

Die Ganzheits- schule

Blätter
zur Theorie
und Praxis
neuzzeitlicher
Bildungsarbeit

Herausgegeben von Artur Kern

Mitbegründet von Alois Gorges †

Pädag. Hochschule
München
Bibliothek

Inhalt

Heinz Schlechtweg,

Zur Mathematisierung der Didaktik
des Rechenunterrichts aller Schularten

Artur Kern,

Parallel-Test (P) zum Grundleistungstest (GLT)

Wilhelm Bührlé,

Ein Verfahren zur Berechnung des Unterschieds

Walter Riethmüller,

Soll der Lehrer ein neues Lesestück vorlesen?

Erwin Schwartz,

Im Niemandsland zwischen Medizin und
Pädagogik

Verlag Herder Freiburg

Postverlagsort: 7800 Freiburg im Breisgau

1

Jan./Febr. 1967
16. Jahrgang

Inhalt		
	<i>Heinz Schlechtweg, Zur Mathematisierung der Didaktik des Rechenunterrichts aller Schularten</i>	1
	<i>Artur Kern, Parallel-Test (P) zum Grundleistungstest (GLT)</i>	7
	<i>Wilhelm Bührlé, Ein Verfahren zur Berechnung des Unterschieds</i>	13
	<i>Walter Riethmüller, Soll der Lehrer ein neues Lesestück vorlesen? (Ein Beitrag zur Didaktik der literarischen Erziehung)</i>	18
	<i>Erwin Schwartz, Im Niemandsland zwischen Medizin und Pädagogik (Seifenblasen — ferngesehen)</i>	21
	<i>Artur Kern, Lehrfilm zum Rechenkasten von Prof. Artur Kern</i>	23
	<i>Buchbesprechungen — Bucheingänge</i>	24

- Anschriften der Mitarbeiter dieses Heftes:** Rektor Wilhelm Bührlé, 7320 Göppingen-Jebenhausen, Schulstraße 6 — Professor Artur Kern, 7815 Kirchzarten bei Freiburg, Bahnhofstraße 48 — Dozent Walter Riethmüller, 6900 Heidelberg, Humboldtstraße 19 — Professor Dr. habil. Heinz Schlechtweg, 4307 Kettwig, Hauptstraße 8 (am Hexenberg) — Professor Erwin Schwartz, 2900 Oldenburg, Hoyersgang 23.
- Herausgeber:** Professor Artur Kern, 7815 Kirchzarten bei Freiburg, Bahnhofstraße 48
- Beiräte:** Stadtoberschulrat Hans Elschenbroich, Garmisch-Partenkirchen — Schulrat Hans Erben, Wittlich — Hochschul-Dozent Hans Gieding, Heidelberg — Rektor Heinrich Jeanjour, Ehrang/Trier — Professor Erwin Kern, Heidelberg — Schulrat Paul Nentwig, Ahaus/Westfalen — Professor Dr. Karl Samstag, Saarbrücken — Schulrat Hermann Wacker, Alfeld/Lelne
- Schriftleitung:** Professor Wilhelm Straub, 7600 Offenburg, Telchstraße 38
- Beiträge, Zuschriften** sind zu richten an die Schriftleitung der Ganzheitsschule in Offenburg
- Besprechungsstücke** werden von Fall zu Fall von der Schriftleitung angefordert
- Erscheinungsweise:** Zweimonatlich in jährlich 6 Heften, jeweils zum 15. des Ausgabemonats, Jahrgangsbeginn im Januar
- Bezugspreis:** Jahresabonnement 8.00 DM, Einzelheft 1.60 DM, zuzüglich Zustellgebühr
- Bestellung und Versand:** Bezugsmöglichkeiten durch den Buchhandel, die Post oder durch den Verlag Herder, 7800 Freiburg im Breisgau, Hermann-Herder-Straße 4. Bei Rückfragen und Zahlungen bitten wir, stets die Kundennummer anzugeben. Kündigungen nur zum Jahrgangsende
- Abbestellungen:** jeweils bis Ende Dezember für den folgenden Jahrgang
- Zahlungsmöglichkeiten:** auf Postscheckkonto Karlsruhe Nr. 389 03 und auf Konto Nr. 10 100 der Öffentlichen Sparkasse Freiburg i. Br.

Verlag Herder Freiburg
 Druck: Herder Druck Freiburg im Breisgau

Empirische Ansätze zur Erfassung des Methodeneffektes (beim Erstlese- und Erstschreibunterricht) und ihre Problematik

Von Kurt Heller

Einleitung

In letzter Zeit wurden von verschiedener Seite, in Buchpublikationen, Zeitschriften, Tageszeitungen, Boulevardblättern u. ä. m., mit mehr oder weniger oder auch gänzlich fehlender Sachkenntnis Nutzen und Erfolg der Ganzheitsmethode in Frage gestellt. Da in der „Ganzheitsschule“ wiederholt hierzu Stellung bezogen wurde, können wir uns auf einen einzigen Hinweis beschränken, die Monographie von *Paul Fleig*: „Ist die Synthese tot?“¹ Vom polemischen Grundcharakter und den unqualifizierten Auslassungen gegen die Gebrüder *A.* und *E. Kern* einmal abgesehen, demonstriert hier *Fleig* recht eindrucksvoll, wo die Grenzen spekulativer Beweisführung in der wissenschaftlichen Entscheidung der anstehenden Methodenfrage gelegen sind. Eine wissenschaftlich verbindliche Klärung wird u. E. hier nur via Empirie zu erzielen sein. Obwohl wir den sachlogischen Ausführungen von *F. Kron*² weitgehend zustimmen, können wir seine resümierende Schlußfolgerung nicht akzeptieren: „Die Unterrichtswirklichkeit bleibt also einer empirisch exakten Erforschung zunächst entzogen und mit ihr zugleich die exakten Aussagen über Erfolg oder Mißerfolg der Ganzheitsmethoden.“³ Wir sind vielmehr davon überzeugt, daß auch diese Frage empirisch beantwortet werden kann, eine endgültige und wissenschaftlich bindende Entscheidung dürfte sogar ausschließlich auf dem Boden pädagogischer Tatsachenforschung fallen. Dabei werden profunde Sachkenntnis oder persönliche Erfahrung eines einzelnen keinesfalls als gering erachtet; im wissenschaftlichen Beweisverfahren gewinnen sie jedoch bestenfalls untergeordnete Bedeutung, weil sie dem wichtigsten Wissenschaftskriterium, der Nachprüfbarkeit und Kontrollierbarkeit wissenschaftlicher Aussagen, nicht oder nur unzulänglich genügen⁴. Somit fällt,

wie in den meisten offenen didaktischen, schulpädagogischen und -psychologischen Fragen und Problemen, auch in der anstehenden Methodenentscheidung hinsichtlich des Nutzeffektes der empirischen Tatsachenforschung die Hauptentscheidungslast zu. Daß diese Entscheidung bereits nach dem heutigen Stande der Wissenschaft möglich sein dürfte, sollen unsere folgenden Ausführungen verdeutlichen. Dabei sind wir uns völlig im klaren darüber, daß es sich hier um eines der schwierigsten (freilich nicht unmöglichen) Unterfangen handelt.

In den letzten Jahren sind im deutschsprachigen Raume zwei verdienstvolle Buchpublikationen zum angezeigten Thema erschienen⁵. An Hand dieser empirischen Modellarbeiten (beide sind als psychologische Dissertationen entstanden) wollen wir uns die Probleme, die im Zusammenhang eines solchen Vorhabens auftreten, vergegenwärtigen, wobei wir zusätzlich auf eigene Untersuchungen aus den Jahren 1957—1959⁶, soweit sie hier einschlägig erscheinen, zurückgreifen werden. Eine solche wissenschaftstheoretische resp. -methodische Besinnung ist besonders aus zwei Gründen notwendig:

1. Der Empiriker, sofern er als Wissenschaftler tätig wird, muß das methodische Rüstzeug kennen, das ihm im (aktiven) Wissenschaftsvollzug zur Verfügung steht. Die genaue Kenntnis der einschlägigen Verfahrensweisen und Methoden (Methode der Beobachtung und Deskription, Beurteilungstechniken, Experiment, statistische Verfahrensweisen, Skalierungsmethoden, aber auch Testtechniken) wird hier unerlässlich.

gründet sind, daß jeder Schritt — jedenfalls prinzipiell — nachvollziehbar ist.“ Im Sinne dieser Wissenschaftsauffassung stellt die Kontrollmöglichkeit die *Conditio sine qua non* dar.

