⁶ günstigen Antworten.

Mit der Zeit wurden weitaus strengere Konstruktionsmethoden eingeführt. Für die Nichtspezialisten sind die von Thurstone und Likert⁸⁷ am leichtesten zugänglich. Wir geben hierüber eine kurze Übersicht.

1. Die Methode von Thurstone

Anstatt direkt eine logisch aufgebaute Skala zu konstruieren, beginnt der Untersucher damit, eine große Anzahl von Feststellungen zu sammeln — stets mehr als 100 —, die sich auf die zu untersuchende Einstellung beziehen. Diese Feststellungen findet er in der Literatur oder in den im Verlauf der Interviews gemachten Notizen; vielfach bringt er selbst eine bestimmte Anzahl zu Papier.

Die gesammelten Feststellungen sind etwa von dieser Art:

- „Die älteren Schüler sollten den jüngeren beim Lösen von Aufgaben, die sie nicht verstehen, helfen.“
- „Die Älteren sollten die Jüngeren im Straßenverkehr beaufsichtigen.“
- „Die Älteren sind selbst noch Kinder und sollen sich darum nicht der Jüngeren annehmen.“
- „Die Älteren sollen sich bei den Jüngeren Respekt verschaffen.“
- „Wenn nötig, dürfen die Älteren die Jüngeren schlagen, damit sie ihnen gehorchen.“
- „Die Älteren sollen alles dransetzen, um den Jüngeren zu helfen.“

Ein solches Verzeichnis von Feststellungen wird sodann einer großen Anzahl von Gutachtern (ca. 100) unterbreitet, die sie im allgemeinen in elf Kate-

86 Beispiele: M.H. Harper, *Social Beliefs and Attitudes in American Educators*, New York, Teachers College, Columbia Univ., 1927. — J.W. Wrightstone, *Wrightstone Scale of Civic Beliefs*, Yonkers, World Book Co., 1938.

87 L.L. Thurstone u. E.J. Chave, *The Measurement of Attitude*, Chicago, Univ. of Chicago Press, 1929.

R. Likert, A Technique for the Measurement of Attitude (*Archives of Psychology*, Nr. 141, Columbia Univ., 1932, 55 Seiten).

Über den gegenwärtigen Stand der Technik siehe vor allem A.L. Edwards, op. cit. Die sehr komplexe von Guttman entwickelte Technik wird hier nicht in Betracht gezogen. Siehe hierzu: L. Guttman, A basis for Scaling Qualitative Data (*American Sociological Review*, Nr. 9, 1944, S. 139–150). Die Methode von Guttman wird von H. Remmers u. N. Gage verständlich dargelegt in: *Educational Measurement and Evaluation*, New York, Harper & Br., 1955, S. 392–400.

gorien⁸⁸ aufteilen: fortschreitend von der günstigsten bis zur ungünstigsten Feststellung.

In der Praxis werden die Feststellungen einzeln auf verschiedene Kärtchen übertragen, dabei werden drei Gesichtspunkte berücksichtigt:

- Kategorie 1: Die günstigsten Beurteilungen;
- Kategorie 5: Durchschnittliche oder neutrale Stellungnahme;
- Kategorie 11: Die ungünstigsten Beurteilungen.

Sodann ist es wichtig festzustellen:

1. Welche Rangstellung die Gutachter insgesamt jedem Item in dem Kontinuum von 11 Intervallen zuordnen (Errechnung des Medianwertes: 50. C).

Zahl der Beurteiler: 98

Kategorie (Klassen)	Intervall- Grenzen	Feststellung 1		Feststellung 2	
		Wahl- Häufig- keiten	Kumulierte Häufig- keiten	Wahl- Häufig- keiten	Kumulierte Häufig- keiten
1	0- 0,9	2	2	9	9
2	1- 1,9	2	4	17	26
3	2- 2,9	5	9	19	45
4	3- 3,9	4	13	17	62
5	4- 4,9	9	22	14	76
6	5- 5,9	23	45	11	87
7	6- 6,9	18	63	3	90
8	7- 7,9	13	76	3	93
9	8- 8,9	11	87	2	95
10	9- 9,9	6	93	2	97
11	10-10,9	5	98	1	98

Intervall 1	N 98	N 98
	Q_2 (50. Centilw.) = 6,2	Q_2 (50. Centilw.) = 3,2
	Q_1 (25. Centilw.) = 5,1	Q_1 (25. Centilw.) = 1,9
	Q_3 (75. Centilw.) = 7,8	Q_3 (75. Centilw.) = 4,8
	$\frac{Q_3 - Q_1}{2} = 1,35 (Q)$	$\frac{Q_3 - Q_1}{2} = 1,45 (Q)$

88 Es werden auch Skalen à 5, 7 und 9 Gradierungen konstruiert.

2. In welchem Maße jedes Item von den Gutachtern sinngemäß übereinstimmend verstanden und klassifiziert wurde. Man errechnet zu diesem Zweck einen Zweideutigkeitsindex: die Hälfte des Abstands zwischen dem 25. und 75. Zentilwert (Mittlerer Quartilabstand).

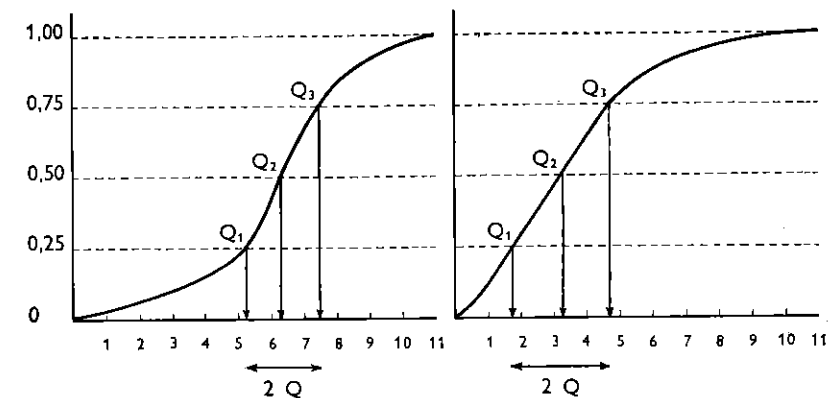
Zur Vervollständigung des Skalenaufbaus beruft man sich auf die den Punkten 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 und 11 am nächsten gelegenen Feststellungen. Wenn mehrere Feststellungen um denselben Punkt herum liegen, greift man auf diejenige zurück, die den kleinsten Zweideutigkeitsindex aufweist.

Das folgende Beispiel zeigt, daß sich für Feststellung 1 die von den Beurteilern vorgenommene Klassifizierung mehr konzentriert (zwischen der 5. und 7. Kategorie) als für Feststellung 2 (zwischen der 2. und 5. Kategorie) — ein Hinweis darauf, daß ihnen die erste Feststellung eindeutiger zu sein schien als die zweite.

Die polyvalenten Items, d. h. solche, die mehrere Einstellungen gleichzeitig messen, werden sodann aufgrund komplexer Methoden ausgeschaltet (Errechnung des Polyvalenzkoeffizienten), um die Skala möglichst eindimensional zu gestalten.

Die Reihenfolge, in der man die Items vorlegt, ist willkürlich.

Außerdem besteht beim Ausgehen von einer großen Zahl von Feststellungen oftmals die Möglichkeit, Parallelskalen aufzubauen.



Auszug aus einer nach der Methode von Thurstone konstruierten Skala, die sich auf die Einstellung der Lehrer bezieht⁸⁹.

89 D. Ryans, Characteristics of Teachers, Washington, American Council on Education, 1962, 2. Aufl. S. 166-167.

Zur Abb.:

Einverstanden nicht einverstanden

- 4. Heutzutage neigt man dazu, die Kinder zu frei zu erziehen (Punktwert: 4,1)
- 9. Das Unterrichten ist eines der besten Mittel im Dienst an der Menschheit (Punktwert: 1,0)
- 10. Ein Lehrer findet immer Zeit, den Problemen der Kinder Gehör zu schenken (Punktwert: 1,2)
- 12. Das Unterrichten führt zu einer zynischen Lebenseinstellung (Punktwert: 4,4)

2. Die Methode von Likert

Zur Konstruktion einer Skala nach Likert bringt man zunächst eine große Anzahl von Feststellungen zusammen, die bald eine günstige, bald eine ungünstige Einstellung ausdrücken.

In einer Voruntersuchung bringen die Beurteiler ihre Meinung in fünf Kategorien zum Ausdruck:

- 1. Stimmt entschieden zu (St. e. z.)
- 2. Stimmt zu (St. z.)
- 3. Unentschieden (Une.)
- 4. Lehnt ab (L. a.)
- 5. Lehnt entschieden ab (L. e. a.)

Die Antworten werden gewichtet: 5 Punkte werden der für die untersuchte Einstellung günstigsten Position zugeteilt usw.

Beispiel: Einstellung gegenüber der Schülermitverwaltung.

- Die Schüler müssen unbedingt an der Schulverwaltung aktiv beteiligt sein.
Antwort: Stimmt entschieden zu = 5 Punkte.
- Die Schüler haben noch nicht den Reifegrad, um sich an der Schulverwaltung aktiv zu beteiligen.
Antwort: Lehnt entschieden ab = 5 Punkte.

Man behält sich für die endgültige Skala lediglich solche Items vor, die die stärkste Übereinstimmung bei den Antworten aufweisen. Die Übereinstimmung wird ausgewertet, indem man die Korrelation zwischen jedem einzelnen Item und der Gesamtpunktzahl errechnet.

Skalenauszüge

a) Einstellung der Schüler zum technischen Unterricht⁹⁰

	+2	+1	0	-1	-2
4. Der Schulunterricht trägt dazu bei, dem Geist der Jugendlichen entsprungene Ideen festzuhalten.					
11. Den Informationen, die man uns über die Geschäftswelt gibt, fehlt es vielleicht an Objektivität.					
20. Die technische Schule gewöhnt die Schüler hinreichend an die Gruppenarbeit.					
22. Die Lehrer, die uns unterrichten, sind ihrer Aufgabe nicht ganz gewachsen.					

b) Einstellung der Lehrer zu homogenen Klassen⁹¹

	St.e.z.	St.z.	Une.	L.a.	L.e.a.
1. Es wäre sinnvoll, einen Förderkurs für die überdurchschnittlich Begabten einzurichten.					
2. In den Klassen für überdurchschnittlich Begabte ist der Wettstreit zu stark.					
3. Die Erzieher, die in den homogenen Klassen unterrichten, leisten mehr undankbare Arbeit als die anderen.					

