

Bodo von Borries (Hamburg)

**Vergleichendes Gutachten zu zwei empirischen
Studien über Kenntnisse und Einstellungen von
Jugendlichen zur DDR-Geschichte**

Hamburg, im Februar 2008

Inhalt

Abstract: Zusammenfassung in zehn Thesen	4
1. Stiftung zur Aufarbeitung der SED-Diktatur: "Schülerbefragung 2005 zur DDR-Geschichte" (Ulrich Arnswald 2006)	7
1.1. Stichprobe und Erhebung	7
1.2. Instrument und Auswertung	9
1.3. Befunde und Interpretation	15
1.4. Folgerungen und Fazit	19
2. Forschungsverbund SED-Staat: "Das DDR-Bild von Schülern in Berlin" (Monika Deutz-Schroeder und Klaus Schroeder 2007)	22
2.1. Stichprobe und Erhebung	23
2.2. Instrument und Auswertung	27
2.3. Befunde und Interpretation	32
2.4. Folgerungen und Fazit	34
3. Vergleich und Beurteilung	35
3.1. Einordnung in Hintergrundtheorien	35
3.2. Fehlende Einbeziehung des Forschungsstandes	38
3.3. Detailvergleich und Länderrangreihe	40
3.4. Schlussbemerkung zur "DDR-Aufarbeitung"	44
Erwähnte Literatur	45

Anhang: Beispiele früherer Befunde (v. Borries u.a. 1995, 1999)

"Geschichtsbewußtsein Jugendlicher in Ost- und Westdeutschland" (1992),

Monografie 1995	47
- Bewertungen der deutschen Teilstaaten (S. 80-88)	48
- System und Regionaleffekte (S. 351-359)	52
- DDR-Nostalgie und Neofaschismus (S. 404-407)	57
"Youth and History" (1994/95), Monografie "Jugend und Geschichte" (1999)	59
- Osten und Westen im innerdeutschen Kulturvergleich (S. 346-352)	60
- Grafiken (S. 361-362)	62
- Änderungen der Sicht von Nationalsozialismus und Deutscher Spaltung 1949/89 (zwischen 1992 und 1994)? (S. 323-327)	63

Abstract: Zusammenfassung in zehn Thesen

2006 hat Ulrich Arnswald im Auftrag der Stiftung zur Aufarbeitung der SED-Diktatur eine Studie "*Schülerbefragung 2005 zur DDR-Geschichte*" vorgelegt. 2007 folgte seitens des Forschungsverbundes SED-Staat eine Publikation "*Das DDR-Bild von Schülern in Berlin*" von Monika Deutz-Schroeder und Klaus Schroeder, die zwar nur eine Vorveröffentlichung für ein Bundesland darstellt, der aber rasch der - auf bereits erhobenen Material beruhende - Vergleich mit Bayern, NRW und Brandenburg folgen soll. In der Öffentlichkeit ist der Eindruck entstanden, beide Studien widersprüchen sich in den Befunden lebhaft, z.B. hinsichtlich der Stellung Berlins in einer Rangreihung der Bundesländer (bezüglich der Kenntnisse und Einstellungen seiner Jugendlichen zur DDR). Deswegen wurde eine sorgfältige Prüfung und eine detaillierte Bewertung in Auftrag gegeben. Das macht es nötig, zuvor beide Studien einzeln gründlich vorzustellen und zu analysieren, bevor sie abschließend verglichen werden. Das Ergebnis soll als "Abstract" kurz vorweggenommen und dann ausführlich begründet werden.

1. Vor einer Einschätzung der Verlässlichkeit, Gültigkeit und Reichweite der Studien muss ihre methodische Anlage im Detail geprüft werden. Politische Aufregung und mediale Kampagnen mit Warnungen vor angeblich gravierenden Fehlentwicklungen ("whistle blowing") sind unangemessen oder jedenfalls überzogen, wenn keine belastbaren empirischen Daten vorliegen bzw. vorgelegt werden können.
2. Beide Studien leiden unter elementaren methodischen Mängeln im sozialwissenschaftlichen Handwerk; sie sind z.B. nicht repräsentativ, betreffen keine klar definierte Grundgesamtheit (mehrere Altersjahrgänge bzw. Klassenstufen in wechselnder Stärke), besitzen keine sorgfältige Entfaltung der Fragestellung und keine zureichende Operationalisierung bei der Erstellung des Erhebungsinstruments (Fragebogen), keine klare Trennung der Analyseebenen, z.B. "historische Wahrnehmung", "historische Deutung" und "historische Orientierung" nach Rüsen (bzw. "Sachanalyse", "Sachurteil" und "Werturteil" nach Jeismann) oder auch nur "Kenntnisse" und "Einstellungen", außerdem keine elaborierte Auswertung (z.B. multivariate Analyse).
3. Das heißt nicht, dass in den Studien nicht wichtige - oder auch unerfreuliche - Teile der sozialen Wirklichkeit - wenngleich wenig perfekt - abgebildet würden. Viele - die meisten - Befunde sind allerdings nicht eigentlich neu. Frühere empirische Studien und differenzierte-

re Ergebnisse sind einfach nicht präsent; eine Vernetzung mit dem bisherigen Forschungsstand (z.B. "Shell-Studien", "IBM-Jugendstudien", "Geschichtsbewusstsein in Ost- und Westdeutschland", "Youth and history") findet nicht statt.

