

wurden, dürfen aber wiederum nicht dazu führen, daß unabhängig von Art und Umfang der Befunde und des jeweiligen Kontextes die Interpretation erfolgt. Das Beispiel seiner Bewertung im Bereich der Erziehungswirkungen für die Hauptschulen zeigt, daß der Kritiker der Gefahr einseitiger Interpretation, die er den Untersuchungen glaubt vorhalten zu müssen, hier wie auch an anderen nachweisbaren Stellen selbst unterliegt. Andererseits wird nicht bestritten, daß die vorgelegten Befunde von Seiten ihrer statistischen Kennwerte (Mittelwerte, Häufigkeiten etc.) auch die Möglichkeit zu Interpretationen und Wertungen geben, die von denen der Berichtsautoren abweichen. Hier spielen sicher aus bevorzugten Theoriekonzepten und Forschungsschwerpunkten resultierende Evidenzkriterien eine Rolle. Allerdings sollte dann hierfür gegenseitige Offenheit bestehen und nicht mit Unterstellungen und Diskreditierungen gearbeitet werden.

Ohne Zweifel besteht eine Strategie der Kritik darin, die Maßstäbe hoch anzusetzen, um dann den zur Diskussion anstehenden Gegenstand - aus welchen berechtigten, weniger gerechtfertigten Gründen oder möglicherweise außerwissenschaftlichen Motiven auch immer - um so leichter und schärfer kritisieren zu können. Sachgeleitete Kritik würde sich u.a. ebenso darin zeigen, daß sie nem Untersuchungsansatz auch dann gerecht zu werden versuchen, und sich bemüht zu erkennen, was trotz dessen möglicher Begrenzung, Einseitigkeiten und methodischer Probleme mit seiner Hilfe erarbeitet und zutage gefördert werden konnte.

#### LITERATUR

- AURIN, K., SCHWARZ, B. & THIEL, R.-D. (1986a). *Gegliedertes Schulsystem und Gesamtschule. Vergleichsuntersuchung des Landes Baden-Württemberg*, (Band I: Forschungskonzept und Methoden). Freudenstadt-Grüntal: Verlag & Druck GmbH.
- AURIN, K., SCHWARZ, B. & THIEL, R.D. (1986b). *Schulsystem und Gesamtschule. Vergleichsuntersuchung Landes Baden-Württemberg*, (Band II: Unterrichtserfolge und Förderwirkungen). Freudenstadt-Grüntal: Verlag & Druck GmbH.
- AURIN, K. (1986). Verzerrte Darstellung der baden-württembergischen Vergleichsuntersuchung - eine Richtigstellung. *Gymnasium in Baden-Württemberg*, 10, 4-10.
- CAMPBELL, D.T. & STANLEY, J.C. (1970). Experimentelle und quasiexperimentelle Anordnungen in der Unterrichtsforschung (deutsche Bearbeitung von E. Schwarz). In K. INGENKAMP & E. PAREY (Hrsg.), *Handbuch der Unterrichtsforschung* (Teil I, S. 451-631). Weinheim: Beltz.
- HAENISCH, H. & LUKESCH, H. (1980). *Ist die Gesamtschule besser?* München: Urban & Schwarzenberg.

- LUKESCH, H. (1986). *Die Schwierigkeiten der systemvergleichenden Schulforschung. Eine Kritik der baden-württembergischen Gesamtschulstudie*. Ludwigsburg: Süddeutscher Pädagogischer Verlag.
- SCHNEEWIND, K. & LUKESCH, H. (1978). *Familiäre Sozialisation*. Stuttgart: Kohlhammer.
- SPD-LANDTAGSFRAKTION BAD.-WÜRTT. (Hrsg.). (1986). *Gesamtschule in Baden-Württemberg, Mitteilungen zur Landespolitik in Baden-Württemberg*. Protokoll der Anhörung, (S. 1-25).
- STOLZ, G.E. (1987). Längsschnittanalysen im Schulvergleich. In K. Aurin (Hrsg.), *Schulvergleich in der Diskussion*. Stuttgart: Klett.