⁵ *E. Schmalohr*, Psychologie des Erstlese- und Schreibunterrichts. Beiheft 16 der Ztschr. „Schule und Psychologie“, München 1961. — *H. Müller*, Methoden des Erstleseunterrichts und ihre Ergebnisse. Schriftenreihe der HIPP in Ffm, Meisenheim am Glan 1964.

⁶ *K. Heller*, Korrelationen zwischen akustischer und optischer Gliederungsfähigkeit und Rechtschreibleistung in analytischen und synthetischen Klassen. Wiss. Hausarbeit zur II. Dienstprüfung für das Lehramt an Volksschulen, Heidelberg 1959 (unveröffentlicht).

¹ *P. Fleig*, Ist die Synthese tot? Bochum 1965.

² *F. W. Kron*, Doch besser lesen lernen nach Ganzheitsmethode, in: Ganzheitsschule, Jg. 15 (1966), H. 4, S. 78 bis 84.

³ Ebd., S. 84.

⁴ *C. F. Graumann* definiert z. B.: „Wissenschaft ist ein System eindeutig formulierter Erkenntnisse, die so be-

Wir können in dem hier zur Verfügung stehenden Rahmen bestenfalls gelegentlich einige dieser Punkte berühren; die Hauptaufgabe der Information liegt bei den Pädagogischen Hochschulen, sofern sie — neben ihrer Ausbildungsfunktion — im engeren und weiteren Bezirk der Schulpädagogik echte Wissenschaftsarbeit betreiben⁷. Daß diese (schulpädagogische) Forschungsarbeit mehr und mehr den Pädagogischen Hochschulen als Aufgabe zufallen wird, ist unsere Überzeugung; ein Blick in die relevanten Universitätsdisziplinen und ihre laufenden Forschungsvorhaben stützt unsere Behauptung weitgehend. Ausnahmen (siehe *Schmalohr* und *Müller*) bestätigen auch hier nur die Regel. Doch wird (und braucht) keineswegs jeder angehende oder praktizierende Pädagoge aktiv wissenschaftlich tätig zu werden; eine solche Forderung würde die realen Gegebenheiten falsch einschätzen.

2. Unabdingbar erscheint uns jedoch das Postulat an jeden Schulpädagogen, sofern er mehr als nur „Handwerker“ sein will, sich die Voraussetzungen für ein kritisches, aber sachliches Urteil über die aktuellen Probleme seines Fachbereiches, z. B. die angeschnittene Methodenfrage, anzueignen. Hierzu gehören neben den notwendigen „handwerklichen“ Voraussetzungen und einigen unentbehrlichen Grundkenntnissen bestimmte Einsichten in den Aufbau und die Struktur empirischer Forschungsarbeiten; erst dann wird ein sachangemessener, „objektiver“ Standpunkt bezüglich Ergebnis und Wert solcher Arbeiten auch vom „Nur-Praktiker“ bezogen werden können. Vorab in dieser Intention möchten wir unsere Erörterungen verstanden wissen.

Hauptteil

Alle empirischen Vorgehensweisen beinhalten drei Hauptproblemkreise: Organisation der Stichprobe, Probleme des methodischen (operationalen, statistischen) Ansatzes und schließlich die Interpretation des Untersuchungsbefundes (Frage nach der Generalisationsbasis). Unter diesen Gesichtspunkten wollen wir jetzt die vorgenannten Arbeiten analysieren und außerdem einige neue, in Deutschland bislang wenig bekannte, u. E. jedoch hier relevante, Verfahren diskutieren.

⁷ Für eine erste Orientierung empfehlen wir: *W. Traxel*, Einführung in die Methodik der Psychologie, Bern-Stuttgart 1964. — *W. D. Fröhlich*, Forschungsstatistik, Bonn 1965(4). — *E. Mittenecker*, Planung und statistische Auswertung von Experimenten, Wien 1963(4).

1. Probleme der Stichprobenbildung

Bei einem bestimmten Forschungsvorhaben ist es in der Regel unmöglich, alle Personen, die genau bezeichnete Merkmale tragen, also einer wohldefinierten Gruppe angehören, bei der anstehenden empirischen Untersuchung zu berücksichtigen. So können z. B. bei einem Lehrmethodenvergleich in praxi nicht sämtliche Schüler des 2. Volksschuljahres in der Bundesrepublik, also nicht die gesamte Population⁸, sondern nur ein kleiner Teil hiervon, eine Stichprobe⁹ von vielleicht wenigen hundert oder auch einigen tausend Zweitkläßlern untersucht werden. Damit nun die Ergebnisse, die durch Untersuchung einer solchen Stichprobe gewonnen wurden, auch für die übrigen Angehörigen desselben Kollektivs, die nicht in die Stichprobe gelangten, aussagekräftig sind, müssen bestimmte Regeln und Kautelen bei der Stichprobenbildung eingehalten werden. Es gibt zwei verschiedene Arten von Stichproben:

- a) die Zufallsstichprobe,
- b) die repräsentative oder organisierte oder stratifizierte Stichprobe.

Eine Zufallsstichprobe liegt dann vor, wenn jedes Mitglied der Population die gleiche Chance hat, in die Stichprobe zu gelangen. Zufallsstichproben können durch Los- oder Lotterieverfahren, durch Benutzung einer Tabelle, die Zufallsfolgen von Ziffern enthält (vgl. *E. Mittenecker*, a. a. O., S. 154, Tafel II), u. ä. zusammengestellt werden. Häufiger findet jedoch die stratifizierte oder Repräsentativstichprobe Anwendung, wie wir sie besonders aus der Markt- und Meinungsforschung kennen. Hierbei ist freilich die Kenntnis der „strata“, der relevanten bzw. interessierenden Schichten, z. B. Alter, Geschlecht, Schulbildung, Klasse (Schuljahr), Methodenzugehörigkeit, sozioökonomischer Status des Elternhauses (Vaterberuf), soziogeographische Bedingtheiten (Stadt-Land) u. a. m., unerläßliche Voraussetzung. Ohne die Kenntnis der Verteilung dieser Kriterien, ihres relativen Anteils an der Gesamtpopulation (0/o-Satz) usw. kann keine repräsentative Stichprobe organisiert werden. M. a. W.: Die

⁸ Population oder Kollektiv kann man als Gesamtheit aller Individuen, die gleiche Merkmale oder Merkmalskombinationen tragen, definieren. Eine Population ist also keine natürliche Gruppe; sie entsteht vielmehr aufgrund einer Definition, z. B. alle Volksschüler der Grundschule oder alle Zweitkläßler der Volksschule in der Bundesrepublik, alle PH-Studenten, u. ä.

⁹ Eine Stichprobe ist eine relativ kleine Gruppe von Angehörigen einer bestimmten Population, die zum Zwecke einer näheren Untersuchung herausgegriffen wird.

Strata der Population müssen in gleichen Anteilen in der kleineren Repräsentativstichprobe vertreten sein; wenn keine kongruenten Verhältnisse vorliegen, sind demnach die an der Stichprobe gewonnenen Untersuchungsergebnisse von keinem oder nur sehr zweifelhaftem Wert. Häufig werden jedoch in der Praxis Mischformen beider Stichprobentypen verwendet, z. B. der Modus der „geschichteten Auswahl“, wie ihn René König beschrieben hat¹⁰. Auch die erwähnten Arbeiten von Schmalohr, Müller und des Verfassers basieren auf einer geschichteten Zufallsauswahl.