4. Die Schlussfolgerungen auf curriculare bzw. schulpolitische Forderungen sind daher ein unzulässiger Fehlschluss. Dabei werden zudem frühere Ergebnisse zu Einstellungen, Kenntnissen und Verarbeitungsweisen bei anderen - umstrittenen und neuralgischen - (Zeit-)Geschichtsthemen, z.B. "NS" und "Demokratie", schlicht nicht zur Kenntnis genommen. Zu diesen Problemen gibt es zahlreiche ältere und neuere empirische Studien, die alle im Kern eine eher bescheidene Einflussmöglichkeit des Geschichtsunterrichts belegen.
5. Es herrscht eine recht naive Vorstellung von einer Art "Nürnberger Trichter" beim Geschichtslernen, so als könnten die Geschichtslehrer(innen) einfach etwas in die Köpfe der Schüler(innen) hinein füllen, der Erfolg ließe sich dekretieren und - bei genügend heftiger Curriculumrevision, d.h. Dosiserhöhung - trete automatisch eine Besserung ein. *Lehren heißt aber nicht Lernen machen.*
6. Das (Zeit-)Geschichtsbewusstsein von Jugendlichen entsteht primär nicht durch Geschichtslehrer, Geschichtsschulbücher und Geschichtsunterricht, sondern in einem komplexen Prozess vor allem auch aus der lebensweltlichen Verarbeitung (und Fehlverarbeitung) im kommunikativen Gedächtnis, d.h. durch die Erwachsenen in Familie und Nachbarschaft. Dabei spielen im Falle der DDR - größtenteils negative - Gegenwartserfahrungen vieler Ostdeutscher nach 1990 und um-konstruierende Vergangenheitswahrnehmungen - zwecks Bewahrung von "Anerkennung" und Schutz der "biografischen Integrität" - eine entscheidende Rolle.
7. Starke Tendenzen zu einer gewissen "DDR-Nostalgie" bei Kindern und Jugendlichen konnten schon 1992 (nur drei Jahre nach der Wende!) und erneut 1994 (etwas abgeschwächt) gültig und zuverlässig gemessen werden (z.B. v. Borries u.a. 1995, 80ff., 351ff., 404ff., v. Borries u.a. 1999, 323ff., 346ff., 361f., vgl. auch IBM 1992, 78ff., 88ff., IBM 1995, 81ff., 99ff., 110ff., 115ff.). Die Unterschiede der Jugendlichen aus Ost- und Westdeutschland waren und sind - hinsichtlich der DDR-Geschichte - massiv; sie bilden aber nur die Unterschiede ab, die - das weiß doch jede(r) - auch ihre Eltern und Großeltern im Wahlverhalten seit 1990 an den Tag legen. Jede Vermutung, Geschichtsunterricht könne mit ein paar "Lektionen" diese elementare Spiegelung der Erwachsenenwelt bei den Jugendlichen einfach auslösen, ist unglaublich naiv.

8. Für einen exakten Ländervergleich der Kenntnisse und Einstellungen zur DDR-Geschichte (z.B. zwischen Brandenburg und Ost-Berlin, zwischen NRW, Bayern und West-Berlin) eignen sich die Studien beide nicht besonders gut, auch weil sie - vgl. These 2 - ihr Material ganz unvollkommen auswerten und präsentieren (kein exaktes Spezifizieren auf Altersgruppen, Geschlechter, Berliner Stadtteile usw.).
9. Das gilt auch für die angeblich widersprüchlichen Befunde beider Studien. Die - ohnehin nicht besonders validen (vgl. These 2) - Fragestellungen, d.h. Einzelitems der beiden Fragebogen, sind nur selten direkt vergleichbar. Die unzureichende Form der Veröffentlichung (z.B. fehlende Trennung der Befragten aus Ost- und Westberlin bei Arnswald 2006) tut ein Übriges; mit dem Urmaterial müsste - z.B. bei Beschränkung auf die 11. Klasse - mehr zu machen sein. Die heikle - auch wichtige? - Frage von Bundesländer-Rangreihen würde aber letztlich eine wirklich gültige und verlässliche neue Untersuchung voraussetzen.
10. Eine politische Forderung nach Vermehrung des Stoffes aus der DDR-Geschichte in den Lehrplänen oder nach Integration beider Teilstaaten-Geschichten zu einer gemeinsamen deutschen Nachkriegs-Geschichte lässt sich aus den Studien nicht verantwortbar herleiten. Für eine gemeinsame, interdependente Version der Geschichte beider Teilstaaten 1949/89 sprechen überragende andere (fachliche und pädagogische) Gründe. Gegen eine quantitative Aufwertung der DDR-Geschichte (die qualitative Bewertung muss ohnehin in jeder einzelnen Klasse "verhandelt" statt "dekretiert" werden!) gibt es, falls nicht die Stundenzahl des Faches "Geschichte" in der Stundentafel insgesamt erhöht wird, massive Gründe. In der gegenwärtigen Phase der "Globalisierung" müssen Provinzialismus und nationsbezogene Enge - und seien sie in bester Absicht hergestellt - verhindert werden.

1. Stiftung zur Aufarbeitung der SED-Diktatur: "*Schülerbefragung 2005 zur DDR-Geschichte*" (Ulrich Arnswald 2006)

1.1. Stichprobe und Erhebung

Im Jahre 2005 wurde mit 5.616 Jugendlichen aus 58 Schulen eine recht große Stichprobe befragt, zudem noch in allen Bundesländern. Es handelt sich aber ausschließlich um Schüler(innen) von Gymnasien. Das erweckt - wie leider oft - den Eindruck, als käme es den politisch und didaktisch Verantwortlichen - und den wissenschaftlichen Forschern - auf die größere Hälfte der Lernenden, die nicht die privilegierteste Schulform besuchen, überhaupt nicht mehr an. Hier liegt eine erste, aber schlechterdings entscheidende Schwäche der Studie (zumal im Folgenden stillschweigend meist von einer möglichen Hochrechnung auf alle Jugendlichen ausgegangen wird).