Unter den neueren Arbeiten sollte man sich die Untersuchung von Edwards und Wilson⁹² merken, die sich auf sehr weitentwickelte mathematische Techniken stützt.

Zur Bestimmung der Schülereinstellung gegenüber den Unterrichtsstoffen haben Edwards und Wilson Skalen konstruiert, die es erlauben, die übliche Orientierung der Entscheidungen und Wahlen zu bewerten. Sie berücksichtigen fünf Orientierungsdimensionen und untersuchen sie anhand von sechs verschiedenen Skalen:

⁹⁰ A. Cheniaux, unveröffentlichte Denkschrift, Univ. Lüttich, 1961.

⁹¹ Wrightstone, Justman u. Robbins, Evaluation in Modern Education, New York, American Book Co., 1956, S. 371.

Abkürzungen vgl. S. 152.

⁹² T. Edwards u. A. Wilson, The Development of Scales of Attitudinal Dimensions (Journal of Experimental Education, Bd. 28, Sept. 1959, S. 3-36). Diese Skalen beleuchten gleichzeitig die Berufsinteressen.

1. Analytische Orientierung auf die nicht-sozialen (Naturwissenschaften und Mathematik) oder auf die sozialen Bereiche (Humanwissenschaften).⁴

Itemsbeispiele:

Die technische Errungenschaft, die die Konstruktion der Pyramide darstellt, interessiert mich mehr als die soziale Struktur oder das Wirtschaftssystem von Ägypten, die ihren Bau ermöglichten. Ich würde lieber Naturwissenschaften unterrichten als Forschung betreiben.

St.e.z.	St.z.	Une.	L.a.	L.e.a.

2. Orientierung auf die sozialen Bereiche oder auf das Gebiet der Ästhetik.

Itemsbeispiele:

Man sollte nicht zulassen, daß zur Herstellung von Baumaterial die Bäume unserer schönen Wälder gefällt werden. Die medizinischen Versuche an Tieren sind grausam und unmenschlich.

St.e.z.	St.z.	Une.	L.a.	L.e.a.

3. Analytische Orientierung auf die sozialen Bereiche oder auf die unmittelbare Praxis.

Itemsbeispiele:

In hoffnungslosen Fällen sollte man die Kranken nicht leiden lassen: aus Mitleid muß man ihrem Leben ein Ende setzen. Ein Geschäftsmann muß Entscheidungen treffen, die mit den Interessen seines Unternehmens übereinstimmen. Er darf sich nicht um die Auswirkungen seiner Maßnahmen auf die Löhne und Preise auf nationaler Ebene kümmern.

St.e.z.	St.z.	Une.	L.a.	L.e.a.

4. Theoretische oder praktische Orientierung.

Itemsbeispiele:

Der Mathematikunterricht sollte der Heranbildung des Menschen als Geschäftsmann oder Verbraucher mehr Rechnung tragen. Im Kino verliere ich manchmal den Faden der Geschichte, weil ich mir Gedanken über die Beleuchtungstechnik mache.

St.e.z.	St.z.	Une.	L.a.	L.e.a.

5. Ästhetische oder praktische Orientierung.

Itemsbeispiele:

Die chemischen Experimente sind interessant, wenn sie Explosionen auslösen oder schöne Farben erzeugen. Wenn ich die Sterne betrachte, mache ich mir mitunter Gedanken über das Funktionieren des Universums.

St.e.z.	St.z.	Une.	L.a.	L.e.a.

6. Theoretische oder ästhetische Orientierung.

Itemsbeispiele:

Ich fahre lieber alleine mit dem Segelboot, als daß ich mir ein Fußballspiel ansehe. Ich würde *Evangéline* von Longfellow vorziehen (als Lektüre), wenn die Liebesgeschichte in Prosa erzählt wäre.

St.e.z.	St.z.	Une.	L.a.	L.e.a.

3. Von der Likert-Methode abgeleitete Formen
Anstatt die Einstellung durch Formulierungen wie „Stimmt entschieden zu“, „Stimmt zu“ usw. auszuwerten, wird der Proband aufgefordert, zwischen fünf Lösungen zu wählen.

*Beispiel: Contemporary problems*⁹³ (Aktuelle Probleme)

Ein kurzer Text gibt bekannt, daß die Klasse demnächst einen wichtigen Ausflug unternehmen wird, und betont, daß es gut wäre, ein Schülerkomitee zu wählen, das die Vorbereitung dafür übernimmt.

Frage: Welcher Vorschlag ist zur Gewährleistung einer erfolgreichen Wahl der Komiteemitglieder am wichtigsten?

- Jeder Schüler kann Komiteemitglied werden, da wir alle an der Exkursion teilnehmen.
- Nur die besten Schüler sollten dem Komitee angehören.
- Man sollte solche Schüler wählen, die an allen früheren Exkursionen teilgenommen haben.
- Die Komiteemitglieder sollten von dem Klassenlehrer gewählt werden.
- Man sollte solche Schüler wählen, deren Eltern großes Interesse an der Schule zeigen.

⁹³ Veröffentlichung des Teachers College, Columbia Univ., in Anlehnung an Wrightstone, Justman u. Robbins, op. cit., S. 363.

Wie bei der klassischen Skala, teilt man den verschiedenen Antworten einen Notenwert von 1 bis 5 zu. Die Gesamtzahl der Punkte stellt die Rohpunktzahl des Probanden dar.

II. Allgemeine Kritik

In zahlreichen Fällen geben die Einstellungsskalen dem Untersucher wichtige Informationen an die Hand.

Man darf jedoch nicht vergessen, daß die Verhaltensweisen bei weitem nicht immer die wahren Gefühle des untersuchten Probanden wiedergeben. Auch die Resultate einer Einstellungsskala müssen anhand anderer Informationen, die man über den Probanden besitzt, systematisch validiert werden.

III. Die Einstellungsskalen und die Pädagogische Forschung

Die objektive Untersuchung der Einstellungen ist für die Schule in mancherlei Hinsicht von Bedeutung:

1. Die Erziehung steht unter dem Einfluß der Einstellungen von Lehrern, Eltern, Schülern und der Gesellschaft ganz allgemein.
 - Die Orientierung des Geschichtsunterrichts differiert, jenachdem, ob der Lehrer radikal, konservativ oder eklektisch eingestellt ist.
 - Die Reaktion der Eltern gegenüber der Schule ist ähnlich geartet.
 - Der Lerneifer der Schüler hängt teilweise von ihrer Einstellung gegenüber dem entsprechenden Fach oder Lehrstoff ab.
 - Die Unterstützung, die der Staat dem Bildungswesen gewährt, hängt von seiner grundsätzlichen Einstellung ab.
2. Die Schulbildung leistet einen wesentlichen Beitrag zur Bildung neuer Einstellungen beim Kind, einmal durch den direkten Unterricht, zum andern durch das Nachahmungsspiel oder das, was man als „ansteckende“ Verhaltensweisen bezeichnen könnte.
3. Die Schule macht es sich auch zur Aufgabe, die unerwünschten Einstellungen zu bekämpfen, insbesondere die ungerechtfertigten Ressentiments, die Vorurteile gegenüber Rassen, Menschen und Dingen („Alle Deutschen sind kriegerisch“. „Die schwarze Rasse ist weniger intelligent als die weiße“ usw.)

Es besteht daher ein großes Interesse, die Einstellungen zu messen und ihre Entwicklung objektiv zu beobachten.

Anwendungsbeispiele:

- Untersuchung des Einflusses eines Unterrichts für politische Bildung, der die internationalen Beziehungen zum Gegenstand hat. Die Schüler beantworten vor Beginn des Kurses eine Einstellungsskala; die gleiche Skala oder eine Parallelförmige wird ihnen am Ende des Schuljahres vorgelegt; man stellt fest, ob die festgestellten Differenzen statistisch signifikant sind⁹⁴.
- Auswirkung bestimmter Lesestoffe.
- Beeinflussung durch eine Reihe von Erziehungsfilmen⁹⁵.
- Auswirkungen der Reklame, einer Aussprache oder einer unparteiischen Beurteilung.
- Untersuchung des Einflusses, den die pädagogische Bildung auf die Einstellungen zu Erziehungsproblemen hat.

Einige publizierte Einstellungsskalen:

- *Master Attitude Scales*: Beginnend mit 12 Jahren (Division of Education, Purdue University).
- *What would you do?* Beginnend mit 12 Jahren (Harvard School of Education).
- *Illinois Opinion Inventories*: Beginnend mit 11 Jahren (World Book Co).
- *Behavior Preference Record*, von H. B. Wood: Ende des Volksschul- und Realschulunterricht (National Foundation for Educational Research in England and Wales).

94 Bei einem solchen Experiment ist die festgestellte Entwicklung nicht unbedingt auf den Einfluß des Unterrichts zurückzuführen. Andere Faktoren — wie Beeinflussung durch Film, Fernsehen usw. — können eine entscheidende Rolle spielen. Komplexe experimentelle Schemata erlauben es, derartige Fehlerquellen in weitem Maße zu vermeiden (siehe Kapitel Statistik).

95 R. Peterson u. L. Thurstone, *Motion Picture and Youth*, New York, Macmillan, 1933. C. Hovland, A. Lumsdaine u. F. Sheffield, *Experiments in Mass Communication*, Princeton Univ. Press, 1949.

4. Die Untersuchung von Werten

Der Wert einer Sache wird dadurch bestimmt, inwieweit sie erwünscht und bevorzugt wird.

Man hat vielfach versucht, bei den Werten, die ein einzelner als wichtig erachtet, dominierende charakteristische Merkmale zu erkennen: religiöse, ästhetische Werteinschätzung usw. Dies scheint tatsächlich auf eine Identifizierung der intensivsten, der am tiefsten verankerten Einstellungen hinauszulaufen, d. h. derjenigen, die eine der Grundlagen der Persönlichkeit bilden.

Es ist in jedem Fall schwierig, das Wertgefühl klar von der Einstellung zu unterscheiden. Darum soll dieses kurze Kapitel als Nachtrag zu dem Vorhergehenden aufgefaßt werden. Es wird im Folgenden nur durch einen allgemein bekannten Test illustriert.