Die Stichprobe von Schmalohr umfaßte 400 Probanden (200 ganzheitlich nach Kern und 200 lautsynthetisch geschulte) aus dem 5. Schuljahr der Volks- und Oberschule. Die untersuchten Sextaner und Volksschüler der 5. Klasse waren im Elementarunterricht mindestens 2 Jahre lang bei den gleichen Lehrkräften nach der ganzheitlichen oder synthetischen Methode unterrichtet worden. Da Knaben und Mädchen jeweils zu gleichen Anteilen vertreten waren, konnten Lehrmethodengruppen zu 50, 100 und 200 Pbn vergleichend gegenübergestellt werden. Sämtliche Schüler stammten aus Krefeld, so daß schon hier der Verdacht angemeldet werden soll, daß die Untersuchungsbefunde der Schmalohrschen Querschnittsanalyse möglicherweise nur für den Raum Krefeld, bestenfalls noch für groß- und randstädtische Verhältnisse Gültigkeit haben werden. Eine Kontrolle der Intelligenzverteilung in den Vergleichsgruppen via Test fand nicht statt. Trotzdem meint Schmalohr: „Die Beteiligung dieser Anzahl (400 Pbn; d. Verf.) zufällig ausgewählter Schüler aus je 12 dem Stadtbezirk nach parallelen Klassen mit insgesamt 24 Grundschullehrern läßt erwarten, daß in den Leistungen der Kinder der Lehrmethodeinfluß isoliert worden ist und von der Lernfähigkeit der Schüler und der Einwirkung der Lehrer her gleichwertige Bedingungen in den Untersuchungsgruppen vorlagen.“¹¹

Wesentlich detaillierter sind die Angaben Müllers über seine Stichprobenkriterien. Müller bildete aus vollausgebauten Stadt-schulen Wiesbadens und Frankfurts Stichproben des 2. Schuljahrs (N = 576) und des 4. Schuljahrs (N = 358), wobei er jeweils drei Untergruppen vergleichend gegenüber-

stellte: synthetische Lautiermethode (sy), Ganzwortmethode (Gw) und Ganzsatzmethode (Gs). Die Kontrolle der Variablen: Geschlecht, Intelligenz und sozioökonomischer Status des Elternhauses ergab, daß diese wichtigsten relevanten Faktoren in den einzelnen Vergleichsgruppen nicht gleichgehalten werden konnten, worauf bereits Klauer hingewiesen hat¹². In der Reihenfolge Gs — Gw — sy waren die Knaben in folgender Überlegenheit vertreten: 64 % — 54 % — 52 %; die sozial schwächeren Pbn verteilen sich in der gleichen Reihenfolge: 72 % Kn / 76 % Md — 62 % Kn / 67 % Md — 56 % Kn / 50 % Md; auch die Intelligenzschwachen (nach Progressive Matrices von Raven) waren anteilmäßig am stärksten bei den Ganzheitlern vertreten: 39 % Kn / 48 % Md — 35 % Kn / 45 % Md — 30 % Kn / 35 % Md¹³. Für die Stichprobe des 4. Schuljahres ist noch anzumerken, daß die nachträgliche Lehrerbefragung bezüglich der „Methodenreinheit“ der betr. Pbn-Klassen im 1. und 2. Schuljahr nur sehr vage Ergebnisse erbrachte, was die Zuverlässigkeit dieser Stichprobe weiter herabsetzt¹⁴. Ferner ist eine Benachteiligung der Gs-Klassen gegenüber den sy-Klassen in bezug auf Lehrerwechsel im sample des 2. Schuljahres feststellbar.^{14a} Zusammenfassend können wir bereits jetzt schon festhalten: Während das Schmalohrsche sample unter der Kautel, daß die (spärlichen) Angaben zur Stichprobenkontrolle zutreffen — die zugehörigen Unterlagen fehlen in der Publikation größtenteils —, einigermaßen Gleichverteilung in den wichtigsten Populationsvariablen erreichte, offenbarte die Sample-Analyse bei Müller eindeutige Benachteiligung der ganzheitlich geschulten Pbn, wobei am einschneidendsten die Kinder der Ganzheitsmethode betroffen wurden. Die Müllersche Stichprobe repräsentiert somit die bedeutendsten Einflußvariablen bezüglich des Methodeneffektes in den einzelnen Vergleichsgruppen sehr einseitig zugunsten der synthetischen Schüler, so daß bereits von der Zusammensetzung der Stichprobenkriterien her das Untersuchungs-

¹² K. J. Klauer, Methoden des Erstleseunterrichts und ihre Ergebnisse. in: Ganzheitsschule, Jg. 15 (1966), H. 4, S. 85 f. — Vgl. auch die Buchrezension von K. J. Klauer, in: Päd. Rundschau, 19. Jg. (1965) H. 11, S. 798 f.

¹³ H. Müller, a. a. O., S. 84. — Siehe hierzu auch K. J. Klauer, a. a. O., S. 86, dem wir die %-Zahlen entnommen haben.

¹⁴ H. Müller, a. a. O., S. 59.

^{14a} Ebd., S. 71.

¹⁰ R. König (Hrsg.), Soziologie. Fischer Lexikon 10, Frankfurt a. M. 1959, S. 205.

¹¹ E. Schmalohr, a. a. O., S. 56.

ergebnis in etwa abzuschätzen ist. „Von hier aus läßt sich voraussagen, daß die Gruppe, welche nach der Ganzsatzmethode unterrichtet worden ist, wegen ungünstigerer Lernvoraussetzungen die schlechtesten Ergebnisse bringen muß, während jene Gruppen, welche nach der synthetischen oder nach der Ganzwortmethode unterrichtet worden sind, einigermäßen gleiche Ergebnisse erbringen, nämlich ohne signifikante Differenzen

oder mal zugunsten der einen und mal zugunsten der anderen Gruppe. Genau dies ist das Ergebnis von Müller.“¹⁵

Unsere eigene Stichprobe aus dem Jahre 1959 umfaßte insgesamt 728 Pbn, 360 Schüler des 2. Schuljahres und 368 Schüler des 3. Schuljahres. Diese verteilen sich folgendermaßen auf die drei Methodengruppen Gs, Gw, sy:

Tabelle 1

2. Schuljahr:

159 Pbn Gs (nach Kern)
88 Pbn Gw (nach Kade)
113 Pbn sy

3. Schuljahr:

71 Pbn Gs
106 Pbn Gw
191 Pbn sy

Die folgenden Ausführungen gelten nun der Kontrolle unserer Stichprobe nach Lern-dispositionen, sozialökonomischem Status und Geschlecht der Pbn.

Zum Zeitpunkt der Untersuchungserhebungen (1957/58), jeweils in der 2. Hälfte des 2. bzw. 3. Schuljahrs, standen im deutschsprachigen Raume kaum bewährte, standardisierte und ökonomische Gruppen-Intelligenztests für die benötigten Altersstufen zur Verfügung. Wir kontrollierten deshalb nicht die Intelligenz via Test, sondern das Lehrerurteil (Zeugniszensur) über die Schülerleistung im Rechtschreiben. Dies ist insofern berechtigt, als die

Rechtschreibleistung — und nur dieser Aspekt unserer Arbeit interessiert in diesem Zusammenhang hier — wiederholt als wichtiges Indiz eines (möglichen) Methodeneffektes in der Literatur genannt wird. Da die Verlässlichkeit des Lehrerurteils (LU) im allgemeinen der Aussage eines standardisierten Testverfahrens unterlegen ist, korrelierten wir LU mit den Resultaten unseres Testdiktates „Till Eulenspiegel“ (nach A. Kern), um auf diesem Wege die Gültigkeit des LU abzuschätzen. Folgende sehr signifikante Korrelationsergebnisse wurden ermittelt¹⁶:

Tabelle 2

Gs im 2. Schuljahr: $r = 0.56$
Gw im 2. Schuljahr: $r = 0.67$
sy im 2. Schuljahr: $r = 0.53$
Gs, Gw, sy im 2. Schuljahr: $r = 0.60$

Gs im 3. Schuljahr: $r = 0.74$
Gw im 3. Schuljahr: $r = 0.73$
sy im 3. Schuljahr: $r = 0.61$
Gs, Gw, sy im 3. Schuljahr: $r = 0.67$

Die Zusammenhangskoeffizienten erlauben in diesem Falle dem LU gegenüber Vertrauen; bezüglich LU liegen die Korrelationswerte normalerweise zwischen 0.40 und 0.60¹⁷.