Die Ausklammerung von Haupt- und Realschule sowie vergleichbaren Schulformen (wohl auch Gesamtschule) lässt sich auch nicht damit rechtfertigen,¹ dass man unterstellt, entscheidend seien erst Geschichtsbewusstsein und Geschichtseinstellungen der schon beinahe Wahlberechtigten im *Alter der Sekundarstufe II*. Wenn man das vertreten würde, dann müsste man eben im beruflichen Schulwesen befragen, wie wir selbst das 1992 bei den Älteren selbstverständlich getan haben (vgl. v. Borries u.a. 1995, 17ff., 24ff). Es bleibt dabei: Entweder die Wissenschaftler um Arnswald haben den Aufwand gescheut, die - dort natürlich schwierigeren - Befragungen in anderen Schulformen (wie auch bei PISA üblich) durchzuziehen, - oder "Nicht-Gymnasialst(inn)en" werden nicht als vollwertige Bürger(innen) angesehen.

Die Stichprobe ist auf keine bestimmte Klassestufe festgelegt, sondern streut von der 9. bis zur 13. Klasse (S. 144), allerdings mit einem Übergewicht der 11. Klasse (fast 65%). Man fragt sich unwillkürlich, was hier die Logik ist, über die man nämlich nichts erfährt. Offenbar hat man auch hier - aus unbekanntem Gründen - den Weg der Bequemlichkeit statt einer strengen und nachvollziehbaren (auch replizierbaren) Methode gewählt.² Im ganzen wird man wohl so interpretieren müssen, als seien 11. Klassen befragt worden, doch das Resultat sei erheblich verschmutzt und verunklärt (lautes "Rauschen"), zumal ja - beim Ländervergleich - nicht gesichert

¹ Mit der Auslassung der Sonderschulen kann man sich - angesichts der Leseschwierigkeiten in vielen (nicht allen) Fällen dieser Schulart - einverstanden erklären.

² Vermutlich haben hier die Schulen selbst mitbestimmt; das wirft - falls zutreffend - das Problem systematischer Verzerrungen auf, d.h. die mögliche Wahl von als besonders leistungsstark geltenden Klassen ("Schokoladenklassen"), die die Repräsentativität verfälschen können.

und eher unwahrscheinlich ist, dass die Abweichungen von der 11. Klasse in allen Bundesländern parallel vorkommen, was (S. 169) auch tatsächlich nicht der Fall ist.

Die Schulen (nicht aber die Klassen, was ein zusätzliches Klumpungsrisiko bedeutet!) sind zufällig gezogen (S. 111), aber offenbar nicht in einer nach Ländern "geschichteten" Stichprobe. Sonst könnten nicht aus Baden-Württemberg mehr Befragte stammen als aus dem weit größeren Bayern (und aus dem noch viel kleineren Niedersachsen fast ebenso viele). Über die Zahl der Verweigerungen durch Schulleiter, Schulkonferenz oder Eltern (das Genehmigungsverfahren scheint in diesem Falle korrekt angewandt worden zu sein, vgl. S. 111) und den Umgang damit wird nichts mitgeteilt; es ist aber ganz unplausibel, dass es nirgends Ausfälle gegeben hat (wurde dann nachgezogen?).

Die Abweichungen von den prozentualen Anteilen Gleichaltriger sind beträchtlich (so hat Sachsen-Anhalt viel mehr Befragte als das wesentlich volkreichere Sachsen, Hamburg - fast dreimal so groß wie Bremen - bleibt der Schwesterstadt gegenüber in der Teilnehmer(innen)zahl zurück, S. 114f.). Man wird das als eine unperfekte (aber noch erträgliche) an Zufall angelehnte Stichprobenziehung bezeichnen können,³ aber - auch abgesehen von der Beschränkung auf Gymnasien - den ausdrücklichen Anspruch auf "*eine repräsentative Befragung in allen deutschen Bundesländern, die die Schüler der Gymnasien ab 9. Klasse sowie der höheren Klassen (in Grund- oder Leistungskursen) erfassen soll*" (S. 108), deutlich zurückweisen müssen.⁴

Ein Kommentar ist auch zur Befragungsdurchführung selbst nötig. Die Schulen bekamen die Fragebögen zugeschickt, ließen sie selbst ausfüllen und sandten sie mit vorfrankierten Briefumschlägen zurück; über die Rücklaufquote und die Vollständigkeitsquote (fehlende Schüler) verlautet nichts. Das lässt eine Menge Fragen offen: unerlaubte, aber gut gemeinte Lehrerhilfen, Perfektion der Anonymitätswahrung, kaum standardisierte Durchführungsbedingungen. Natürlich sind geschulte und einheitlich verfahrenende Testleiter(innen) bei weitem vorzuziehen (so die

³ Die ungleichmäßige Erfassung ist dann nicht dramatisch, wenn sich die Länderunterschiede als relativ belanglos erweisen sollten (dann sind sie freilich in einem Ländervergleich auch nicht sicher abbildbar bzw. nicht statistisch "signifikant"), so dass man auf Rangreihen verzichten muss. Die Konfidenzintervalle werden eben groß; damit liegen die Signifikanzschwellen hoch. Da es für eine optimale Stichprobe übrigens nicht primär auf die Größe, sondern auf die Heterogenität der Grundgesamtheiten ankommt, hätte man auch in jedem Bundesland - für den Ländervergleich - gleich viele Probanden befragen und - für den Bundesdurchschnitt - die Länderwerte anteilig gewichten können.