Die Beziehung zwischen den Werten und den Interessen ist ebenfalls evident.

Study of Values von Allport, Vernon und Lindzey, 1960, 3. Auflage (Houghton Mifflin Co.).

Stark beeinflusst von Sprangers *Lebensformen*, hat dieser Fragebogen zum Gegenstand, im Hinblick auf die Persönlichkeit sechs dominierende Wertebereiche zu messen: den theoretischen, ökonomischen, sozialen, politischen, ästhetischen und religiösen⁹⁶.

Der Test ist anwendbar bei Studenten, die im Hochschulstudium stehen und bei Erwachsenen, die zumindest teilweise ein Hochschulstudium betrieben haben.

Die Fragen beziehen sich auf vertraute Situationen. Im ersten Testteil wird eine zweifache Beantwortungsmöglichkeit angeboten: Der Proband kann seine Wahl treffen zwischen 3 = einverstanden, 0 = nicht einverstanden, 2-1 = leichter Vorzug für 2, 1-2 = leichter Vorzug für 1.

Im zweiten Teil werden vier Wahlmöglichkeiten angeboten. Man ordnet sie in der Reihenfolge ihrer Priorität.

Insgesamt gibt es 120 Fragen: 20 pro Wertebereich.

Man stellt fest, daß sich die Wertprofile nach den Berufen unterscheiden.

Dieser Test fand Verwendung bei der Messung der Entwicklung der Wertmaßstäbe im Laufe des Studiums, sowie bei der Beurteilung, in welchem Maße Werte innerhalb der Familienmitglieder oder eines Freundeskreises gemeinsam auftreten.

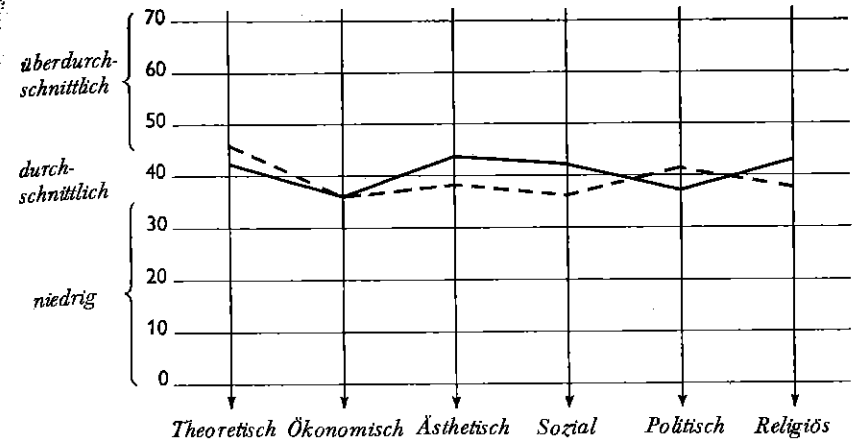
⁹⁶ Bemerkenswert ist die enge Verwandtschaft mit den Skalen von Edwards u. Wilson, die am Ende des Kapitels über die Einstellungen dargelegt wurden.

Itemsbeispiele:

I. 12. Ihre Morgenzeitung enthält die beiden folgenden Schlagzeilen:

- a) Der Oberste Gerichtshof hat seinen Beschluß bekanntgegeben
- b) Eine neue wissenschaftliche Entdeckung

Welchen der beiden Artikel lesen Sie zuerst?



— Durchschnittsprofil von 1000 Medizinstudenten, Ende des Studiums (USA, Quotenstichprobe).
 - - - Durchschnittsprofil von 68 graduierten Pädagogen der Universität von Harvard*.

18. In einem Wartezimmer liegen zwei Zeitschriften zu Ihrer Verfügung. Welche würden Sie wählen?

- a) Das Zeitalter der Naturwissenschaft
- b) Kunst und Dekoration

II.1. Sind Sie der Meinung, daß es das eigentliche Bestreben einer Regierung sein sollte, dafür zu sorgen:

- a) daß die Armen, Kranken und Alten mehr unterstützt werden;
- b) daß das Handels- und Industriegewesen weiter ausgebaut wird;
- c) daß hohe moralische Grundsätze in die Politik und Diplomatie eingeführt werden;
- d) sich eine Vormachtstellung innerhalb der Gemeinschaft oder Nationen zu erkämpfen?

14. Wünschen Sie sich, daß Ihr zukünftiger Ehemann:

- a) im Berufsleben erfolgreich ist und von den anderen bewundert wird;
- b) den andern gerne hilft;
- c) eine hohe Lebensauffassung vertritt;
- d) künstlerische Fähigkeiten besitzt?

* Siehe Testhandanweisung, S. 14.

Faktorenanalysen haben aufgezeigt, daß die Unterscheidung zwischen den sechs dominierenden Wertbereichen mehr in der Theorie als in der Praxis besteht.

5. Die Identifizierung der Interessen

Die Identifizierung und das Messen von Interessen sind für die praktische wie für die experimentelle Pädagogik von einstimmig anerkannter Bedeutung.

Aber die Auswertung der eigentlichen Interessen ist schwierig, denn das Übereinstimmungsverhältnis zwischen den Tendenzen des Prüflings und den Gegenständen erklärt sich letztlich nur aufgrund einer Erforschung seiner ganzen Person: seiner körperlichen Konstitution, seiner Intelligenz und seiner Persönlichkeit.

Wenn darüber hinaus das Interesse theoretisch eine breitere Grundlage besitzt, wenn es funktioneller erscheint als die Einstellung — auf affektivem Niveau enger lokalisiert — ist es dennoch nicht leicht, eine klare Grenze zwischen dem Interesse und der Einstellung zu ziehen. In dieser Hinsicht ist es symptomatisch, daß L. Delys die in dem Fragebogen zur Messung der Berufsinteressen von Strong⁹⁷ gegebenen Antworten als Einstellungs-termini interpretieren konnte.

Die Relation zwischen den Prüfungen von Strong, Kuder und dem MMPI⁹⁸ einerseits und den Wertuntersuchungen andererseits wurde übrigens von verschiedenen Autoren beleuchtet (insbesondere von C. Bursch, J. Darley und G. Garman).

Wie dem auch sei, der Pädagoge weiß aus Erfahrung, daß das Interesse — ob spontan oder provoziert — einen außerordentlichen Einfluß auf den Lernvorgang ausübt, und er muß Mittel und Möglichkeiten suchen — und sei es auf empirischem Wege —, diese starke Antriebskraft kennenzulernen.

97 L. Delys, Une application du questionnaire d'intérêts vocationnels de E. K. Strong Jr. avec interprétation des résultats en termes d'attitudes, Brüssel, C. N. R. P. S., 1957.

98 Inventaire multphasique de Personnalité du Minnesota (MMPI), von J. McKinley, Psychological Corporation — Centre de Psychologie Appliquée.

Die zur Diskussion stehenden Prüfungen gehen von dieser pragmatischen Einstellung aus. Die Indikationen, die sie liefern, sollen mit jener Umsicht interpretiert werden, die für die gesamte Persönlichkeitsforschung gilt.

Wenn der Proband gut mitgearbeitet hat, zeigen die Interessentests — vorbehaltlich einer weitmöglichst guten Validierung — Tendenzen an, die der Pädagoge für den praktischen Gebrauch ausnutzen wird.

Ein Interesse untersuchen bedeutet, sich bemühen, sowohl eine weitgehend klare Vorstellung von seinem Charakter zu gewinnen, als auch etwa die fünf folgenden Fragen zu beantworten:

1. Handelt es sich um ein flüchtiges bzw. vorübergehendes oder um ein dauerhaftes Interesse?

Die Forschungsergebnisse bezüglich des Alters, in dem das Interesse erstmalig auftrat, und der Dauerhaftigkeit sind zugegebenermaßen wenig übereinstimmend.

Nach Fryer wäre die Chance, daß die von einem Jugendlichen geäußerten Interessen länger als ein Jahr bestehen, nur 50 : 100⁹⁹. Murphy jedoch hat gezeigt¹⁰⁰, daß dem Wandel unterliegende Interessen sich dennoch stets auf das gleiche Wertesystem, auf die gleiche Grundeinstellung beziehen können. Ein Geltungsbedürfnis kann die einzige Erklärung sein für das Interesse eines Probanden an einem Verein, in dem er eine Rolle spielt, an einer politischen Partei oder einem Beruf, die insgesamt gesehen Ziele verfolgen oder Ideologien vertreten, die im Widerspruch zueinander stehen.

2. Besteht ein ernsthaftes oder ein oberflächliches Interesse?

3. Ist es weit- oder enggespannt?

4. Ist es egoistisch oder altruistisch? Auf den einzelnen oder auf die Gruppe zentriert?

5. Sind die Interessen des Probanden insgesamt gesehen gut ausgewogen?

Darüber hinaus unterscheidet D.E. Super nach den zur Identifizierung der Interessen eingesetzten Mitteln¹⁰¹:

1. Die offenkundigen Interessen: Solche, die man beim Beobachten des Probanden während der Arbeit oder während seiner Freizeit feststellt (direkte Beobachtung der Verhaltensweisen).

99 D. Fryer, The Measurement of Interests, London, Harrap, 1931.

100 G. Murphy, Personality, A Biosocial Approach to Origins and Structures, New York, Harper, 1947.

101 D.E. Super, Appraising Vocational Fitness by Means of Psychological Tests, New York, Harper, 1949, S. 376–379.

2. Die geäußerten Interessen: Solche, denen der Proband den Vorzug gibt, wenn er die freie Wahl hat (Fragebogen, Interviews).

3. Die in einem Verzeichnis festgehaltenen Interessen: Solche, die der Proband angibt, wenn man ihm eine Aufstellung über Betätigungsmöglichkeiten vorlegt (Inventarien).

Vielfach gibt es keine völlige Identität zwischen diesen drei Typen, und die Art ihrer Relationen ist bisher noch unzureichend definiert.

Schließlich möchten wir noch bemerken, daß man sich bis heute scheinbar mehr um die Diagnostizierung der spezifisch beruflichen Interessen bemüht hat, als darum, die allgemeineren Motivationen zu identifizieren, und man kommt immer mehr zu der Überzeugung, daß diese für die einzelnen Menschen die wesentlicheren sind. Jedenfalls scheint es bedauerlich, daß die Interessen bei Kindern und Jugendlichen nicht systematischer untersucht werden.