#### Helmut Lukesch

### Schulsystemevaluation - ein hoffnungsloses Unterfangen? Zur Entgegnung Kurt Aurins zu vorgebrachten Mängeln an der baden-württembergischen Vergleichsuntersuchung<sup>1</sup>

Die Bildungskommission des DEUTSCHEN BILDUNGSRATS (1969, S. 141ff) hatte ein breit angelegtes Experimentalprogramm angeregt, von dem die Entscheidungsgrundlagen, eventuell auch ein empiriegeleiteter Konsens über das Für und Wider Integrierter Gesamtschulen erwartet wurden. Die angeregten Vergleichsstudien hatten Möglichkeiten für empirisch arbeitende Pädagogen und Sozialwissenschaftler geschaffen, wie sie nie zuvor gegeben waren und wie sie vermutlich für die absehbare Zukunft nicht mehr kommen wer-

<sup>1</sup>Der Verfasser wurde im März 1986 von der GEW Baden-Württemberg und der Max-Träger-Stiftung gebeten, eine kritische Würdigung der von einer Projektgruppe an der Universität Freiburg und dem Landesinstitut für Erziehung und Unterricht in Stuttgart unter der Leitung von Prof. Dr. Kurt Aurin erstellten Vergleichsstudie vorzunehmen, die am 7.3.1986 der Presse vorgestellt worden ist (AURIN et al., o.J.) und später publiziert wurde (AURIN et al., 1986a, 1986b). Voraussetzung für die Übernahme des Auftrages war die Zusicherung, daß das zu erarbeitende Gutachten ohne Zusätze oder Kürzungen der Öffentlichkeit zugänglich gemacht wird. Für die Drucklegung des Gutachtens (LUKESCH, 1986) sei den anfangs genannten Institutionen nachdrücklich gedankt.

den. Die Hoffnung auf konsensfähige Ergebnisse über vorhandene bildungspolitische Standpunkte hinweg hat allerdings getrogen, kontrovers blieb nicht nur die Gewichtung der Ziel- und Vergleichskriterien, kritisiert wurde auch die Qualität vorgelegter Ergebnisse. Ein herausragendes Beispiel dafür ist und bleibt die baden-württembergische Vergleichsuntersuchung, die um es vorweg zu nehmen - trotz siebenjähriger Projektlaufzeit keine verlässlichen empirischen Evidenzen für die Bewertung der beiden Schulsysteme schaffen konnte. Dazu im einzelnen:

AURIN (1987, S. 4) meint: "es liege ein Mißverständnis vor, wenn LUKESCH (1986, S. 16, 123) behauptet, vom Untersuchungsauftrag sei 'nur die Hälfte' realisiert und dieser 'mangelhaft erfüllt' worden". Um ein Mißverständnis handelt es sich hierbei wohl nicht, sondern um eine Tatsachenfeststellung. Neben nicht realisierten Untersuchungsteilen (vgl. im einzelnen LUKESCH, 1986, S. 13ff), bleibt es Tatsache, daß die zugesagte Abiturientenstudie nicht ausgewertet wurde und die geplante Vernetzung von Leistungsdaten mit denen zu den Erziehungswirkungen nicht vorgelegt wurde.

Gravierender jedoch sind methodische Mängel: AURIN (1987, S. 5) gesteht für seine Untersuchung zu, daß Repräsentativität hinsichtlich des gegliederten Schulsystems in Baden-Württemberg nicht gewährleistet ist. Dies wurde von LUKESCH (1986, S. 24ff) im Detail nachgewiesen. Welche Aussagekraft will man aber einer Untersuchung zugestehen, von der der Projektleiter selbst sagt, ihre empirische Grundlage sei nicht für das gegliederte Schulwesen repräsentativ. Darüber hinaus wird die Kritik hinsichtlich der Auswahl der Schulen für das gegliederte Schulwesen durch die von Aurin vorgetragenen Argumente in keiner Weise relativiert; auch wenn z.B. zu einem späteren Zeitpunkt das "erweiterte Bildungsangebot" an allen Hauptschulen Baden-Württembergs eingeführt wurde, bleibt es Tatsache, daß zum Zeitpunkt der Untersuchung die einbezogenen Schulen Modellschulen waren. In diesem Zusammenhang ist nochmals auf die Art zu verweisen, wie die Schulen des gegliederten Schulsystems ausgewählt wurden (FEND, 1987, S. 14), nämlich über Empfehlungen der Oberschulämter; allein dies dequalifiziert die mit der Untersuchung angestrebten Vergleichsaussagen.

Neben der von Aurin selbst zugestandenem Aussage, daß die Stichprobe aus dem gegliederten Schulsystem dieses nicht repräsentiert, kommt hinzu, daß auch Vergleichbarkeit mit den integrierten Gesamtschulen ebenfalls nicht gewährleistet ist. Die als Ausweg gewählte Methode, Untergruppen miteinander zu vergleichen, könnte geeignet sein, diesen grundlegenden Fehler im nachhinein zu heilen, produziert aber weitere Probleme, die unter dem Stichwort der Vergleichsgruppenbildung früher dargelegt worden sind (LUKESCH, 1986, S. 35ff). Der Versuch, bei den Leistungsvergleichen die bestehenden Unvergleichbarkeiten kovarianzanalytisch in den Griff zu bekommen, war von vorne herein eine Hilfskonstruktion, die letzt-

endlich wegen der Datenausfälle in bezug auf die Stichprobenmerkmale zu gravierenden Interpretationsproblemen führen mußte.