⁴ So war es "*vorgesehen*" (S. 108); die Realisierung für die Klasse 9 ist bei 0,1% aller Befragten, die eine 9. Klasse besuchen (so S. 114), ein Witz. Denn das wären bei 5.616 Befragten 5 bis 6 Personen; so kleine Klassen sind ausgesprochen selten; es liegt also nicht einmal eine intakte Klasse vor - oder ein Eingabe- bzw. Rechenfehler. Ein Sample, das zu 0,1% aus Neuntklässlern und zu 64,9% aus Elftklässlern besteht, ist jedenfalls in hohem Maße "*schief verteilt*" und eben nicht "*repräsentativ*" für 9. bis 11. Klasse. Das sollte im DIPF in Frankfurt bekannt sein?

eigenen Befragungen von 1992 und 1994, vgl. v. Borries u.a. 1995, 1999), allerdings teurer. Auch bei PISA sind geschulte externe Testleiter(innen) eingesetzt worden.

Zwei Klassenkamerad(inn)en werden sich - wenn denn Unterricht überhaupt etwas bewirkt - in Geschichtskennntnissen und -einsichten ähnlicher sein als zwei beliebige Lernende der Klassenstufe. Das bedeutet *Klumpungseffekte* bei Befragung in *intakten Klassen* und umgekehrt dadurch eine geringere "effektive" Stichprobengröße (das Phänomen kann als "Designeffekt" statistisch kontrolliert werden und ist keineswegs immer gleich groß). Bei Schulen als Grundeinheiten (statt Klassen) verschärft sich das Problem noch; in ganz Sachsen z.B. wurden nur zwei Gymnasien mit fast 300 Schüler(inne)n berücksichtigt, in Niedersachsen fünf mit fast 550. Die relevanten Unterschiede zwischen einer traditionsreichen Eliteschule, einem Vorstadtgymnasium in einer Villengegend und einem ums Überleben kämpfenden Gymnasium in einem "sozialen Brennpunkt" reduzieren die Aussagemöglichkeiten drastisch (über eine Kontrolle dieser Effekte verlautet nichts).

Nimmt man die zuvor erwähnten Methoden-Unschärfen (Klassenstufe, Befragungssituation) hinzu, dann kann man - auch ohne Detailrechnung - ziemlich sicher feststellen, dass Länderunterschiede zwischen einzelnen West-Ländern oder zwischen einzelnen Ost-Ländern mit dem Datensatz nicht seriös untersucht werden können. Vermutlich sind aber - und darauf beschränkt sich ja auch der Autor (S. 158-166) - Ost-West-Differenzen mindestens wahrscheinlich zu machen, vielleicht als "signifikant" nachzuweisen;⁵ dazu fehlen freilich in der Veröffentlichung - trotz der generellen Behauptung (S. 162) - alle Detailangaben.⁶

1.2. Instrument und Auswertung

Der Fragebogen wurde von einer Expertengruppe beraten, in der merkwürdigerweise die Geschichtsdidaktik nicht vertreten gewesen ist (S. 110). Das hat gravierende Folgen, vor allem bei den Aufgabeninhalten (s.u.). Doch vorher sind noch einige Bemerkungen zur Aufgabenform nötig: Es handelt sich um einen ausnahmslos "geschlossenen" Fragebogen, d.h. die Lernenden

⁵ Es ist daran zu erinnern, dass die Stichprobe selbst nicht zufällig ist und nach Klassenstufen erheblich "schlackert". Signifikanzformeln (z.B.: $p < 0.01$, S. 162) setzen jedoch stets Zufallstichproben aus eindeutig definierten Grundgesamtheiten voraus. Streng genommen muss es also heißen: "Die tatsächlich untersuchten, u.U. 'bias-behafteten' Gruppen unterscheiden sich 'signifikant' (aber wie stark?); ob Unterschiede beider Ländergruppen, die durch die Befragten repräsentiert werden sollen, 'signifikant' wären, kann man nicht wissen. Es wäre der Fall, wenn in einer 'repräsentativen Zufallsstichprobe' dieselbe Mittelwertdifferenz gefunden würde wie in der verwendeten (unperfekten) 'Konvenienzstichprobe'."

⁶ Eine ehrliche Signifikanzprüfung täte gut daran, nur die Daten der 11. Klassenstufe zu verwenden und die Klassenklumpung und Schulklumpung ("Designeffekte") ausdrücklich zu berechnen.

mussten nur ankreuzen, nicht selbst schreiben.⁷ Das macht die Handhabung leicht und auch die Auswertung ökonomisch und "objektiv" (vom Auswerter unabhängig), freilich lassen sich höhere kognitive und evaluative Leistungen damit nur sehr schwer erfassen bzw. abbilden.⁸ Dazu wird auch keinerlei ernsthafter Versuch gemacht.

Als Frageform sind meist vierstufige Likertskalen gewählt, d.h. Antworten von "*trifft gar nicht zu*" über "*trifft eher nicht zu*" und "*trifft eher zu*" bis "*trifft voll zu*", und das bei einer zusätzlichen Möglichkeit der einfachen Auslassung (in der Auswertung heißt das dann "*keine Angabe*"), aber ohne Chance von "*teils - teils*" oder "*kann mich nicht entscheiden*". Dieses Vorgehen, das die Befragten zu einer Stellungnahme zwingen möchte und die Ausweichmöglichkeit minimiert, ist möglich und wird oft verwendet, da es die Varianz künstlich erhöht.⁹

Neben den Likert-Skalen werden auch Zuordnungsaufgaben, z.B. die ein-eindeutige Verbindung von Ereignissen und Jahreszahlen (S. 129, 134) oder von Namen und politischen Funktionen (S. 125), eingesetzt. Auch das ist eine übliche und zulässige Form, bei der man allerdings - besonders wenn beide Listen gleich lang sind - berücksichtigen sollte, dass sie ziemlich leicht ist (hohe Ratewahrscheinlichkeit durch Ausschlussverfahren) und insofern auch vages Ahnen meist zu "richtigen" Antworten macht. Dazu kommen vereinzelte ganz normale "multiple-choice-Aufgaben", so zum Autor eines Zitates oder zum biografischen Zeitpunkt der Jugendweihe (S. 118, 126). Damit sind - unbeschadet der Verbreitung und Zulässigkeit dieser Form - natürlich komplexere Leistungen kaum zu erfassen.