Zunächst geben wir drei Untersuchungsbeispiele zur Identifizierung der allgemeinen Interessen; sodann betrachten wir kurz die beruflichen Interessen.

A. Untersuchung der allgemeinen Interessen

1. *Interest Finder* von A. Jersild und R. Tasch¹⁰²

Als gut ausgearbeitetes Untersuchungsverfahren, anwendbar bei 6- bis 18jährigen Probanden, besteht der *Interest Finder* aus einer Reihe von Sätzen, die zu vervollständigen sind.

- Meine drei liebsten Wünsche sind ...
- Ich würde in der Schule gern etwas mehr erfahren über ...
- Ich habe keine Lust ... zu lernen.
- Am liebsten habe ich in der Schule ...
- Am wenigsten mag ich in der Schule ...
- Neben der Schule habe ich am liebsten ...
- Wenn ich groß bin, werde ich ... sein.
- Wenn ich groß bin, werde ich ... machen.
- Im Laufe der letzten Woche war von dem, was wir in der Schule gemacht haben, am interessantesten ...
- Einer der schönsten Tage meines bisherigen Lebens war ...

Die bei einem solchen Fragebogen gegebenen Antworten können sowohl hinsichtlich der Interessen als auch der Persönlichkeit aufschlußreich sein.

102 A. Jersild u. R. Tasch, *Children's Interests and what they suggest for Education*, New York, Teachers College, Columbia Univ., 1949.

Ja selbst wenn die Erforschung der Interessen es dem Pädagogen auch erlaubt, seine Schüler besser kennen und verstehen zu lernen, bedeutet dies doch keineswegs, daß nur die augenblicklichen Interessen der Klasse — sie sind überdies so verschieden, wenn nicht sogar widersprechend — den Lehrer dazu verpflichten, den Unterricht nach ihnen auszurichten und zu gestalten.

Es gehört mit zu den Aufgaben des Pädagogen, eine Kompromißlösung zu finden, die die spontane Begeisterung des Kindes mit den späteren Anforderungen, die das Leben stellt, in Einklang bringt.

Eine umfangreiche von Jersild und Tasch¹⁰³ durchgeführte Umfrage kommt zu folgenden Schlußfolgerungen:

1. Die Interessen der Kinder variieren erheblich von Schule zu Schule und Klasse zu Klasse. Dies bestätigt die Theorie, daß beim Aneignen von Interessen das Milieu eine wichtige Rolle spielt.
2. In dem Maße, wie die Kinder älter werden, läßt auch ihr Interesse an der Schule nach.
3. Die geäußerten Interessen entsprechen kaum den Grundproblemen, die die Kinder zu lösen haben; ihre tiefer gegründeten Belange werden nicht identifiziert, was den pädagogischen Wert der erhaltenen Resultate erheblich herabsetzt.
4. Das Fehlen von Interessen bei den Eltern kann ebenfalls Interessenlosigkeit beim Kind zur Folge haben (siehe 1).
5. Die identifizierten Interessen können nur schwerlich zur Erarbeitung eines Lehrprogrammes herangezogen werden.

M. Almy¹⁰⁴ bemerkt überdies hinsichtlich des *Interest Finder*, daß die Reaktionen des Kindes durch die nach seiner Meinung vom Lehrer erwartete Antwort stark beeinflußt werden. Ein Vergleich der Ergebnisse mit anderen Informationsquellen ist daher notwendig.

Es ist weiterhin bekannt, daß die Antworten, die die Wünsche betreffen, nichts über ihre Intensität aussagen; man weiß jedoch, daß sie von Fall zu Fall erheblich variiert.

Schließlich ist es mitunter schwierig, festzustellen, inwieweit die Äußerungen rein stereotype Antworten sind. Almy zitiert das Beispiel eines kleinen Kindes, welches seine drei Wünsche so ausdrückt: „dem lieben Gott wohlgefallen — immer brav sein — in den Himmel kommen“.

2. *Questionnaire d'intérêts* für Jungen von 10—12 Jahren, BINOP¹⁰⁵

Dieser Fragebogen (72 Items) wurde an Klassen im Raum von Paris erprobt. Die interne Untersuchung hat zu homogenen Fragengruppen geführt. Eichungen in fünf normalisierten Kategorien.

103 Nach R. Travers, *Educational Measurement*, New York, Macmillan, 1955, S. 290 bis 291.

104 M. Almy, *Ways of Studying Children*, New York, Teachers College, Columbia Univ., 1966, S. 123—124.

105 Siehe S. Larcebeau, *Etude d'un questionnaire d'intérêts pour garçons de 10 à 12 ans* (BINOP, 1958, Nr. 4, S. 232—254).

5. Untersuchung anhand von Titeln fingierter Bücher, von R. Thorndike¹⁰⁶

Ursprünglich dazu bestimmt, die literarischen Interessen beim Kind zu untersuchen, wurde diese Prüfung späterhin für eine viel weiter gefasste Erforschung der Interessen verwendet.

Die Aufstellung von Thorndike enthält eine Reihe Titel von fingierten Werken, denen jeweils eine frei erfundene Zusammenfassung folgt.

Das Kind muß angeben, welche Bücher es gerne lesen würde.

Der Test erstreckt sich auf die Interessen, für die das Kind im allgemeinen abgeschlossen ist: Betätigung draußen im Freien, kleine Abenteuer usw., und die Technik läßt sich leicht nachahmen.

Um eventuelle Unaufrichtigkeiten aufzudecken, enthält die Liste sechs Titel, die normalerweise für 10- bis 15jährige Probanden uninteressant sind: „Berühmte Predigten von Hohen Geistlichen“. Man schaltet die Antworten aus, die vier oder fünf derartiger Titel beinhalten.

B. Untersuchung der Berufsinteressen¹⁰⁷

Es erscheint durchaus wahrscheinlich, daß im Laufe der nächsten Jahre der Begriff selbst der Berufsinteressen und demzufolge auch die zu ihrer Auswertung eingesetzten Meßinstrumente von Grund auf revidiert werden müssen. Auf den Gebieten der Landwirtschaft und der Industrie hat der technologische Fortschritt eine rasche Entwicklung der Berufe zur Folge; einige bilden sich neu heraus, andere sterben aus. Neben der Vielfalt neuer Produkte macht sich ein Trend zur Vereinheitlichung der Produktionstechniken bemerkbar. Die Berufsausbildung stellt daher in zunehmendem Maße das Erlernen polyvalenter Methoden und Techniken in den Vordergrund. Donald Super vertritt die Meinung, daß man zu einer sinnvollen Hinlenkung des Jugendlichen auf eine ihm angemessene Berufslaufbahn nicht mehr eine bestimmte, enggefaßte Leistungskapazität zu identifizieren hat — was übrigens stets recht illusorisch und vielfach unheilvoll war —, sondern daß es vielmehr gilt, die allgemeine Tendenz seiner Interessen aufzuklären (wissenschaftlich, altruistisch, literarisch, systematisch usw.). Die Einstellungsskalen von Edwards und Wilson, die wir im vorhergehenden Kapitel be-

106 R.L. Thorndike, A Comparative Study of Children's Interests, New York, Teachers College, Columbia Univ., 1941.

Anm. d. dtsh. Bearbeiters: Vgl. Bücherkatalogtest von Tramer, Zürich, 1953.

107 Über das Wesen und die Untersuchung der Berufsinteressen siehe: R. Pasquasy, Les intérêts professionnels et leur mesure. Mont-sur-Marchienne. Verlag: Application des Techniques Modernes 1961 (Veröffentlichungen über moderne Anwendungstechniken). Anm. d. dtsh. Bearbeiters: Vgl. auch: BIT (Berufsinteressentest), M. Irlé, Göttingen, 1955.

handelt haben, scheinen gegenwärtig dieser neuen Orientierung am ehesten zu entsprechen.

Es folgen einige Testbeispiele:

1. Fragebogen zur Messung der Berufsinteressen (*Vocational Interest Blanks, VIB*), von E.K. Strong jr., französische Bearbeitung von L. Delys (Clerebaut, Brüssel, 1954)

Form M für männliche Probanden umfaßt Skalen für 49 Berufe, 6 Beschäftigungsgruppen und 4 spezielle Variablen (Intensitätsgrad der Interessen, Beschäftigungsniveau, Spezialisierungsniveau, maskulin-feminin).

Form W für weibliche Probanden umfaßt 30 Berufe. Dieser Berufsinteressentest ist in den USA vermutlich am besten erprobt, und sein Voraussagewert ist sehr hoch.

Auszüge aus Form M

I. Geben Sie für die folgenden Berufe an, ob Sie sie „mögen“ — „nicht mögen“ — sie Ihnen „gleichgültig“ sind:

- 1) Schauspieler
 - 2) Werbefachmann
 - 3) Architekt
- usw.

II. Ebenso für folgende Lehrstoffe:

101. Algebra
 102. Landwirtschaft
 103. Arithmetik
- usw.

III. Desgleichen für die folgenden Tätigkeiten:

187. Vergaser regulieren
 188. Elektrischen Kurzschluß reparieren
- usw.

IV. Geben Sie in der folgenden Liste in Kolonne 1 jeweils durch ein (x) die drei Tätigkeiten an, die Sie am liebsten ausüben möchten, und in Kolonne 3 diejenigen, die Ihnen am wenigsten zusagen würden.

(1) (2) (3)

281. () () () Zur Inbetriebsetzung einer neuen Maschine eine Theorie aufstellen.
 282. () () () Sich der Maschine bedienen.
 283. () () () Einen Verbesserungsvorschlag zu der Maschine machen.
 284. () () () Die Betriebskosten der Maschine errechnen.
 285. () () () Die Fabrikation der Maschine leiten.
- Usw.

VIII. Beantworten Sie folgende Fragen:

- | | |
|--|---------------|
| 362. Ich arbeite regelmäßig und nicht intuitiv oder stoßweise, | ja — ? — nein |
| 363. Ich schließe leicht Freundschaften, | ja — ? — nein |

364. Ich bringe andere leicht dazu, nach meinem Willen zu handeln.
usw.

ja — ? — nein

Ohne die maschinelle Unterstützung ist die Auswertung dieser Prüfungen sehr zeitraubend. Der Testautor hat vor allem die freien Berufe und die Geschäftswelt im Auge gehabt; daher berücksichtigt der Fragebogen vornehmlich diese Berufe.