Wenn AURIN (1987, S. 6) zum wiederholten Male behauptet, es sei den Schulen aus Überlastungs- oder Motivationsgründen freigestellt worden, die Leistungserhebungen als Klassenarbeit oder als anonyme Übungsarbeit durchzuführen, so ist darin ein Mangel an Durchführungsobjektivität der Studie zu sehen; dieser Aspekt ist vor allem deswegen anrühlich, weil in dem der Presse vorgestellten Bericht (AURIN et al., o.J., S. 6) gesagt wurde, "eine genaue Kenntnis darüber, wie die einzelnen Schulen die Möglichkeit nutzten, hatte die Projektgruppe jedoch nicht", in der gedruckten Version diese entlarvende Passage aber entsprechend *nachgebessert* wurde (AURIN et al., 1986b, S. 6f: "Zwar war der Projektgruppe bekannt, ob die Tests als Übungs- oder Klassenarbeit verwendet wurden ..").

Die kritisierten Datenausfälle sind, ausgehend von der ursprünglichen Stichprobe, offenkundig. Es wird zwar in AURIN et al. (1986b, S. 5) davon gesprochen, daß ab der 2. Haupttestphase "unterschiedliche Testmodelle" verwendet wurden (vgl. auch 1986b, S. 54; gemeint sind vermutlich Erhebungsmodelle), worin diese 'Modelle' aber genau bestanden, wird an keiner Stelle beschrieben. Da auf der anderen Seite herausgestellt wurde, auf welcher eine imposante Datenbasis sich die Untersuchung stütze (AURIN et al., 1986b, S. 2), wird vergessen gemacht, daß für einzelne Aussagen nur 5% der ursprünglichen (Teil-) Stichproben zur Verfügung standen. Einem genaueren und kritischen Leser mußte außerdem wieder der Formulierungsunterschied zwischen dem vorläufigen Bericht (AURIN et al., o.J., S. 6) und der gedruckten Version (AURIN et al., 1986 b, S. 6) auffallen. Und zwar wurde in die spätere Publikation eingefügt, es seien Ausfallanalysen gemacht worden, ohne dies aber zu belegen; dies muß bei so einem gravierenden Faktum mißtrauisch stimmen. Die von STOLZ (1987, S. 98ff) vorgelegte Ausfallanalyse ist verdienstvoll und geeignet, Sachfragen zu klären, es wird aber auch in dieser Darstellung eine gewisse Konfusität während der Erhebungsphase belegt.

Daß bei der Analyse der *Erziehungswirkungen* neben sachlich widersinnigen Vergleichen (Gesamtschule als eine Schulart neben den drei anderen) auch von der Sache her begründbare Auswertungen gemacht wurden, soll nicht bezweifelt werden. Betrachtet man jedoch die Ergebnisse aus dem Kurs-Schulart-Vergleich unvoreingenommen, so ist trotz der vielen nichtsignifikanten Ergebnisse der Vorteil auf Seiten der Gesamtschüler nicht in Abrede zu stellen (LUKESCH, 1986, S. 76f). Es werden dabei zusätzlich Schwachstellen im gegliederten Schulsystem nicht gesehen, wo sie offensichtlich sind (keine C-Kurs-Problematik, sondern eine Hauptschulproblematik, im Querschnittvergleich sind 15 Effekte zugunsten der C-Kurs-Schüler und 7 zugunsten der Hauptschüler zu verzeichnen, und dies trotz der für C-Kurs-Schüler ungünstigeren Bezugsgrup-

peneffekte; besonders ungünstige "Erziehungswirkungen" bei den Schülern des gegliederten Schulsystems auf dem mittleren und höheren Leistungsniveau, *Gymnasialproblematik*).

Was letztendlich den Schulleistungsvergleich betrifft, so ist das Vorhaben, niveaubezogene Erhebungsinstrumente zu konstruieren und eine Längsschnittstudie zu realisieren, zu würdigen. Die Probleme, die man sich dabei einhandelt, sind aber beträchtlich: Letztendlich konnte eben doch - schon aus methodischen Gründen - keine echte Längsschnittanalyse durchgeführt werden, es wurde die hohe Mobilität an Gesamtschulen unzureichend berücksichtigt, es bleiben auch differentielle Fragen des Systemvergleichs unbeantwortet (LUKESCH, 1986, S. 19f).