Viel wichtiger als die Aufgabenformen sind - wie oben angedeutet - die Aufgabeninhalte bzw. die verlangten geistigen Operationen. Eine wirklich detaillierte Ausführung zu den Zielen fehlt; die Formel lautet ziemlich knapp: "*Erkenntnisse zum realen Wissensstand der Schüler über die DDR-Geschichte zu gewinnen und an ausgewählten Fragestellungen Haltungen zur SED-Diktatur zu ermitteln.*" (S. 108)" Wenige Zeilen darauf wird noch einmal von "*Wissen und Einstellungen*" gesprochen.¹⁰ Dem Anspruch nach sollen die dann formulierten Aufgaben auf

⁷ Als Ausnahmen waren das Datum des Kriegsendes und die Parteien, die in der SED aufgingen, aufzuschreiben (S. 116).

⁸ Das kann aus langer und leidvoller eigener Erfahrung sicher gesagt werden (vgl. v. Borries u.a. 1992, 1995, 1999, 2005).

⁹ Persönlich lehne ich diese Frageform aber ab und setze stets Likertskalen mit ungerader Zahl (meist fünfstufig) ein. Denn im Leben kann man sich - außer als Prüfer und als Richter - fast immer enthalten, schon gar bei skalierbaren Fragen. Nur bei reinen Ja-Nein-Entscheidungen sieht es manchmal anders aus ("ein bisschen verheiratet" oder "ein bisschen schwanger" gibt es nicht; deshalb kann man sich da auch wirklich nicht enthalten).

¹⁰ Es ist bezeichnend, dass es Arnswald ausdrücklich - und selbstverständlich - um "*die auf Wissen gegründeten Einstellungen, Meinungen und Haltungen der Schüler*" geht (S. 108); auf die Idee eines umgekehrten Zusam-

die Curricula abgestimmt sein (S. 109), was bei sechzehn Bundesländern bereits eine Quadratur des Zirkels darstellt; das Ergebnis wird denn auch nicht im einzelnen erläutert und begründet, sondern gesetzt und enthält fünf curriculare Hauptpunkte ("*1. Politischer Neubeginn, 2. Das Machtsystem, 3. Gesellschaftsgeschichte, 4. Widerstand und Opposition, 5. Friedliche Revolution*", S. 109),¹¹ zu denen jeweils ganz wenige *Wissensaufgaben* und zahlreichere *Einstellungsaufgaben* gefunden werden.

Alles kommt also auf die sprachliche und didaktische Qualität des Fragebogens an: Was wird warum - in welcher Verständlichkeit und Genauigkeit - erfragt? Erfahrungsgemäß schleichen sich auch bei gründlichster Kontrollarbeit und sorgfältiger Pilotierung (eine solche wurde unternommen) einzelne schwache Items ein. Das gilt wohl für Frage 24: "*Worin unterschied sich die Freizeitgestaltung von Jugendlichen in West und Ost?*" Das Item 4 "*durch die Ausrichtung auf westliche Kultureinflüsse*" ist - genau betrachtet - logisch unklar und - wenn man erst einmal aufmerksam wird - auch historisch ambivalent: Waren DDR-Jugendliche nicht auf Fernsehen, Mode, Musik des Westens vergleichbar "heiß" wie BRD-Jugendliche selbst? Auch das Item "*durch die ideologische Ausrichtung*" ist unklar, weil Tatsache, Intensität und Richtung von "ideologischer Ausrichtung" unterschieden werden müß(t)en. Oder will jemand ernsthaft behaupten, im Westen habe es - mitten im "Kalten Krieg" - keine ideologische Ausrichtung gegeben? Beispiele dieser Art ließen sich - bei einer Detailanalyse - häufen.

Wichtiger aber ist der unklare Status der meisten Fragen.¹² Sie sind einerseits Einstellungsfragen, andererseits haben sie auch Anteile von Kenntnisfragen, dazu Anteile von Geschichtsdeutungen und auch Anteile von Orientierungsentscheidungen. Diese drei - wenn auch interdependenten - Ebenen unterscheiden zu können, ist eine wichtige Aufgabe des Erwerbs historischer Kompetenz.¹³ In den gestellten Fragen gehen die Teiloperationen historischen Denkens jedoch leider durcheinander. Statt dessen wird ja nur zwischen "Kenntnissen" und "Einstellungen"

menhangs, eines durch Einstellungen, Meinungen und Haltungen gesteuerten "Kenntnisstandes" kommt er gar nicht.

¹¹ Abgeleitet sind diese fünf Themenkomplexe allerdings nicht aus den Richtlinien, sondern aus der vorangegangenen Schulbuchanalyse (S. 110).