2. *Contrôle des intérêts professionnels* von R. Derivière (Editest, 3. Aufl. 1961)

Diese in Belgien gern angewandte Prüfung wertet die Erfahrungen von Desautels, Thurstone, Kuder und Strong aus, auf die sich die Forschungen des Autors offenbar gründen.

Seinen mehr spezifischen Orientierungscharakter entnimmt das Verfahren von Derivière der Methode von Strong, während seine Erschließung der Interessen auf breiter Ebene von der Anregung von Kuder herrührt.

Form der Prüfung für jüngere Schüler (Ende der Volksschulzeit)

Der Autor, der sich über die Fluktuation der Interessen in dieser Altersgruppe im klaren ist, beschränkt sich darauf, die prinzipiellen Neigungen zu identifizieren (ist Interesse an höherer Schulbildung vorhanden oder nicht), sowie der drei generellen Tendenzen zu diagnostizieren:

- formell (vor allem verbal);
- technisch (Umwandlung der Materie);
- praktisch.

Itembeispiel: (forcierte Wahl; 00 gibt die grundsätzliche Ablehnung an und 0 die Ablehnung an zweiter Stelle).

- A. 1. Rechtsanwalt, Lehrer, Journalist.
- 2. Übersetzer, Ansager (in) beim Radio, Bibliothekar(in)
- 3. Museumsaufseher, Reiseleiter, Gerichtsvollzieher, Fremdenführer, Telefonist (in)

Form des Tests für Junioren (Ende der höheren Schule — Unterstufe)

Form des Tests für Senioren (Ende der höheren Schule — Oberstufe)

Hier versucht man, die Interessen nach neun verschiedenen Gebieten aufzuteilen:

- a) Sprachen (Sprachkenntnisse, mündliche und schriftliche Ausdrucksweise, Interesse für Geisteswissenschaften);
- b) Naturwissenschaften (wissenschaftliche Forschung und angewandte Naturwissenschaften);
- c) Praktische Berufe (körperliche Tätigkeit, Befehlsgewalt, Ausführung);
- d) Mathematik;
- e) Soziale Berufe;
- f) Technische Berufe (Umwandlung von Stoffen);
- g) Graphik;
- h) Verwaltung;
- i) Handel;

Zu jeder vorgeschlagenen Tätigkeit äußert sich der Proband anhand einer in fünf Stufen unterteilten Skala (Ablehnung, Gleichgültigkeit, Akzeptierung, Interesse, Bevorzugung). Das Antwortblatt dient gleichzeitig der Konstruktion eines Profils (halbe Standardabweichungen).

Itembeispiele (Junioren)

I. Welches sind Ihre liebsten Unterrichtsstunden ?

- 1 A. Französischer Aufsatz;
- 2 A. Alte Sprachen;

usw.

II. Welche Art von Freizeitbeschäftigung bevorzugen Sie ?

- 14 A. Lesen von Romanen, Reiseberichten oder Abenteuer Geschichten;
- 15 B. Chemische oder physikalische Experimente;
- 16 C. Rote-Kreuz-Veranstaltungen;

usw.

III. Welche Veranstaltungen würden Sie wählen ?

- 29 G. Dokumentarfilme über Kunststädte;
- 23 A. Theaterstücke.

IV. An welchen Wettbewerben würden Sie gerne teilnehmen ?

V. Würden Sie gerne an untenstehenden Untersuchungen teilnehmen ?

VI. Welche Aufgaben würden Sie bei der Organisation einer Festlichkeit am liebsten übernehmen ?

VII. Was machen Sie während Ihres Urlaubs am liebsten ?

VIII. Welchen Dienst würden Sie sich beim Militär wählen ?

IX. Welchem der Entwicklungsländer würden Sie sich zuwenden ?

X. Welchen Beruf würden Sie von den im folgenden angeführten wählen ?

3. *Test projectif d'intérêts vocationnels* von F. Bemelmans (O. P. Lüttich, 3. Aufl.)¹⁰⁸

Bei diesem Test werden Photographien von Personen bei der Ausübung ihres Berufes vorgelegt, die in ihrer Wiedergabe mehr oder minder klar sind. Die Projektion wird somit in Richtung auf den beruflichen Aspekt gelenkt.

Darunter sind einige klar erkennbare Photos (Apotheker in seinem Labor); andere stellen lediglich das Portrait von Persönlichkeiten dar; wieder andere weisen eine vollkommen verwirrte Darstellung auf, und zwar derart, daß man unmöglich irgendeine Tätigkeit daraus erkennen kann.

Zunächst wurde eine Einzelprüfung vorgenommen. Ausgehend von 60 Photos (30 Männer und 30 Frauen) stellte der Autor die folgenden Fragen:

- 1. Welche Tätigkeit übt diese Person aus ?
- 2. Welche Materie bearbeitet sie ? (Holz, Eisen, eine geistige Materie usw.)
- 3. Welchen Beruf übt sie aus ?
- 4. Welchen Beruf oder welche Stellung kann sie sich für später erhoffen ?

¹⁰⁸ Siehe auch: M. Viglietti u. P. Garcia Castelblanco, Le Test projectif d'intérêts professionnels de F. Bemelmans (Bulletin d'Orientation Scolaire et Professionnelle, 7. Jahrg. Nr. 3, 1958).

5. Was fehlt ihr, damit sie vollkommen glücklich und mit ihrem Schicksal zufrieden ist?

6. Gefällt Ihnen diese Idee?

usw.

Die Inhaltsanalyse wurde nach einem klaren Plan durchgeführt (siehe Handbuch, S. 5–9).

Kollektivform: Sie umfaßt 40 Photos. „Für jedes Bild hat man von den während der ersten Untersuchung der jeweiligen Person zugeordneten Aktivitäten eine bestimmte Itemanzahl, entsprechend derjenigen des Kuder'schen Tests, gewählt. Dies hat es erlaubt, aufgrund der Faktorenanalyse, jedem Item eine mutmaßliche Bedeutung beizumessen“ (Handbuch S. 13).

Jedem Photo ist eine bestimmte Anzahl von Vorschlägen zugeordnet, von denen der Proband jeweils einen wählen muß. Er soll weiterhin angeben, ob die betreffende Tätigkeit ihm sehr zusagt, ihm gleichgültig ist oder ihm mißfällt.

Itembeispiele:

1. (Das Photo stellt ziemlich klar erkennbar einen Mann dar, der in einem chemischen Labor arbeitet.)

1. D. Macht chemische Analysen.

E. Verkauft chemische Produkte.

D. Chemielehrer in einem Labor.

19. (Schattenbild einer aufrecht stehenden Person auf verschwommenem Hintergrund.)

B. Demontiert die Turbine eines Schiffes.

F. Meißelt die Vorderfront einer Statue.

J. Begibt sich zu einem Unfallort, um den Opfern zu helfen.

A. Lichtet und beschnidet die jungen Bäume im Wald.

Die Buchstaben geben die folgenden Tendenzen an:

A. Im Freien.

B. Manuelle Tätigkeit – technisch, mechanisch usw.

C. Rechnungswesen – Mathematik.

D. Naturwissenschaften.

E. Überredung von anderen.

F. Kunst, Zeichnen.

G. Literatur.

H. Musik.

J. Soziale Tendenz.

K. Verwaltung.

Durch Addieren der Punktwerte für jede einzelne Tendenz erhält man eine Punktzahlreihe, die man in ein Profil überträgt.

Zur beruflichen Orientierung genügt es, von Dominanten auszugehen, die im Schaubild hervortreten, und sich eine allgemeine Klassifizierung zu vergegenwärtigen, die nach Kuder und den Arbeiten der „Commission technique générale de classification professionnelle“ (Allgem. Techn. Klassifizierungskomitee) des Belgischen Arbeitsministeriums aufgestellt wurde.

Dieser Test ist in seiner originellen und ansprechenden Ausführung noch nicht hinreichend validiert. Es bestehen auch gewisse Unklarheiten über die Konzeption des

Instruments selbst: Inwieweit bleibt die aufgrund der Wahlen gelenkte Projektion eine echte Projektion? Man wird sich insbesondere daran erinnern, daß der Systematisierungsversuch, den Harrower für den Rorschach-Test vorschlägt, stark angefochten wird.

6. Die Bewertung der Arbeitsmethode der Schüler

I. Die Arbeitsmethode

Bei der Untersuchung der Arbeitsmethode sind zwei Vormerkungen angezeigt: Eine Methode, so gut sie auch sein mag, hat keinerlei Bedeutung, es sei denn, daß bei dem Schüler echte Begabungen zugrunde liegen. Sie läßt sich nicht auf einige allgemeingültige Regel reduzieren. Tatsächlich sind zweifellos Intelligenz und Persönlichkeit gänzlich in die Arbeitsweise mit einbezogen.

Weiterhin bestimmt die Art der auszuführenden Arbeit auch die Methode, nach der man lernt: In einer Schule, die den freien Vortrag verlangt, werden die Schüler unter diesem Aspekt lernen.

Der intelligente und leicht anpassungsfähige Schüler wird von sich aus seine Arbeitsmethode kritisch überprüfen und sich die Mittel und Wege vorbehalten, die ihn in der jeweiligen Situation zu den besten Ergebnissen führen. Dennoch, auf sich selbst angewiesen, erfolgt das Vorgehen der Schüler vor allem aufgrund von Versuch und Irrtum, so daß das Finden der richtigen Lösung zeitraubend und kostspielig sein kann.

Ohne daher zu übersehen, daß die Arbeitsmethode kein isolierter und unabhängiger Vorgang¹⁰⁹ ist, kann die Schule hier eine Doppelrolle spielen: Den Begabtesten kann sie helfen, so rasch wie möglich die für sie geeignetsten Methoden zu finden und den weniger Begabten unmittelbar wertvolle Ratschläge zu sinnvoller Zeiteinteilung, zur Verwendung von Lern- und Arbeitsmitteln sowie hinsichtlich der Arbeitsorganisation zu geben.

¹⁰⁹ Man wird übrigens feststellen, daß sich bei den von uns dargelegten Tests die Fragen teilweise auf die Probleme der Persönlichkeit erstrecken und nur teilweise auf die eigentlichen Arbeitsmethoden.

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 zu viel Zeit und für andere wiederum zu wenig. Ich lerne am liebsten mit einem anderen zusammen.	6 5	5 -4	-11 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. Ich kann nur lernen, wenn ich dabei rauche.	-13 7	0 -5	7 -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 Examens 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 Verwendung, 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.