In der Zwischenzeit wurden die Ergebnisse des Leistungsvergleichs von Mitgliedern der Projektgruppe wesentlich vorsichtiger und bezogen auf schulsystemspezifische Gegebenheiten (z.B. höhere Quoten bei mittleren Bildungsgängen und stärkere Selektion nach unten in den C-Kursen) interpretiert (STOLZ, 1987), wobei auch diese Aussagen noch in Einzelfällen ergänzungswürdig sind (z.B. haben Hauptschüler die Möglichkeit, das Fach Englisch abzuwählen, sie erhalten dann in diesen Stunden Stütz- und Förderunterricht für andere Fächer (RECK et al., 1986b, S. 1), diese Schüler wurden aber naheliegenderweise nicht in den Leistungsvergleich im Fach Englisch einbezogen, was zu einer Benachteiligung der Gesamtschüler führt). Selbst wenn man sich auf den Leistungsbereich als das Kernstück einer Systemevaluation bezieht, sollte man Schuleffekte nicht außer acht lassen, und hierbei war es eben eine Gesamtschule, die in wesentlichen Bereichen die besten Leistungen zeigte (GESAMTSCHULE STUTTGART-NEUGEREUT, 1986).

Auf viele weitere Schwächen der baden-württembergischen Studie kann hier nicht eingegangen werden; da sie in der Replik AURINS (1987) nicht angesprochen werden, sind sie von der Projektgruppe offensichtlich akzeptiert worden. Damit kann man konstatieren, daß trotz der zwischenzeitlich vorgenommenen Relativierungen der Aussagekraft der baden-württembergischen Vergleichsstudie durch ihre Autoren - die bedauerlicherweise auf den wissenschaftsinternen Auseinandersetzungsprozeß beschränkt sind und nicht in den gegenüber einem breiten Publikum gegebenen Stellungnahmen und Interviews zum Ausdruck kommen -, gravierende Qualitätsmängel nicht wegzudiskutieren sind.

#### LITERATUR

AURIN, K., SCHWARZ, B. & THIEL, R.-D. und Mitarbeiter (o.J.): *Gegliedertes Schulsystem und Gesamtschule Unterrichtserfolge und Förderwirkungen* (übergeben an Pressevertreter am 7.3.1986). o.O.

AURIN, K., SCHWARZ, B. & THIEL, R.-D. (1986a). *Gegliedertes Schulsystem und Gesamtschule. Vergleichsuntersuchung des Landes Baden-Württemberg*, (Band I: Forschungskonzept und Methoden). Freudenstadt-Grüntal: Verlag & Druck GmbH.

AURIN, K., SCHWARZ, B. & THIEL, R.D. (1986b). *Gegliedertes Schulsystem und Gesamtschule. Vergleichsuntersuchung des Landes Baden-Württemberg*, (Band II: Unterrichtserfolge und Förderwirkungen). Freudenstadt-Grüntal: Verlag & Druck GmbH.

AURIN, K. (1987). Stellungnahme zu der an der baden-württembergischen Vergleichsuntersuchung von Helmut Lukesch vorgebrachten Kritik. *Empirische Pädagogik*, 1, 264-272.

DEUTSCHER BILDUNGS RAT. (1969). *Einrichtung von Schulversuchen mit Gesamtschulen. Empfehlungen der Bildungskommission*. Stuttgart: Klett.

FEND, H. (1986). Stellungnahme bei einer Anhörung der SPD-Landtagsfraktion am 4. Juli 1986 im Landtag. In SPD-LANDTAGSFRAKTION BADEN-WÜRTTEMBERG (Hrsg.). *Gesamtschule in Baden-Württemberg* (S. 11-18). Stuttgart: Eigenverlag.

GESAMTSCHULE STUTTGART-NEUGEREUT (4.4.1986). *Vergleichsuntersuchung Gesamtschulen-Regelschulen*. Stuttgart: Eigenverlag.

LUKESCH, H. (1986). *Die Schwierigkeiten der systemvergleichenden Schulforschung. Eine Kritik der baden-württembergischen Gesamtschulstudie*. Ludwigsburg: Süddeutscher Pädagogischer Verlag.

RECK, U., AURIN, K. & EHRINGER, M. (1986). Darstellung der untersuchten Schulen beider Schulsysteme. In K. AURIN, B. SCHWARZ & R.-D. THIEL. *Gegliedertes Schulsystem und Gesamtschule. Vergleichsuntersuchung des Landes Baden-Württemberg*, (Bd. II: Unterrichtserfolge und Förderwirkungen; Anhang S. 1-38). Freudenstadt-Grüntal: Verlag & Druck GmbH.

STOLZ, G.E. (1987). Längsschnittanalysen im Schulvergleich. In K. AURIN (Hrsg.), *Schulvergleich in der Diskussion* (S. 77 - 103). Stuttgart: Klett.