¹² Grundsätzlich ist zu fragen: Ist das Gefragte "Wissen"? Wenn ja, "Wissen welcher Art"? Es geht um *Deutungen* oder *Sachurteile*. Da müsste eigentlich formuliert werden: "*Bitte gib an, für wie zutreffend du die folgenden Charakterisierungen des Unterschieds hältst!*"

¹³ Das Modell ist nicht ganz neu oder neumodisch; es findet sich als drei Ebenen des historischen Denkens bei Jeismann ("Sach[verhalts]analyse", "Sachurteil", "Bewertung", vgl. Jeismann 1985, 2000) und Rösen ("historische Wahrnehmung", "historische Deutung" und "historische Orientierung", vgl. Rösen 1994b/b) seit den siebziger/achtziger Jahren und ist im Verbundprojekt "FUER Geschichtsbewusstsein" ausführlich reformuliert und entfaltet worden (vgl. Schreiber 2003, Schreiber/Körber 2006, Körber u.a. 2007). Die Unterscheidung ist also keine Quisquilie einzelner weltfremder Theoretiker, sondern eine Grundfigur des Geschichtslernens.

unterschieden, was sich aber ebenfalls als unscharfe Trennung erweist.

Das wirkt sich dann auf die Auswertung aus, wo einerseits von "zutreffenden" und "der realen Situation entsprechenden" Antworten der Schüler die Rede ist, andererseits aber *verschiedene zulässige* "Deutungen" und auch "Einstellungen" oder "Einschätzungen" gar nicht erwähnt werden. *Sind auch Gesinnungen schlicht als richtig oder falsch zu bewerten?* Der Geschichtsbegriff, die Erkenntnislogik der Historie selbst, ist bei der Testerstellung nicht zureichend durchdacht worden. Zwar ist es lobenswert, dass verschiedene Antworten nebeneinander als richtig angekreuzt und anerkannt werden können (Mehrdimensionalität, Multikausalität), aber eine tiefgründige Abbildung von "Multiperspektivität", "Kontroversität" und "Pluralität" der Historie wird dadurch noch keineswegs erreicht. Auch die Forderung danach gilt - seit Jeismann und Rüsen - in der Geschichtsdidaktik als Selbstverständlichkeit.¹⁴

Hinzu kommt - und das ist ein weiterer ernster Mangel -, dass die Items fast durchgehend nur hoch triftige, noch akzeptable oder jedenfalls diskutabile (wenigstens ehemals diskutierte) Deutungen aufzuführen trachten, nicht aber hoch ideologische, absurde oder bösertige Thesen. Es gibt also keine "Känguruhs", keine "Nonsensantworten", durch die u.a. die Ernsthaftigkeit der Beantwortung geklärt werden könnte. Auch verbreitete Legenden und Mythen, die in großen Teilen der Bevölkerung präsent sind, werden - vielleicht als eines "wissenschaftsorientierten" Geschichtsunterrichts unwürdig - kaum berücksichtigt. Auch Gymnasiasten hängen aber vielfach absurden oder untriftigen Positionen an - und diese sollten diagnostiziert werden.

Wenigstens ein Beispiel soll gebracht werden: Frage 28 lautet: *"Wie ging die Staatsmacht der DDR gegen die oppositionellen Kräfte vor?"* Vier Wege werden angeboten und sind einzuschätzen (Bespitzelung, berufliche Repression, Privatsphären-Eingriff, Auslandsreisen-Verbot). Alle vier treffen, wenn auch vielleicht in etwas unterschiedlichem Maße, zu.¹⁵ Solche Items erzeugen geringe Varianz und wenig interpretierbare Information;¹⁶ den Jugendlichen wird nur einfach zugestanden, sie hätten *"das Vorgehen der Staatsmacht gegen oppositionelle Kräfte (...) sehr realitätsnah bewertet"* (S. 151). Schöner kann man die - analytisch unklare - Vermischung von Kenntnissen und Einstellungen nicht ausdrücken (aber *ohne Vorsicht bei Folgerungen!*).

¹⁴ Wenn man das ernst nimmt, steht die Gültigkeit der Studie - und damit auch ihre Brauchbarkeit für Schlussfolgerungen - in Frage. Darf man - aufgrund solcher Items - überhaupt von "zu wenig" oder "unzureichendem" Wissen sprechen?

¹⁵ Es ist also kein Zufall, dass die Datenausfälle stets weit unter 10% bleiben und Ablehnung meist nur bei weniger als 20% erfolgt (Ausnahme "berufliche Repressionen" mit 25%).

¹⁶ Schärfer gesagt: *Eigene Wertungen* sind bei diesen Items kaum zu belegen; "Fehlbewertungen" im Sinne der Autoren wären allenfalls als Ablehnung der Items möglich.

Wie wäre es mit weiteren Items, z.B. *"Bestechung und Abwerbung durch versprochene Privilegien"*, *"Festhalten in geheimen Arbeitslagern ohne Gerichtsurteil"* oder *"Abordnung in die Sowjetunion zwecks Kennenlernen des 'Mutterlands aller Werktätigen'"*, auch *"Folter und Psychiatrieeinweisung"*?¹⁷ Beim *"Führungsanspruch der SED"* (Frage 13) liegt das Problem ganz ähnlich. Alle vier Bereiche (*"Staat"*, *"Bildung"/"Kultur"*, *"Wirtschaft"* und *"innere Sicherheit"/"Landesverteidigung"*) sind offenkundig mit hohen Werten anzukreuzen, was die diagnostische Qualität gravierend mindert. Warum werden nicht z.B. auch *"christliche Kirchen"*, *"Wohnblocks und Nachbarschaften"* sowie *"Sportvereine"* genannt?