Um nun die Verteilung der Schulleistungsvariablen in unseren Lehrmethodengruppen des 2. und 3. Schuljahrs zu überprüfen, geben wir die errechneten LU-Mittelwerte bekannt:

Tabelle 3

Gs im 2. Schuljahr: $M = 2,4$
Gw im 2. Schuljahr: $M = 2,2$
sy im 2. Schuljahr: $M = 2,9$

Gs im 3. Schuljahr: $M = 3,1$
Gw im 3. Schuljahr: $M = 3,0$
sy im 3. Schuljahr: $M = 3,0$

Sämtliche M-Differenzen sind statistisch nicht zu sichern, also zufallsbedingt. Hieraus schließen wir, daß die Schülerdispositio-

nen, soweit sie als Bedingungsfaktoren für unseren Untersuchungsgegenstand Bedeutung erlangen, in den einzelnen Methoden-Ver-

¹⁵ K. J. Klauer, a. a. O. S. 86.

¹⁶ K. Heller, a. a. O., vgl. hierzu Nachtrag 1959, bes. S. 4 und 7 ff.

¹⁷ Es besteht die begründete Annahme, daß LU im Recht-

schreiben und Rechnen am zuverlässigsten sind, da Schulleistungen in diesen Fächern am leichtesten quantifizierbar sind. Vermutlich dürfte das LU im Lesen oder Aufsatz eine wesentlich geringere Verlässlichkeit ausweisen.

gleichsgruppen ausreichend konstant gehalten werden konnten¹⁸.

Ebenfalls möchten wir die Gleichverteilung der sozialökonomischen Statusvariablen in unseren Vergleichsgruppen annehmen. Als Belege fügen wir die Herkunftsorte unserer ausgewählten Schulklassen einschließlich deren Einwohnerzahlen und Klassenlehrerstellen an. Die Kontrolle der sozioökonomischen und -geographischen Verhältnisse dieser Schulorte an Hand der Unterlagen des

Statistischen Landesamtes Baden-Württemberg in Stuttgart bestätigt im wesentlichen unsere Annahme. Diese Voraussetzung gewinnt auch durch den Auswahlmodus der Test-Schüler hohe Wahrscheinlichkeit, da wir von den in die Auswahl einbezogenen Schulen resp. Schulorten immer sämtliche Pbn, also alle Klassen der betr. Schuljahre für die Untersuchung berücksichtigten; irgendwelche Manipulationen bezüglich der Klassenselektion liegen also nicht vor.

Tabelle 4

2. Schuljahr Gs-Klassen:

	Einwohner 1950	1961 ¹⁹	Kl. Lehrer- stellen:
Kl. 2 (35 Pbn) Adelsheim ^{20a}	2473	2829	7
Kl. 2a (34 Pbn) Buchen Stadt	4023	4670	9
Kl. 2 (9 Pbn) Heidersbach	573	510	2
Kl. 2 (42 Pbn) Osterburken	2438	2805	7
Kl. 2 (16 Pbn) Rosenberg	1037	952	3
Kl. 2 (23 Pbn) Seckach	1337	1847	4
Summe: 159 Pbn			

2. Schuljahr Gw-Klassen:

Kl. 2a (30 Pbn) Walldürn	6084	7356	17
Kl. 2b (28 Pbn) Walldürn	6084	7356	17
Kl. 2c (30 Pbn) Walldürn	6084	7356	17
Summe: 88 Pbn ²⁰			

2. Schuljahr sy-Klassen:

	(Kn)		
Kl. 2b (24 Pbn) Buchen Stadt	4023	4670	9
Kl. 2a (25 Pbn) Hardheim	3574	3879	11
	(Kn)		
Kl. 2b (26 Pbn) Hardheim	3574	3879	11
	(Md)		
Kl. 2 (21 Pbn) Mudau	1747	1755	4
Kl. 2 (17 Pbn) Schlierstadt	924	755	2
Summe: 113 Pbn			

3. Schuljahr Gs-Klassen:

Kl. 3 Kn (18 Pbn) Hardheim	3574	3879	11
Kl. 3 (10 Pbn) Heidersbach	573	510	2
Kl. 3 (43 Pbn) Osterburken	2438	2805	7
Summe: 71 Pbn ²¹			

3. Schuljahr Gw-Klassen:

Kl. 3a (36 Pbn) Walldürn	6084	7356	17
Kl. 3b (35 Pbn) Walldürn	6084	7356	17
Kl. 3c (35 Pbn) Walldürn	6084	7356	17
Summe: 106 Pbn			

3. Schuljahr sy-Klassen:

Kl. 3b (20 Pbn) Adelsheim	2473	2829	7
Kl. 3a (34 Pbn) Buchen Stadt	4023	4670	9
Kl. 3b (38 Pbn) Buchen Stadt	4023	4670	9
Kl. 3 Md (30 Pbn) Hardheim	3574	3879	11
Kl. 3 (24 Pbn) Mudau	1747	1755	4
Kl. 3 (11 Pbn) Rosenberg	1037	952	3
Kl. 3 (11 Pbn) Schlierstadt	924	755	2
Kl. 3 (23 Pbn) Seckach	1337	1847	4
Summe: 191 Pbn			

¹⁸ Der im 2. Schuljahr durchweg feststellbare Trend zur positiv-schiefen Beurteilung der Schülerleistung durch LU findet sich gleichsinnig auch bei den Erhebungen von Schmalobr und Müller. Hierfür sind wohl in erster Linie pädagogische Gründe verantwortlich. Einmal möchte der Lehrer auf dieser frühen Altersstufe weniger Leistungszeugnisse als vielmehr pädagogische Anreize setzen, zum anderen spielt wohl aber auch eine gewisse Scheu des Lehrers hier eine Rolle, einen Schüler bereits zu Beginn seiner Schulbildung eventuell zu negativ zu beurteilen und damit indirekt seine (positiven) Anlagepotenzen in ihrer Entfaltung zu beeinträchtigen.

¹⁹ Gemeindestatistik Baden-Württemberg 1960/61. Teil 1: Bevölkerung und Erwerbstätigkeit. Hrsg. vom Stat. Landesamt Baden-Württemberg, Stuttgart 1964. Vgl. S. 66—73. Differenziertere Angaben (Stand 1949—1963) über den Kreisbezirk Buchen, aus dem sämtliche Pbn unseres samples entnommen worden sind, finden sich in: Die Stadt- und Landkreise Baden-Württembergs in Wort und Zahl. Heft 7: Landkreis Buchen/Odenwald. Hrsg. vom Innenministerium und Wirtschaftsministerium in Baden-Württemberg. Bearbeitet vom Stat. Landesamt, Stuttgart o. J.

^{20a} Alle Orte im Kreis Buchen/Odenwald.

²⁰ Gw-Klassen standen zum Zeitpunkt der Untersuchung im Kreis Buchen ausschließlich an der Volksschule Walldürn zur Verfügung. Kn = Knabenklasse; Md = Mädchenklasse; alle übrigen sind gemischte Klassen.

²¹ Im gesamten Kreis Buchen/Odw. waren 1957/58 ausschließlich diese drei Klassen des 3. Schuljahres nach der Kern-Methode unterrichtet worden. Die Lehrerbefragung bezüglich „Methodenreinheit“, die bei allen untersuchten Schulklassen des 2. und 3. Schuljahres durchgeführt worden ist, ergab für das Gs-sample des 3. Schuljahres folgende Auskünfte: Klasse 3 in Heidersbach hatte zwei Methodenwechsel (im 1. Schuljahr ganzheitlich, im 2. Schuljahr und 1. Hälfte 3. Schuljahr synthetisch, ab Mitte 3. Schuljahr wieder ganzheitlich geführt); Klasse 3 in Hardheim und Osterburken wurden nur im 1. Schuljahr ganzheitlich, danach, also fast zwei Jahre lang, synthetisch unterrichtet. Aus diesen Gründen und bedingt durch die kleine Pbn-Zahl, wobei 1 Klasse (Hardheim) zudem noch eine reine Knabenklasse darstellt, entfällt dieses sample als Methodenvergleichsgruppe. Siehe auch K. Heller, a. a. O., bes. S. 20 ff.