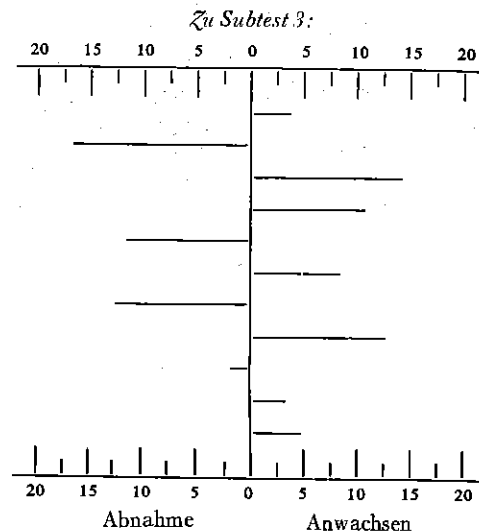
(Siehe Abb. S. 176.)

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?

Österreich
 Birma
 Finnland
 Frankreich
 Deutsche Bundesrepublik
 Indien
 Japan
 Holland
 Schweiz
 Großbritannien
 U.S.A.



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

5. Fähigkeit im Aufnehmen von Notizen (Erkennen des Wesentlichen)
 Anhand von kurzen Texten, bei denen die jeweils zugrunde liegenden Ideen numeriert 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 begrifflich 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 Studienzzeit 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 vortragen 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. 347ff).

126 Siehe S. 83ff.

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 Ann. d. dtsh. Bearbeiters: Durch die Unterrichtsmitschau (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 Beobachtungsminuten

Gesamt		Kontrollierende Handlungen	Anordnungen	Helfende Handlungen	Entwicklungen	Persönliche Beantwortungen	Handlungen der Affektivität	
							+	-
Gesamtzahl der Handlungen	26385	12402	738	1918	4131	1289	3243	2664
%	100	47	3	7	16	5	12	10
Durchschnittszahl pro Erzieher	753,9	354,3	21,1	54,8	118	36,8	92,7	76,1

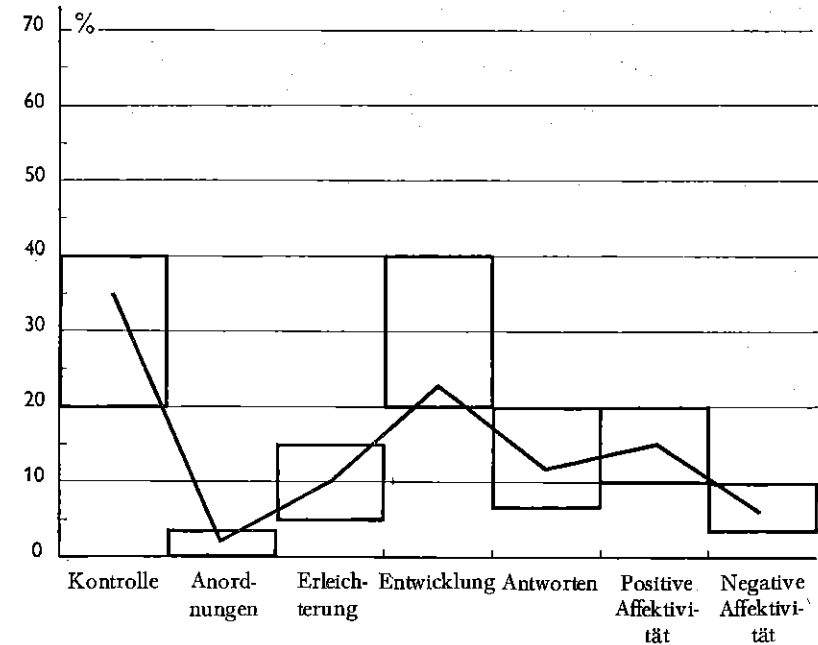
Nach einigen ergänzenden Untersuchungen schlagen M. Hughes u. a. als Modell eines guten Unterrichts vor:

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 beurteilen, 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 anfechtbar; es beruht – wie gesagt – auf der Beobachtung von Erziehern, die von ihren Direktoren 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 Vorzug, einen präzisen Vorschlag zu bringen, und es ist wegberetend 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 Beobachtungen 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 Verhaltensweise 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 Education, 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 Güntilinge 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 etlichen 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, Rektoren 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

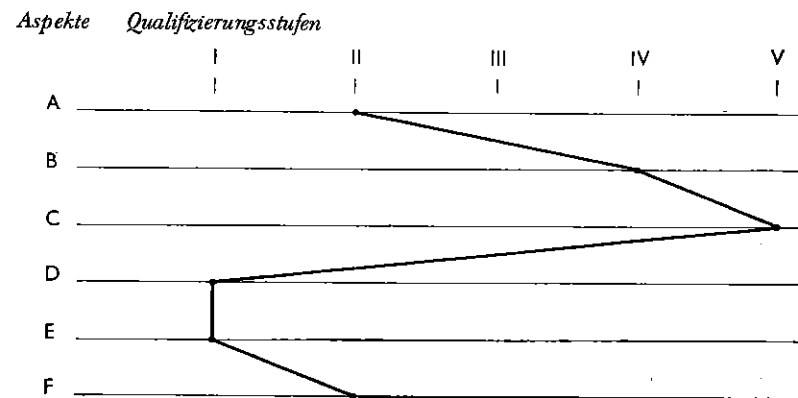
Diese ausführliche Anleitung ist für die Beurteilung der Nebraska-Schulen und ihres Lehrpersonals entworfen und ist tatsächlich nur für diese zu gebrauchen. Dennoch bringen wir eine Darstellung dieses Untersuchungsverfahrens, und zwar in ziemlich eingehender Form, weil es eine nützliche Anregung vermitteln 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 Erfahrungen werden für das gegenwärtige Leben des Kindes in einer Demokratie zu entwickeln. Diese Erfahrungspunkte sind im Hinblick auf die Anforderungen, die Fähigkeiten und Interessen der Gruppemitglieder, die Fähigkeiten und Interessen der Einzelpersonen gewählt.	Das Lehrprogramm enthält Erfahrungsaspekte, die dazu beitragen werden, Lerngewohnheiten, Erkenntnisse, nützliche Einstellungen und Techniken für das gegenwärtige Leben des Kindes in einer Demokratie zu entwickeln. Diese Erfahrungspunkte sind im Hinblick auf die Anforderungen, die Fähigkeiten und Interessen der Gruppemitglieder, die Fähigkeiten und Interessen der Einzelpersonen gewählt.

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>
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.
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: Leistungstests und Intelligenztests zu Beginn des Schuljahres. Die Testdienleistungen der Anpassung des Unterrichtsprogramms an jeden Schüler.	Wie Stufe III, ferner: Interessentests 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 im theoretischen und praktischen Bereich, sowie die Fortschritte im Bereich des mitemenschlichen Verhaltens selbst wahrzunehmen.
<i>N. Zensuren</i>				

Naturwissenschaften (7 Aspekte)

D.

Stufe I	Stufe II	Stufe III	Stufe IV	Stufe V
Wenig oder keine Experimente bzw. Untersuchungsvorhaben. 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 selbstetige 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 Tatbestände 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)

I.

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 angestrebten 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 provozierte 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.

<p><i>Stufe I</i> Gelegentlich werden einige Lieder gesungen.</p>	<p><i>Stufe II</i> Man schenkt der Musik gewisse Beachtung, jedoch ist im Stundenplan kein spezieller Musikunterricht eingeplant.</p>	<p><i>Stufe III</i> 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.</p>	<p><i>Stufe IV</i> Wie III, ferner: Individualisierung der Aktivitäten. Einige Erfahrungen erlauben es, das musikalische Verständnis sowie schöpferische Betätigung zu entwickeln.</p>	<p><i>Stufe V</i> Wie IV. Der Entwicklung des musikalischen Verständnisses und der schöpferischen Tätigkeit wird größere Bedeutung beigemessen.</p>
---	---	--	--	---

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

B.

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

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

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

B.

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

Schulgebäude 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, Begävning 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 Hochschulbildung 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. Dienstpersional	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 einseitig 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>	
Erbter 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	Ein- kom- men	Bil- dung	Pre- stige (NORC)	Sozio- ökon- omischer 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	Ein- kom- men	Bil- dung	Pre- stige (NORC)	Sozio- ökon- omischer 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 Schwierigkeitsgrads 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. VI 5–6 u. VI 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.

Formel¹³: Durchschnittliche Länge der Sätze mal 1,015 = x
Zahl der Silben für 100 Wörter mal 0,846 = y

$$\frac{x + y}{x + y}$$

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

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). Ebenso wenig 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)
	...120...			
6 Jahre	...115...			
7 Jahre	...110...			
8 Jahre	...105...			
9 Jahre	...100...			
10 Jahre	... 90 ...	sehr einfach	Komikhefte	93%
11 Jahre	... 80 ...	leicht	Kioskliteratur ^a	91%
12 Jahre	... 70 ...	ziemlich leicht	Zukunftsromane ^b	88%
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 Gradeneinstellung 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 Mittellungen 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 Fleisch 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 Fleisch 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)	()
Schuljahrgang:	(23)	()
Unterrichtszweig: IG – LN – LM – NSNA – NSW	(24–26)	()
Testdatum: Monat	(27–28)	
Jahr (die 2 letzten Ziffern der Jahreszahl)	(29–30)	()
Testergebnisse		()
– Primary Mental Abilities (PMA) von Thurstone		
Im Prozentrang: Wortverständnis (V)	(31)	()
Räumlicher Faktor (R)	(32)	()
Numerischer Faktor (N)	(33)	()
Sprachflüssigkeit (W)	(34)	()
Gesamtpunktwert (T)	(35)	()
Intelligenzquotient (IQ)	(36–38)	()
– Skala von Wechsler		
Intelligenzquotient, Verbalteil (VIQ)	(39–41)	()
Intelligenzquotient, Handlungsteil (HIQ)	(42–44)	()
Intelligenzquotient: Gesamt (GIQ)	(45–47)	()
– Fragebogen zur Messung der Anpassung von Bell (Rohpunktwerte)		
Familiäre Anpassung (FA)	(48–49)	()
Anpassung nach gesundheitlichen Gesichtspunkten (GA)	(50–51)	()
Soziale Anpassung (SA)	(52–53)	()
Emotionale Anpassung (EA)	(54–55)	()

- c) Karten, die also mit entsprechendem Vordruck ausgestattet sind, erlauben es dem Befrager oder Beobachter, die während eines strukturierter oder halbstrukturierter 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.