Dabei soll nicht bestritten werden, dass in anderen Bereichen gelegentlich auch wenig triftige oder irrtümliche Items eingeführt sind, so bei Frage 37: *"Bei den Demonstrationen der DDR-Bürger im Herbst 1989 wandelte sich die mitgeführte Losung 'Wir sind das Volk' in die Losung 'Wir sind ein Volk'. - Was war der Grund für diese Veränderung?"* Da ist Item 2 *"der Verzicht auf Reformen"* kein Nonsense, aber wenig passend; die beiden anderen *"das Ziel der Wiedervereinigung"* und *"eine Absage an die Eigenstaatlichkeit der DDR"* dürften als plausibel gelten. Aber wie steht es mit weiteren Ideen wie *"nationalkonservative Provokation"*, *"der Wunsch nach DM-Einführung und Wirtschaftsstabilisierung"*, *"Wunsch nach einem neu gestalteten Deutschland jenseits von 'Beitritt/'Anschluss'"* und *"politischer Appell an und moralischer Druck auf die Siegermächte des Zweiten Weltkriegs"*? Das sind natürlich nur - noch nicht ausgereifte - Beispiele zur Verdeutlichung des Gemeinten.

Insgesamt ist der Fragebogen verhältnismäßig eintönig und wenig anstachelnd, wohl auch mit eher heißer Nadel gestrickt. Dadurch kommt er vielleicht manchen Curricula nahe, was aber erst noch einmal im Detail zu belegen wäre. Jedenfalls bewegt er sich ziemlich weit weg von den lebensweltlichen Interessen - und auch den Mythen und Legenden - der Befragten. Wie in der Interpretation *"Kenntnisse"*, *"Operationen"*, *"Deutungskritik und Deutungsannahme"* sowie *"Orientierung"* unterschieden werden sollen, wird nicht recht klar. Eine theoretische Konzeption und eine ausgewiesene Operationalisierung liegen dem Entwurf wohl (nähere Auskünfte, wie erwähnt, fehlen) nicht zugrunde. Statt dessen gibt es offenbar eine - ebenfalls nur vage erwähnte (S. 109) - Vorstellung von einem angeblich idealen Stoffplan.

Die Behauptung, der Fragebogen lehne sich an die drei Anforderungsbereiche der EPA an

¹⁷ Es muss zugegeben werden, dass die Formulierung gar nicht leicht ist, zumal die DDR fast alle Mittel typischer Diktaturen, bis hin zur geheimen Liquidation und zur ebenfalls weitgehend geheimen Todesstrafe (teilweise gefälschte Totenscheine), eingesetzt hat. Eine einigermaßen genaue Analyse müsste wohl auch - likertskaliert? - nach der Häufigkeit fragen.

(S. 109) und sei auf den "*Ausprägungsgrad von Sach-, Methoden und Urteilskompetenz*" ausgerichtet (S. 108), kann nicht nachvollzogen werden, sondern bleibt ein völlig uneingelöster bloßer Anspruch. Es geht durchweg allenfalls um das Wiedererkennen mehr oder weniger geläufiger "Fakten", Anspielungen, Deutungen und Urteile, denen man dann - notfalls auch bloß ratend - mehr oder weniger zustimmen kann.¹⁸ Eine eigene Argumentation, wie sie ja auf Anforderungsebene II und III absolut zwingend ist, wird nirgends verlangt. Sie ließe sich übrigens auch nicht ohne weiteres durch "geschlossene" Aufgabenformen (d.h. bloßes Ankreuzen und Zahlen-Eintragen) abtesten.¹⁹ Das geschichtsdidaktische Reflexionsniveau des Fragebogens ist zweifellos - um es zurückhaltend zu sagen - bescheiden.

Das Buch gibt an, die Befragung sei - nach Eingabe der Daten - mit SPSS gerechnet und ausgewertet worden. Der Publikation - sowie der beigelegten CD-ROM, die nur die Fragebögen enthält - ist aber nicht zu entnehmen, dass irgendwelche komplexeren Analyseschritte durchgeführt worden wären.²⁰ Leider enthält die CD-ROM auch nicht - wie heute durchaus üblich (vgl. z.B. Angvik/v. Borries 1997) - den gesamten Datensatz zu eigenen Analysen seitens der Benutzer(innen). Das ist angesichts der oberflächlichen und wenig professionellen Auswertung besonders dramatisch.

Tatsächlich wird ausschließlich über Prozentwerte der Gesamtpopulation (nicht einmal geschlechtsspezifisch oder klassenstufenspezifisch aufgegliedert!) berichtet (S. 144-157).²¹ Dabei werden sogar noch jeweils die Angaben "*trifft voll zu*" und "*trifft eher zu*" einerseits und "*trifft gar nicht zu*" und "*trifft eher nicht zu*" andererseits zusammengezählt, also erneut wesentliche

¹⁸ Wie später noch einmal den Fragen 31 (Verhältnis Kirche - Opposition), 33 (Arten von oppositionellen Gruppen in der Endphase der DDR), 35 (wesentliche Ursachen für die friedliche Revolution in der DDR) und 37 (Hintergrund für den Losungswechsel von "*Wir sind das Volk*" zu "*Wir sind ein Volk*") bestätigt werden kann, dass "*sie vertieftes analytisches Denken erfordern*" (S. 156), ist schleierhaft. In allen Fällen wird nicht die Analyse abgefragt, sondern das bloße Wiederauftauchen früher gehörter Assoziationen (Reiz-Reaktions-Verbindungen).

¹⁹ Es ist durchaus nicht völlig unmöglich, damit auch auf höhere kognitive Niveaus zu kommen. Aber dann muss man z.B. Materialien (Zitate, Karikaturen, Bilder) auswerten lassen, indem man z.B. Kurzinterpretationen zu ihnen als "trifftig", "fraglich" oder "abseitig" einschätzen lässt. Das bedeutet dann relativ viel Lese- und Denkarbeit.

²⁰ Selbst (gruppenspezifische) Mittelwerte auf den Skalen der Einzelitems fehlen, Standardabweichungen werden nicht angegeben.