Schließlich wäre noch die Verteilung der Geschlechtsvariablen auf die einzelnen Lehrmethodengruppen zu überprüfen. Die Ana-

lyse ergab, daß dieser Faktor nicht konstant gehalten werden konnte.

Tabelle 5

Gs-Pbn im 2. Schuljahr: 76 Kn + 83 Md = 48% Kn + 52% Md	N = 159 Pbn (100%)
Gw-Pbn im 2. Schuljahr: 51 Kn + 37 Md = 58% Kn + 42% Md	N = 88 Pbn (100%)
sy-Pbn im 2. Schuljahr: 71 Kn + 42 Md = 63% Kn + 37% Md	N = 113 Pbn (100%)
Gs-Pbn im 3. Schuljahr: 40 Kn + 31 Md = 56% Kn + 44% Md	N = 71 Pbn (100%)
Gw-Pbn im 3. Schuljahr: 52 Kn + 54 Md = 49% Kn + 51% Md	N = 106 Pbn (100%)
sy-Pbn im 3. Schuljahr: 93 Kn + 98 Md = 49% Kn + 51% Md	N = 191 Pbn (100%)

Die Übersicht über die Verteilung des Geschlechtsmerkmals zeigt eine Benachteiligung der sy- und Gw-Pbn im 2. Schuljahr sowie eine Benachteiligung der Gs-Pbn im 3. Schuljahr. Wir versuchten deshalb, diesen Faktor zu kontrollieren, indem wir die Signifikanzberechnungen der M-Differenzen (in der Diktatleistung) zwischen den einzelnen Methodengruppen zusätzlich getrennt nach dem Geschlecht durchführten. Ohne den späteren Erörterungen hierüber vorzugreifen, möchten wir an dieser Stelle schon betonen, daß die Kontrolle der Geschlechtsvariablen im 2. Schuljahr keine, und im 3. Schuljahr nur geringfügige Veränderungen (zugunsten der Gs-Pbn) im Gesamtergebnis der untersuchten Rechtschreibleistungen bewirkte. Bezüglich der Aussagekraft unseres Sample-Ergebnisses bleibt allerdings kritisch anzumerken, daß sich seine Gültigkeit möglicherweise nur auf ländliche sowie klein- und mittelstädtische Verhältnisse erstreckt. Unsere Stichprobe stellt in soziogeographischer Hinsicht das Pendant zu den Auswahlklassen von *Schmalohr* und *Müller* dar. Eine Extrapolation unserer Untersuchungsbefunde auf großstädtische Schulverhältnisse erscheint freilich eher angezeigt, als die gegenläufige Generalisation bei *Schmalohr* und *Müller*, da wir neben Pbn aus wenig gegliederten Landschulen auch vollausgebaute Schulklassen berücksichtigen konnten, während die genannten Autoren ausschließlich auf solche (Großstadtklassen) zurückgegriffen haben. Trotzdem sollte man bei solchen Verallgemeinerungen eine gewisse Vorsicht walten lassen.

2. Der methodische Zugang zum Untersuchungsgegenstand

Zur Bewältigung des hier anstehenden Methodenproblems bietet die operationale Psychologie vorab zwei Grundverfahren an: Beobachtung und Experiment. Obwohl der phänographische Zugang zur Gegenstandsthematik prinzipiell möglich erscheint, dürften sich in praxi doch erhebliche

Schwierigkeiten ergeben. Ohne ein geschultes Beobachter-Team z. B. würde die wissenschaftliche Deskription (Beobachtung — Beschreibung — Beurteilung) kaum zuverlässige Ergebnisse liefern. Es ist deshalb kein Zufall, daß die bisherigen Arbeiten experimentalphysikalische Untersuchungen darstellen. Obgleich das Experiment (schon seit *Wundt*) als eine Variante der Beobachtungstechnik²² angesehen werden kann, ist es doch eine eigene Methode. Beobachtung, willkürliche Einwirkung, systematische, isolierende Variation und Wiederholbarkeit (Kontrolle) sind die vier Hauptkernfunktionen des Experiments. Im Experiment geht es um die Messung wechselseitiger Bedingtheiten von Aspekten, die zwecks genauer Beobachtung (vorher) festgelegt und während des Verlaufs isoliert werden, wobei die nichtinteressierenden Züge ausgeklammert resp. konstant gehalten werden. Im Blick auf diese Idealform des Experiments wollen wir nun die vorliegenden Felduntersuchungen auf ihre methodische Angemessenheit hin kritisch überprüfen.

Bei der Veri- bzw. Falsifikation von Hypothesen spielt die sogenannte Nullhypothese eine wichtige Rolle. Einer bestimmten Annahme wird hier eine negierende These entgegengesetzt. Zum Beispiel: Zwischen Ganzheitsmethode und synthetischer Methode bestehen keine Unterschiede hinsichtlich der Lese- und Rechtschreibleistungen; die von den Fachdidaktikern und -methodikern verschiedener Provenienz vermuteten oder behaupteten Differenzen sind nicht vorhanden, sind gleich „null“ („Null“-Hypothese). Eine solche Nullhypothese hat erkenntnistheoretische Voraussetzungen. Die Nullhypothese wird aufgestellt, um widerlegt zu werden, d. h., es soll in überzufälliger Weise festgestellt werden, daß man sich in der Nullhypothese geirrt hat. Diesem Vorgehen kommt eine logische Regel entgegen:

²² So definiert z. B. *W. Traxel* das Experiment als „absichtliche, planmäßige Auslösung eines Vorgangs zum Zwecke der Beobachtung“.

Ein einziger Fall genügt bereits, um eine Hypothese ad absurdum zu führen; dagegen bedarf es vieler (oft tausend und mehr) Beleg-Fälle, um einen positiven Beweis führen zu können. Mit anderen Worten: Es ist leichter, etwas zu widerlegen, als eine positive Annahme oder eine vorgefaßte Meinung zu beweisen, die man zudem möglicherweise wieder zurücknehmen müßte. Die Absicht, sich selbst zu widerlegen, ist demnach der Hauptvorteil des Entscheidungsexperiments, das außerdem gegen Zufälligkeiten gesichert ist.

Exemplarisch bieten wir die von Schmalobr geprüften Hypothesen, wobei wir den Autor selbst zu Worte kommen lassen. „Im Für und Wider der Lehrmethodendiskussion überwiegen die folgenden Hypothesen über die Auswirkungen der Verfahrensweisen, die in der Untersuchung an Sextanern und Volksschülern aus dem 5. Schuljahr überprüft wurden:

1. *Die Ganzheitsmethode fördert die Durchgliederungsfähigkeit der Kinder besser als die Lautiermethode.*

2. *Bei Anwendung der Ganzheitsmethode leidet die Arbeitshaltung der Schüler, während die Lautiermethode zu einer konsequenten Arbeitshaltung führt.*

3. *Die Schulleistungen der ganzheitlich unterrichteten Kinder liegen beim Übergang in die höhere Schule und die Oberstufe der Volksschule merklich über denen der lautsynthetisch unterrichteten.*

4. *Besonders im Deutschunterricht schneiden die „Ganzheitler“ besser ab.*

5. *Für den Lateinunterricht schafft die systematisch aufbauende Lautiermethode die günstigeren Voraussetzungen.*

6. *Die ganzheitliche Arbeitsweise führt zu einer bedeutenden Überlegenheit im sinnverständigen Lesen.*

7. *Die Ganzheitsmethode hat eine spürbar bessere Rechtschreibung im Gefolge.*

8. *Nur die gut begabten Kinder werden im Ganzheitsunterricht gefördert.*

9. *Für schwächer begabte Schüler bringt die Lautiermethode Vorteile mit sich.*

Auf den Aufsatzunterricht ist keine Hypothese ausgerichtet, weil ein dafür geeignetes Meßinstrument fehlt. Auch die Annahmen über die Lernfreude der Kinder lassen sich in hier geforderten Gruppentests nur schwer angehen . . . Die verschiedenen Aspekte der frühen Leseleistung sind ebenfalls nicht einzeln beachtet, da es beim reifen Lesen schließ-

lich auf das Leseverständnis ankommt und in Gruppentests nur das stille und nicht das laute Lesen geprüft werden kann.“²³