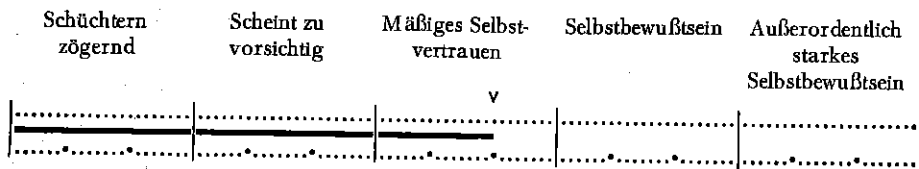
1. a b c d e

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?



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

GESCHLECHT
ALTER

DATE

NAME
BÜRO ODER FIRMA

RATING

Punktzahl

PART I

EXAMPLES

W	a	b	c	d	e
X	a	b	c	d	e
Y	a	b	c	d	e
Z	a	b	c	d	e

1	a	b	c	d	e
2	a	b	c	d	e
3	a	b	c	d	e
4	a	b	c	d	e
5	a	b	c	d	e
6	a	b	c	d	e
7	a	b	c	d	e
8	a	b	c	d	e
9	a	b	c	d	e

10	a	b	c	d	e
11	a	b	c	d	e
12	a	b	c	d	e
13	a	b	c	d	e
14	a	b	c	d	e
15	a	b	c	d	e
16	a	b	c	d	e
17	a	b	c	d	e
18	a	b	c	d	e
19	a	b	c	d	e
20	a	b	c	d	e

21	a	b	c	d	e
22	a	b	c	d	e
23	a	b	c	d	e
24	a	b	c	d	e
25	a	b	c	d	e
26	a	b	c	d	e

PART II

EXAMPLES

W	1	2	3	4	5
X	1	2	3	4	5
Y	1	2	3	4	5
Z	1	2	3	4	5

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

4	1	2	3	4	5
5	1	2	3	4	5
6	1	2	3	4	5
7	1	2	3	4	5
8	1	2	3	4	5
9	1	2	3	4	5
10	1	2	3	4	5
11	1	2	3	4	5
12	1	2	3	4	5
13	1	2	3	4	5
14	1	2	3	4	5

15	1	2	3	4	5
16	1	2	3	4	5
17	1	2	3	4	5
18	1	2	3	4	5
19	1	2	3	4	5
20	1	2	3	4	5

21	1	2	3	4	5
22	1	2	3	4	5
23	1	2	3	4	5
24	1	2	3	4	5
25	1	2	3	4	5
26	1	2	3	4	5
27	1	2	3	4	5
28	1	2	3	4	5
29	1	2	3	4	5

30	1	2	3	4	5
31	1	2	3	4	5
32	1	2	3	4	5
33	1	2	3	4	5
34	1	2	3	4	5
35	1	2	3	4	5
36	1	2	3	4	5
37	1	2	3	4	5
38	1	2	3	4	5
39	1	2	3	4	5
40	1	2	3	4	5

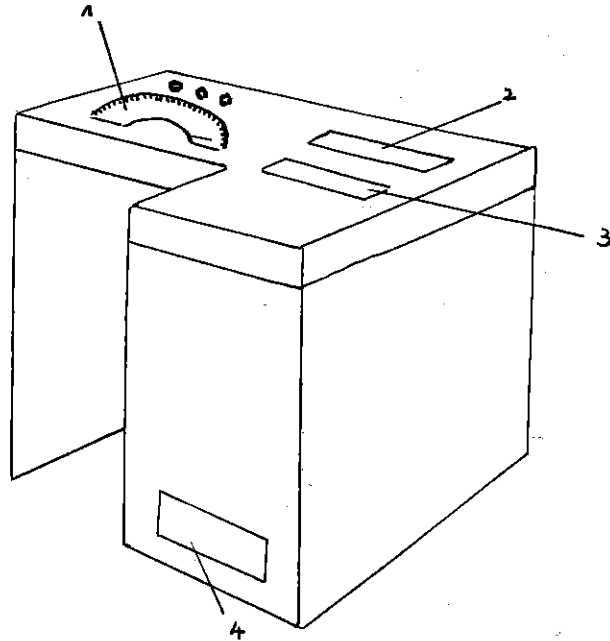
PART III

EXAMPLES

X	a	b	c	d	e
Y	a	b	c	d	e

1	a	b	c	d	e
2	a	b	c	d	e
3	a	b	c	d	e
4	a	b	c	d	e
5	a	b	c	d	e
6	a	b	c	d	e
7	a	b	c	d	e
8	a	b	c	d	e
9	a	b	c	d	e
10	a	b	c	d	e
11	a	b	c	d	e

12	a	b	c	d	e
13	a	b	c	d	e
14	a	b	c	d	e
15	a	b	c	d	e
16	a	b	c	d	e
17	a	b	c	d	e
18	a	b	c	d	e
19	a	b	c	d	e
20	a	b	c	d	e



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 Anzeigernadel 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 1200 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 IARC 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 Relaissschrank „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 1357 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 Abschreibebungen auf die Orthographie einer Klasse gemessen werden. Bei diesem Schema umfaßt das Experiment drei Entwicklungsstufen:

1. Vorausgegangene Durchdührung eines Tests zur Messung der orthographischen Kenntnisse;
2. Untersuchung;
3. Durchführung eines orthographischen Tests — Paralleform 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. Unterricht	Kontrolltest
	Gruppe 6	—	1/2 herk. Unterricht	1/2 masch. Unterricht	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 Zufallstablen), 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 U L E 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. Stüllwold: 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	B ₁	B ₂	B ₁	B ₂	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 Süllwold, 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 verzerrten 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 Lösen, 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 Stratifizierung 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 Unterpulation (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

Weiterhin 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 darstellten.

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 Verfahrensweisen, 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 beimißt“⁴³.

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-selegierten 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:

Fall 1: $50 - 10 = 40$

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

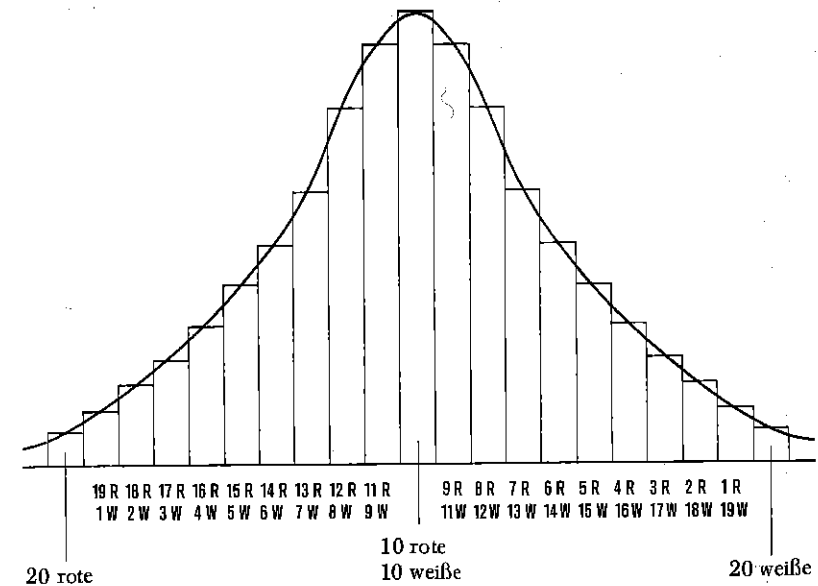
In einem Behälter befinden sich zwei in jeder Hinsicht gleiche Kugeln, doch ist die eine rot und die andere weiß.

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-

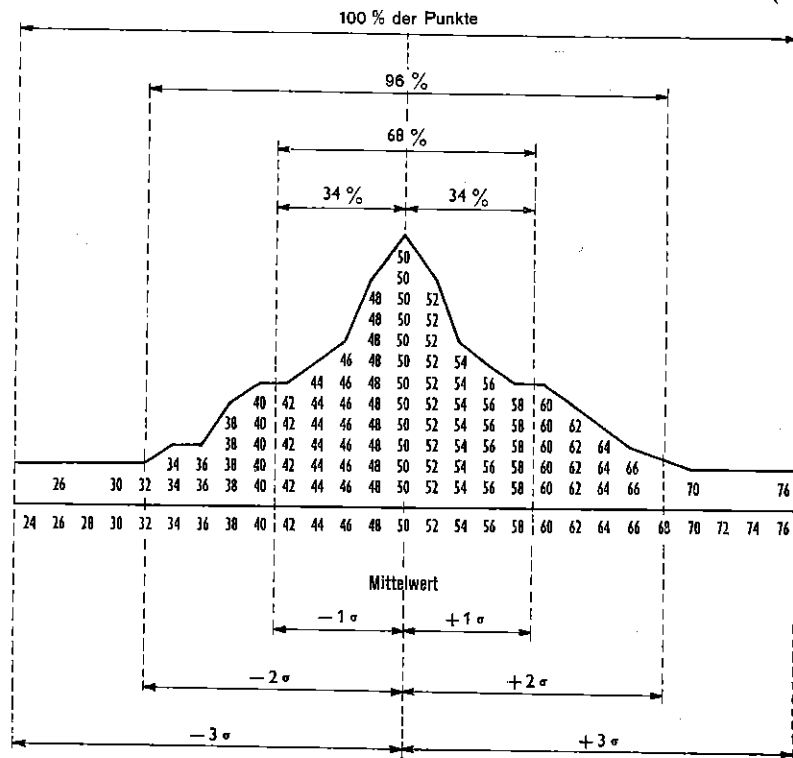
tät sicherlich unermeßliche Dienste und stellt zweifellos eines der wichtigsten Instrumente für den Forscher dar.

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 (\bar{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. dtisch. 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)	Klassen-grenzen	Mittelwert der Klasse (\bar{X}_K)	Häufigkeit (f_K)	Abwei-chung 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. Piéron, Vocabulaire de la psychologie.

47 Bei dem französischen System — das immer weniger angewandt wird — ist das erste Perzentil 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}	$48,5 + \frac{50-44}{22} \cdot 5 = 49,9$	$38,5 + \frac{16-11}{10} \cdot 5 = 41$	$53,5 + \frac{84-66}{18} \cdot 5 = 58,5$

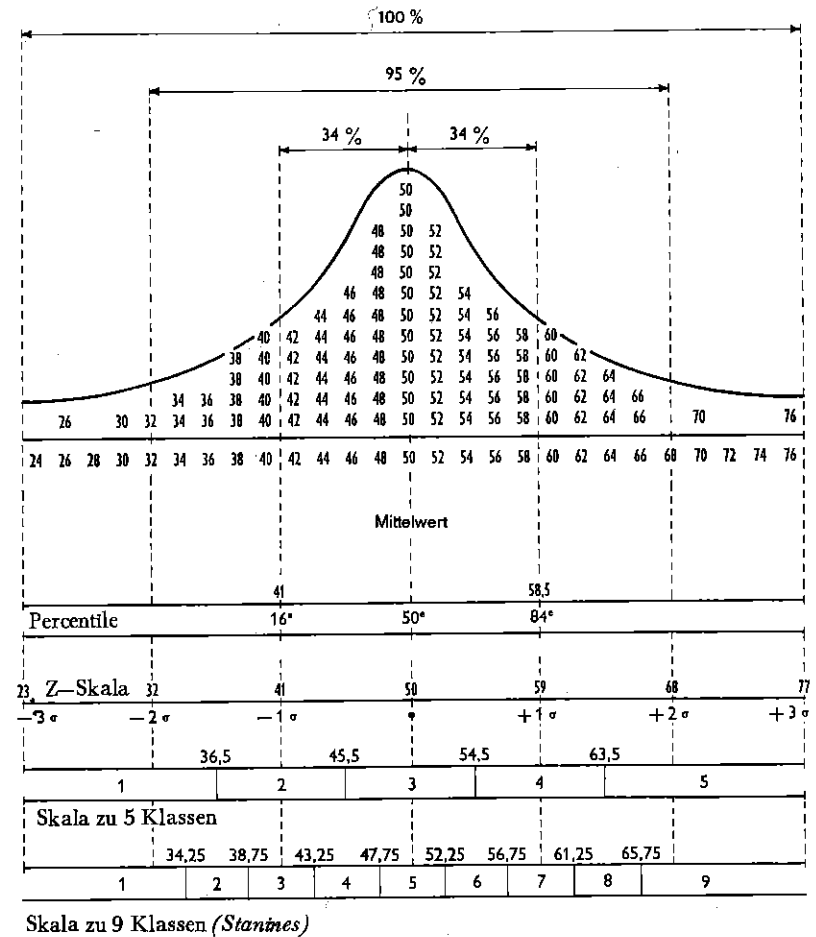
Mittlerer Quartilabstand ($C_{75} - C_{25}$)/2.

$$C_X = \frac{\text{Rang der Maßzahl}}{(N \cdot C)/100}$$

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).



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

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

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

i = Intervallbreite.

2. Die Normalisierung

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

	-2σ	-1σ	0	$+1\sigma$	$+2\sigma$
Percentile	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 verdienstvoller 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:

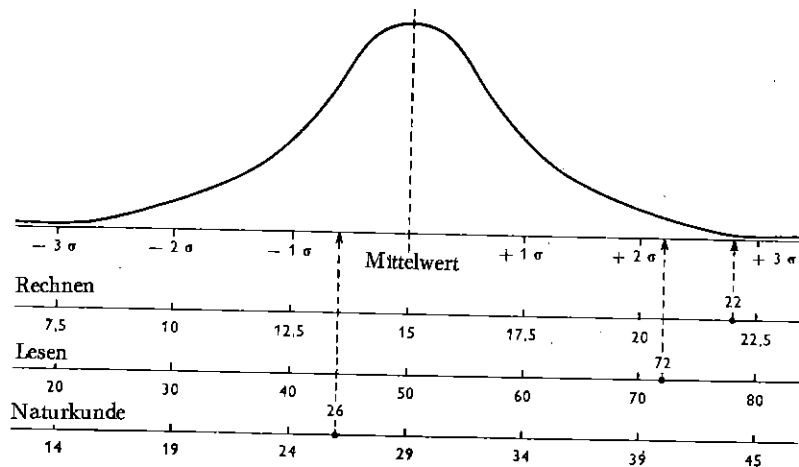
$$X = \bar{X} + z \cdot s$$

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

$$X = 22$$

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 Stehe 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:		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			
1000							
∞							

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
$N = 14$	3448	3395		2479

Mittelwert der Punkte „nachher“: $3448 : 14 = 246,3$.
 Mittelwert der Punkte „vorher“ : $3395 : 14 = 242,5$.
 $\bar{D} = 246,3 - 242,5 = 3,8$.

$$s = \sqrt{\frac{\sum D^2}{N-1}} = \sqrt{\frac{2479}{13}} = 13,8.$$

$$t = \frac{3,8}{13,8/\sqrt{14}} = \frac{38\sqrt{14}}{138} = 1,02.$$

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

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

$$= \frac{(8-8,75)^2}{8,75} + \frac{(10-10)^2}{10} + \frac{(7-6,25)^2}{6,25} + \frac{(16-13,3)^2}{13,3} + \frac{(15-15,2)^2}{15,2}$$

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

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*):

$$TSS = \sum X^2 - (\sum X)^2 / N = 673$$

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		Σ X = 571 Σ 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.

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 einzuhalten, wäre nicht nur ein Betrug, sondern würde auch das Mißtrauen gegenüber 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 Absichten 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 mitteilbar 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

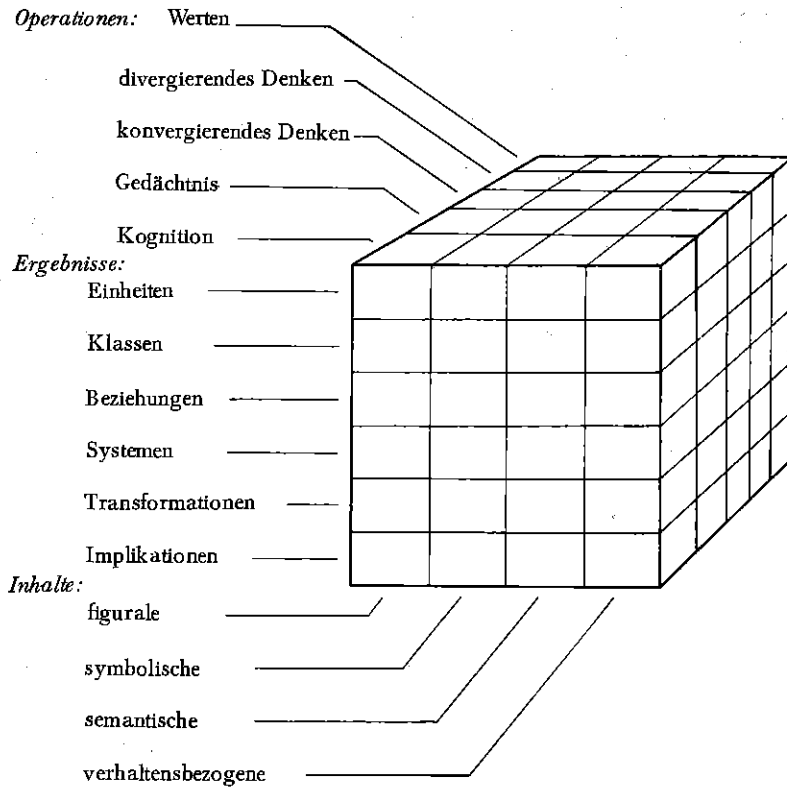
Jedes Ergebnis kann seinerseits abhängig sein von vier Inhaltskategorien:

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. Bearbeiter: 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:



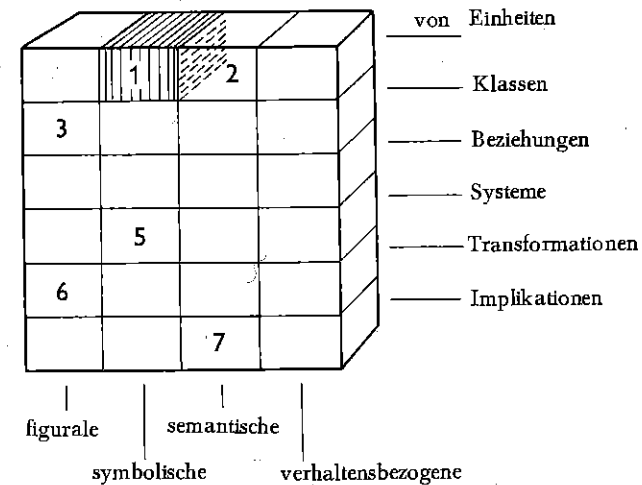
Dieses Modell bietet 120 Kombinationsmöglichkeiten, von denen jede einzelne eine Funktionsart darstellt, die durch einen Test untersucht werden kann.

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.

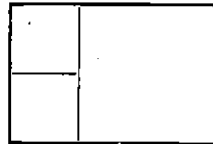
Nikolaus macht unser Tor.

Block 6 (Fähigkeit zu divergierendem Denken von figuralen Transformationen) beinhaltet 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.

18-104

PAE 001

A0341(002)

T08926613

Uni Bw Hamburg

705

705/080082661



PAE 001
A0341(002)

V08926613-1



= > CODE 39 für Fremdbibliotheken

Mit der ständig wachsenden Bedeutung der empirisch-methodischen Betrachtungsweisen in der Pädagogik wird ein leicht zugänglicher und knapp gehaltener Überblick über diese empirischen Verfahren für Studenten, Lehrer und in der Forschung Tätige unerlässlich. Das bietet der Autor – Forscher und Pädagoge zugleich – mit dem vorliegenden Werk, in dem die Leitlinien und das Instrumentarium der modernen pädagogischen Tatsachenforschung zusammengestellt sind. Die Darstellung der modernsten Methoden und Arbeitstechniken umfaßt auch eine Einführung in die elektronische Datenverarbeitung und die statistischen Analysetechniken. Dabei wurden in der deutschen Bearbeitung besonders die in der Bundesrepublik verfügbaren und gebräuchlichen Methoden berücksichtigt. Das Buch wird vielen Lesern eine willkommene Hilfe sein.

Verlag Julius Beltz · Weinheim · Berlin · Basel