Ähnlich formulierte H. Müller 10 Hypothesen, wobei er zusätzlich das mündliche Lesen miteinbezog, bedauerlicherweise aber die Rechtschreibleistung und — aus den schon genannten Gründen — einen Aufsatzvergleich ausschloß. Stellvertretend zitieren wir Hypothese 1 und 6²⁴:

1. *Verschiedene Leselehrmethoden erzielen keinen unterschiedlichen Effekt in bezug auf Fehler / Zeit beim mündlichen Lesen; auftretende Abweichungen sind zufällig . . .*

6. *Verschiedene Leselehrmethoden erzielen keinen unterschiedlichen Effekt in bezug auf das Sinnverstehen beim stillen Lesen; auftretende Abweichungen sind zufällig.*

Von unserer eigenen Untersuchung, die ja in anderem Kontext entstanden ist, interessiert hier nur der Rechtschreibeaspekt. Auf die Ergebnisse werden wir gleich noch zu sprechen kommen.

Sämtliche Thesen wurden als Nullhypothesen formuliert und mit einschlägigen (parametrischen und nonparametrischen) statistischen Verfahren auf ihre Signifikanz hin überprüft. Grundsätzlich ist bei der Aufstellung von Hypothesen darauf zu achten, daß diese auch tatsächlich die aufgestellten Behauptungen und Tatbestände treffen. Auch hier gilt das Postulat, daß das verwendete Modell in Isomorphie zur Sache steht. Da die Art der Hypothesenbildung jedoch immer auch in Rücksicht auf das zur Überprüfung notwendige und jeweils zur Verfügung stehende Instrumentarium geschieht, bleiben oft wesentliche Momente der Gegenstandsthematik unberücksichtigt, was zu einseitigen und verzerrten Ergebnissen führen kann. Die Problematik dieser Situation wird sehr witzig durch eine amerikanische Anekdote, die wir dem Leser nicht vorenthalten wollen, illustriert: Am späten Abend trifft ein Soziologe einen bekannten Psychologen, als dieser im Schein einer Straßenlaterne seinen verlorenen Hausschlüssel sucht. Nach einer Weile gemeinsamen und ergebnislosen Bemühens fragt schließlich der Soziologe den Psychologen, ob er tatsächlich den Gegenstand an dieser Stelle verloren habe. Dieser verneint und begründet dem erstaunten Kollegen sein merkwürdiges Tun damit, daß er den Schlüssel deshalb hier suche, weil gerade

²³ E. Schmalobr, a. a. O., S. 31 f.

²⁴ H. Müller, a. a. O., S. 56 f.

eine Laterne hier stünde und diese Stelle beleuchte; dort, wo er den Schlüssel verloren habe, sei es nämlich dunkel. Im Hinblick auf die Experimentalpsychologie — und nicht nur hierauf — enthält dieses Bonmot sicher ein Körnchen Wahrheit. So stellt der fehlende Vergleich der Aufsatz- und Rechtschreibleistung im Kontext des Methodeneffektes sicherlich ein (vermeidbares) Manko dar. Die von *Schmalohr* in bezug auf den Aufsatzunterricht und die Erfassung seiner Resultate vorgebrachten Argumente sind u. E. nicht stichhaltig. Die Anwendung der verschiedenen Formen der Content-Analyse, wie sie besonders seit dem letzten Weltkrieg vorab im angelsächsischen Raume ausgebaut worden sind, würde sowohl in inhaltlicher als auch in formaler Hinsicht quantifizierbare Vergleiche erlauben. Bei künftigen Ansätzen zur Erfassung des Lehrmethodeffektes sollte deshalb u. a. die Content-Analyse Berücksichtigung finden. Man darf also andererseits bei einer gerechten Beurteilung der Voraussetzungen und Möglichkeiten experimentellen Vorgehens nicht übersehen, daß gerade in jüngster Zeit erhebliche Fortschritte in methodischer und verfahrenswissenschaftlicher Hinsicht erzielt werden konnten, so daß heute ein sehr differenziertes und vielseitig anwendbares Instrumentarium auch zur Untersuchung des erörterten Lehrmethodeffektes zur Verfügung steht. Der Empiriker, sofern er operational die Phänomene resp. Phänomenezüge in den Griff zu bekommen bemüht ist, wird in vielen anhängigen Entscheidungssituationen dem ausschließlich geisteswissenschaftlich orientierten Kollegen gegenüber im Vorteil sein; dies gilt insonderheit für die hier diskutierte Problemlösung und ihren engeren und weiteren Umkreis schulpädagogischer und -psychologischer Fragekomplexe.

Die oben skizzierten Hypothesen wurden überwiegend durch psychologische Testverfahren geprüft. Dabei stützte sich *Schmalohr*²⁵ neben den beiden Frankfurter Tests „Verständiges Lesen“ und „Rechtschreibung“ vorwiegend auf Verfahren, die in der psychologischen Berufseignungsdiagnostik Verwendung fanden bzw. teilweise noch in Ge-

brauch sind²⁶. Zur Wahl der arbeitspsychologischen Eignungstests seien einige kritische Einwände erlaubt: Über die Gütekriterien (Objektivität, Reliabilität, Validität) dieser Techniken ist trotz vielfach behaupteter (oder auch tatsächlicher) Bewährung fast nichts bekannt; diesbezügliche statistische Kennwerte fehlen, so daß man eigentlich nicht von Tests im Sinne von vollstandardisierten Meßverfahren sprechen kann. Diese Unsicherheit des Diagnostikums (in bezug auf fünf von insgesamt sieben benutzten Erhebungstechniken) erfordert erhebliche Einschränkungen bei einer angemessenen Interpretation des Gesamtbefundes der *Schmalohrschen* Untersuchung, zumindest soweit es die Resultate der monierten „Tests“ betrifft. Demgegenüber benutzte *Müller* eine Batterie vollstandardisierter Verfahren mit nachweislichen Gütekriterien²⁷. Eine Ausnahme bildet der mündliche Lesetest von *Walter*, der erhebliche Mängel enthält und auch prompt schiefe Verteilung aufwies; mit den von *Walter* angegebenen Normen konnte keine Übereinstimmung erzielt werden. Auch die Fehlerklassifizierung ist sehr anfechtbar; zur Veranschaulichung genüge ein einziges Beispiel: „Wenn die Probanden ein Wort in der Wiederholung verbesserten, galt nur der beim ersten Lesen gemachte Fehler, eine Wortwiederholung wurde nicht mehr notiert. Die Fehlerkategorisierung berücksichtigt nicht die interessante Feststellung, inwieweit die Kinder ein falsch gelesenes Wort aufgrund einer Sinnsteuerung während des Lesens verbesserten.“²⁸ Auf die einseitige Zusammenstellung der *Müllerschen* Testbatterie haben wir früher schon aufmerksam gemacht. Resümierend stellen wir somit fest, daß die Erhebungstechniken weder des *Schmalohrschen* noch des *Müllerschen* Diagnostikums in ausreichendem Maße das breite Universum möglicher Methodeneffekte repräsentieren. Ihre Ergebnisse können bestenfalls Teilausschnitte betreffen, was für die Interpretation der Arbeitsbefunde Konsequenzen beinhaltet, wie wir gleich noch sehen werden. Zuvor möchten wir jedoch noch einen Vorschlag unterbreiten, der das Dilemma der Auswahl geeigneter und gültiger Untersu-

²⁵ *E. Schmalohr*, a. a. O., S. 34 f.

²⁶ Die von *Schmalohr* verwendeten Eignungsdiagnostika 3, 4 und 6 finden sich in: *A. Huth*, Handbuch psychologischer Eignungsuntersuchungen, Speyer 1953. — Außerdem kamen Test 5 (Druckfehlerlest von *Wartegg*) und Test 7 (*Bourdon*-Test) neben den oben genannten Frankfurter Tests Nr. 1 und 2 zur Anwendung.

²⁷ *H. Müller*, a. a. O., S. 76 ff. — Als Intelligenztests kamen die Progressive Matrices von *Raven* und (für das 4. Schuljahr) der Hamburg-West-Yorkshire-Gruppenintelligenztest zur Anwendung. Außerdem wurden 3 Leseleistungstests benutzt: „Prüfung im mündlichen Lesen“ von *Walter*, „The Chicago Reading Tests“ (in Übersetzung) und der bereits erwähnte Frankfurter Test „Verständiges Lesen“.

²⁸ *H. Müller*, a. a. O., S. 81.

chungsverfahren vielleicht beseitigen könnte, wir meinen den Einsatz der Diskriminanzanalyse in der anstehenden Methodenfrage. Die Diskriminanzanalyse, ein kompliziertes statistisches Verfahren zur optimalen Trennung heterogener Gruppen, sollte zumindest einmal ansatzweise bei zwei kleinen Methodenvergleichsgruppen erprobt werden. Es ist denkbar, daß auf diesem Wege ein „universales“ und zuverlässiges Diagnostikum zur Erfassung möglicher Lehrmethodeeffekte gewonnen werden könnte, ohne sich in die Gefahr der oben angedeuteten Fehlermöglichkeiten zu begeben. Jedenfalls konnten wir in einem ersten Versuch — weitere sind geplant — im Rahmen der Begabungs-

reservenermittlung (z. B. Trennung zwischen Volks- und Oberschulbegabungen) sehr ermutigende Erfolge beobachten. Das Verfahren der Diskriminanzanalyse ist rechnerisch sehr aufwendig und erfordert modernste (IBM- oder Siemens-) Rechenanlagen.

3. Die Untersuchungsbefunde und ihre Interpretation

Die auf 15 Seiten in zahlreichen Tabellen von *Schmalohr*²⁹ vorgelegten Ergebnisse seiner vergleichenden Leistungsmessung lassen erkennen, daß die Ganzheitler insgesamt siebenmal und die Synthetiker nur einmal signifikante Vorteile für sich ausweisen konnten. Folgende Zusammenstellung soll die Übersicht erleichtern³⁰:

Tabelle 6

a) Fremdsprachzensur der männl. Ob.:	G+	s	(X)
b) Verständiges Lesen der weibl. Ob.:	G+	s	
c) Verständiges Lesen der männl. und weibl. Ob.:	G+	s	(X)
d) Verständiges Lesen der männl. und weibl. Vo. und Ob.:	G+	s	(X)
e) Rechtschreibung der männl. Vo.:	s+	s	(X)
f) Listenvergleichen der weibl. Ob.:	G+	s	(X)
g) 50 beste Leistungen im Lesetest 1:	G+	s	
h) 50 schlechteste Leistungen im Bourdon-Test:	G+	s	31

Alle übrigen Differenzen waren statistisch nicht zu sichern. Wie interpretierte nun *Schmalohr* diese Befunde?

„Faßt man die Ergebnisse der vergleichenden Leistungsmessung unter diesen Gesichtspunkten (Hypothesen 1—9; d. Verf.) zusammen, so läßt sich sagen, daß die Untersuchung für keine der Hypothesen, die für die Gesamtheit (Hypothesen 1—7) und die besser begabten (Hypothese 8) der nach den beiden Lehrweisen unterrichteten Kinder aufgestellt worden sind, eine Bestätigung liefert. Im Gegenteil, alle diese Behauptungen der Pädagogen über unterschiedliche Auswirkungen der Lehrverfahren, wie wir sie angetroffen haben, sind durchgängig abzulehnen . . . Der einzige signifikante Unterschied ergab sich bei der Prüfung der Hypothese 9 . . . In ihrer allgemeinen Fassung, die Ganzheitsmethode benachteiligt später die

schwächer begabten Schüler, ist diese Hypothese auf jeden Fall abzulehnen. Der erwähnte Unterschied spricht ja auch gerade im Gegenteil für Vorteile der Ganzheitsmethode bei diesen Schülern. Man kann also Befürchtungen, die Ganzheitsmethode schädige die Arbeitshaltung dieser Kinder, als unbegründet zurückweisen. Was nun die Tendenzen angeht, sprechen die Mittelwertunterschiede (t-Statistik) und die Richtungen der Differenzen (KS-Statistik) für geringfügige Vorteile der Ganzheitsmethode im Lesen der Kinder (Hypothese 6; nur Test 1, nicht Test 3) und ihrer Arbeitshaltung (entgegen Hypothese 2; nur Test 7; nicht Test 3) und für Vorteile des lautsynthetischen Unterrichts im Rechtschreiben (entgegen Hypothese 7; nur in Test 2, nicht Test 5). Für auch nur eben merkliche Vorzüge oder Nachteile einer Lehrmethode im allgemeinen Schul-

²⁹ E. Schmalohr, a. a. O., S. 41—55.

³⁰ G+ = signif. Vorteile für ganzh. Schüler;

s+ = signif. Vorteile für synthetisch unterrichtete Schüler; s = signifikantes Ergebnis (auf dem 5%-Niveau gesichert); ss = sehr signifikantes Ergebnis (auf dem 1%-Niveau gesichert);

Vo. = Volksschüler (5. Schuljahr);

Ob. = Oberschüler (Sextaner);

(X) = die Signifikanzen sind nach dem t-Test gesichert, da jedoch in den vorliegenden Fällen die Normalverteilung (via Chi²-Test) nicht zu belegen war oder heterogene Varianzen vorlagen, wurde zusätzlich der (ungenauere) Kolmogorov-Smirnov-Test — eine nonparametrische Teststatistik —

durchgeführt, der die t-Signifikanzen in den bezeichneten Fällen nicht bestätigte. Das heißt, die t-Signifikanzen sind in diesen Fällen unscharf und nur unter gewissen Vorbehalten zu interpretieren; da sie jedoch zumindest sehr auffällige Tendenzen in den untersuchten Leistungsgruppen anzeigen, sind sie in keinem Falle wertlos.

³¹ Die Herausnahme der Ob.-Mädchen ist u. E. nicht legitim und stellt eine unerlaubte Manipulation der Untersuchungsbefunde dar. *Schmalohr* begründet seine Manipulation mit der vermuteten (aber unbewiesenen) Hypothese, daß Lehrmethoden keine differentiellen Einflüsse auf die Geschlechter ausüben würden. Vgl. S. 54.

erfolg der Kinder, im Deutsch- und Fremdsprachenunterricht der Oberschule oder in den analytischen Fähigkeiten bieten die Resultate keine Anhaltspunkte.“³²

„Danach können die Befürchtungen der Eltern und mancher Lehrer, die neue ganzheitliche Lehrweise im Erstlese- und Schreibunterricht bringe für die Schüler Nachteile hinsichtlich ihrer Schulleistungen mit sich, mit einigem Nachdruck als unbegründet bezeichnet werden. Als ebenso unbegründet erwiesen sich aber auch umgekehrt die Versprechungen der Ganzheitsmethodiker, die nach ihren Anweisungen unterrichteten Kinder hätten gegenüber den lautsynthetisch unterwiesenen unbezweifelbare Vorteile im späteren Schulerfolg . . . Zusammenfassend läßt sich demnach sagen, daß die beiden im pädagogischen Meinungsstreit diskutierten Lehrmethoden, so wie wir sie angetroffen haben, im Hinblick auf den späteren Schulerfolg der Kinder erfolgreich angewendet werden können.“³³

Dieses magere Resumé berücksichtigt u. E. das vorliegende Ergebnis nicht in angemessener Weise. Weder die gesicherten Teilresultate noch die deutlichen Trends wurden genügend gewürdigt. Ob hierbei nicht doch — noch stärker ist der Verdacht bei Müller — ein „bias“ die Interpretation beeinflusst hat? Der Leser mag sich nach unseren bisherigen Erörterungen selbst ein Urteil hierüber bilden. In bezug auf unseren früheren Einwand gegenüber einigen eignungsdiagnostischen Erhebungstechniken sind möglicherweise Bedenken hinsichtlich der Resultate f) und h)

angebracht, doch wird ihre Problematik von *Schmalohr* nicht gesehen resp. in seiner Würdigung der Untersuchungsbefunde ausgeklammert. Unter Berücksichtigung der Kritik sowohl gegenüber der Validität einzelner Testverfahren im *Schmalohrschen* Diagnostikum als auch der angezweifelten Isomorphie dieser Zugangsmodi zum anstehenden Untersuchungsgegenstand könnte man als Quintessenz des *Schmalohrschen* Feldexperimentes operational formulieren: Die Lehrmethodenvergleichsgruppen erbrachten teilweise keine, teilweise jedoch erhebliche Leistungsunterschiede — überwiegend zugunsten der ganzheitlich unterrichteten Schüler — in bezug auf die Dimensionen, die die verwendeten Erhebungstechniken bezielen. Dies wäre wohl die „exakteste“ Aussage. Ob die hier angesprochenen Leistungsdimensionen freilich zentral die aufgestellten Hypothesen bzw. die Unterrichtswirklichkeit treffen, müßte u. E. noch einmal genau überprüft werden. Eine solche Kontrolle sollte jedoch nicht unmöglich sein.

Auf die Müllerschen Arbeitsbefunde und ihre Kritik sind wir bereits bei der Stichprobenanalyse und der ausführlichen Behandlung der Untersuchungsmethoden eingegangen, so daß wir gleich im Anschluß unsere eigenen Ergebnisse bezüglich Rechtschreibleistung in Ganzheits- und synthetischen Klassen berichten können³⁴.

Die Übersicht läßt erkennen, daß im 2. Schuljahr die Gs-Pbn sowohl den Gw-

Tabelle 7

2. Schuljahr:

Rechtschreib. nach Testdiktat „Till Eulenspiegel“:		
Gs Knaben / sy Knaben	Gs +	ss
Gs Mädch. / sy Mädch.	Gs +	ss
Gs Kn + Md / sy Kn + Md	Gs +	ss
Gs Knaben / Gw Knaben	Gs +	ss
Gs Mädch. / Gw Mädch.	Gs +	ss
Gs Kn + Md / Gw Kn + Md	Gs +	ss
sy Knaben / Gw Knaben	Gw +	ss
sy Mädch. / Gw Mädch.	ns (nicht sign.)	
sy Kn + Md / Gw Kn + Md	Gw +	ss

3. Schuljahr:

Rechtschreib. nach Testdiktat „Till Eulenspiegel“:		
Gs Knaben / sy Knaben	sy +	s
Gs Mädch. / sy Mädch.	sy +	s
Gs Kn + Md / sy Kn + Md	sy +	ss
Gs Knaben / Gw Knaben		ns
Gs Mädch. / Gw Mädch.	Gw +	s
Gs Kn + Md / Gw Kn + Md	Gw +	s
sy Knaben / Gw Knaben		ns
sy Mädch. / Gw Mädch.		ns
sy Kn + Md / Gw Kn + Md		ns

als auch den sy-Pbn im Rechtschreiben sehr signifikant überlegen waren. Dieses Ergebnis gilt auch für die nach Geschlecht getrennten Untergruppen. Eben-

falls waren die Lehrmethoden Gruppen der Gw-Knaben und Gw-Knaben + Mädchen den sy-Pbn überlegen; in der Mädchen-Vergleichsgruppe

³² E. Schmalohr, a. a. O., S. 57.

³³ Ebd., S. 63.

³⁴ Dem aufmerksamen Leser fällt eine gewisse Diskrepanz zwischen den hier mitgeteilten Testleistungen und der Verteilung der LU im Rechtschreiben (vgl. Übersicht über die

Verteilung der LU-Variablen in unseren Lehrmethodenvergleichsgruppen) auf. Ähnliche gegenläufige Tendenzen stellte auch Schmalohr bei seiner Untersuchung fest (siehe S. 44). Solche Phänomene sind zwar merkwürdig, sie sind jedoch als Gegebenheiten zu akzeptieren.

konnten die Unterschiede statistisch nicht gesichert werden. Dieser Befund — in der Gs-Stichprobe des 3. Schuljahrs konnten verschiedene, bereits erörterte, relevante Variablen zu den Kontrollgruppen nicht gleichgehalten werden, weshalb ihre Resultate in diesem Zusammenhang wertlos sind — steht in gewissem Widerspruch zu dem *Schmalohrschen* Ergebnis, wonach die synthetische Knaben-Volksschulgruppe (einzig) im Rechtschreiben signifikante Leistungsvorteile aufwies. Allerdings liegen nur Tendenzen (ausschließlich bei Vo.-Knaben) hier vor, und die wegen fehlender Normalverteilung (Prüfung durch χ^2 -Test) angesetzte KS-Statistik konnte diese t-Signifikanz nicht bestätigen. Somit verdienen die von uns ermittelten, statistisch auf dem 1%-Niveau gesicherten Leistungsvorteile der Gs- und teilweise auch der Gw-Pbn des 2. Schuljahrs im Rechtschreiben eindeutig Vertrauensvorrang gegenüber den von *Schmalohr* erfaßten Tendenzen zugunsten der synthetischen Knabengruppe des 5. Volksschuljahrs. Ob unsere Leistungsbefunde auch auf großstädtische Verhältnisse übertragbar sind, können wir im Blick auf die soziogeographische Komponente unserer Stichprobe nicht mit Sicherheit entscheiden. Andere Untersuchungen, deren Resultate zum Teil in der „Ganzheitsschule“ (u. a. von *A. Kern* und *W. Straub*) mitgeteilt worden sind, lassen die Vermutung in diese Richtung eine gewisse Wahrscheinlichkeit gewinnen.

Schlußbetrachtung

An Hand empirischer Untersuchungen zur Erfassung des Lehrmethodeneffektes konnten wir einige relevante Problemkreise — vieles wurde nur flüchtig gestreift oder überhaupt nicht angesprochen —, die sich im Zusammen-

hang solcher Vorhaben ergeben, diskutieren. Wenn bislang offenbar noch keine befriedigende Approximation zwischen den wissenschaftsmethodischen Möglichkeiten und dem hier thematisierten Untersuchungsgegenstand erreicht worden ist, so bedeutet dies nicht ohne weiteres prinzipielle Unmöglichkeit. Unsere Kritik, die keineswegs vorliegende empirische Ansätze „zerreißen“ wollte, galt vielmehr gerade dem Bemühen, die aufgetretenen Fehler zu erkennen, um daraus zu lernen, d. h. letztlich das Ideal einer isomorphen Relation von Empirie und Modell der Verwirklichung ein Stück näherzubringen. In diesem Sinne sind die erörterten Arbeiten in jedem Falle „verdienstvoll“. Man täusche sich nicht im Glauben an eine eventuelle Überlegenheit geisteswissenschaftlicher Voraussetzungen zur Klärung der anstehenden Lehrmethodenentscheidung! Behauptungen, auch wenn sie syntaktisch gesehen wahr und semantisch eindeutig vorgebracht werden, brauchen deshalb noch nicht die Wirklichkeit zu treffen, wie das Beispiel der Sophisterei zur Genüge veranschaulicht. Nicht die Logik entscheidet darüber, was empirisch richtig ist, sondern die Erfahrung selber, die syntaktisch und semantisch gefaßt ist³⁵. An unbewiesenen Behauptungen und Aussagen mangelt es wahrlich nicht, vielmehr gilt es, den Überhang solcher Hypothesen abzubauen und Beweise für ihre Veri- oder Falsifikation zu erbringen. Genau das ist das Anliegen empirischer Wissenschaftsforschung. Daß hierzu sorgfältige und mühevollere Vorbereitungen nötig sind, dürfte aus unseren Überlegungen deutlich geworden sein. Ebenso sollte aber auch Klarheit darüber herrschen, daß von den gegenwärtigen wissenschaftsmethodischen Voraussetzungen und Möglichkeiten her keine unüberwindbaren Barrieren auf dem Wege zum gesteckten Ziele hin bestehen.