²¹ Gleichsam anhanghaft werden die Befunde zu den Wissensfragen nach zwei Klassenstufen-Gruppen (9.-11. und 12./13.) aufgegliedert, wobei eine durchgehende, aber eher geringe Überlegenheit (Erhöhung des Anteils der richtigen Lösungen im Mittel um 5%) der höheren Klassen gefunden wird (S. 166f.). Ergänzend wird - ebenfalls für die Wissensitems - der Zusammenhang mit "*Geschichte als Lieblingsfach*" vorgeführt (ebenfalls leichte Überlegenheit der "Geschichtsfans", S. 168f.). Dabei wird allerdings die Kausalität sträflich einseitig behauptet: "*Das Bekenntnis zum Lieblingsfach Geschichte zeigt somit konkrete Folgerungen* [sic! Es muss logisch wohl "Folgen" heißen]." (S. 169), obwohl der Zusammenhang doch auch umgekehrt sein kann: Wissen spornt Motivation an. Schließlich gibt es einen - unzureichenden - Vergleich der Länder und der Klassenstufenanteile. Demnach sind die befragten Berliner deutlich jünger, die befragten Ostdeutschen deutlich älter als die befragten Westdeutschen (S. 169).

Informationen verschenkt, die freilich weiter vorn (S. 113-143) schon global (ohne Aufgliederung auf Untergruppen) angegeben sind.²² Das geht eigentlich nicht über die Strichlisten hinaus, die man vor Jahrzehnten als Lehrer von Hand führte. In einem späteren Abschnitt (S. 158ff.) werden listenhaft die Items zusammengestellt, in denen sich nennenswerte Unterschiede der ost- und der westdeutschen Befragten ergeben (Berlin wird da, ohne die vermutlich im Urmaterial mögliche Trennung von Ost- und Westteil, gesondert ausgeworfen).

Alle Analysen verbleiben, jedenfalls in der publizierten Fassung, auf der Ebene einzelner Items. Skalen, Faktoren, multiple Regressionen, Korrelationen werden nicht erstellt (von Clusteranalysen zu schweigen). Die Möglichkeiten zu einer multivariaten Analyse, die die eigentliche Stärke des Programmpakets SPSS ausmachen, werden also in keiner Weise ausgeschöpft, was natürlich durchaus kein Selbstzweck ist, aber zu einer besseren Beantwortung von wichtigen Fragen, z.B. nach stabilen Mustern jenseits bloßer äußerer Auffälligkeiten bei einzelnen Items und für strukturelle Analysen sowie die Einschätzung von Effektstärken nötig wäre.

Also: Die oberflächlich-beschreibende Ebene wird an keiner Stelle verlassen. Das ist auf alle Fälle ein großer Verlust, fällt aber wegen der bescheidenen Theoriegrundlagen und Operationalisierungskünste noch besonders ins Gewicht. Hier können gar keine Hypothesen geprüft werden, denn es wurden keine aufgestellt. Die Interpretation des - einigermaßen eintönigen und unklaren - Materials würde nun alle Tricks verlangen, um im Nachhinein noch möglichst viele Zusammenhänge herauszupressen, aus denen dann doch vielleicht noch auf Kausalitäten zu schließen wäre: Korrelationen sind zwar noch keine Kausalitäten, können aber, wenn theoretische Anhaltspunkte dazu kommen, zu ihnen hinführen.

1.3. Befunde und Interpretation

Die Feststellung, eine aufwendige empirische Untersuchung sei in allen ihren Schritten methodisch nicht erstklassig, sondern bleibe erheblich unter dem "state of the art", macht die Studie noch nicht wertlos. Tatsächlich gibt es im Bereich der empirischen Geschichtsdidaktik so wenig belastbares Material, dass man auch mit einem "Fingerhut voll Erkenntnis" zufrieden sein muss, wie Friedrich Lucas (1972, 9ff.) das genannt hat. Es ist übrigens auch viel leichter, perfekte Studien zu fordern, als sie wirklich vorzulegen.²³ Im Reich der Empirie muss es ganz viele und

²² Da die Studie auch vom "Deutschen Institut für Internationale Pädagogische Forschung" in Frankfurt (DIPF) mitverantwortet wird, ist die bescheidene Qualität der Fragebogenkonstruktion (z.B. bei der Operationalisierung) ebenso erstaunlich wie die rein deskriptive, bloße Prozentsätze auswerfende Auswertungsmethode.

²³ Aufgrund von über 15 Jahren Praxis in empirischer Bildungsforschung weiß ich, wovon ich rede.

ganz verschiedene Wohnungen geben. Unter Umständen sind die Befunde so drastisch und eindeutig, dass mangelnde Differenzierung und Genauigkeit - vorläufig? - verschmerzt werden können. Gerade wer solche Studien - *faute de mieux* - benutzen will, muss aber ihre (begrenzte) Belastbarkeit gründlich prüfen und ihre (uneingestanden) Defizite deutlich benennen.

Einige der Fragen werden von Arnswald ausdrücklich als Kenntnisfragen bezeichnet und können auch als beispielhafte Hinweise akzeptiert werden.²⁴ Das genaue Datum des Kriegsendes (8. Mai 1945)²⁵ ist 45,0% der befragten Schüler(innen) bekannt, die beiden Vorgängerparteien der SED 48,5%, Ulbricht als Autor des Zitates "*Es muss demokratisch aussehen*" 14,2%. Die Zuordnung von Personen und Funktionen (z.B. Ulbricht = "*langjähriger SED-Chef...*") gelingt im Mittel weniger als 30%, das gilt auch für die Jahre und die Erhebungsversuche (DDR, Ungarn, CSSR, Polen), während die richtige Zuordnung der Daten im Verlauf der "Wende" 1989/90 durchschnittlich knapp 40% fertigbringen.²⁶

Das kann man - so der Autor - sehr unbefriedigend ("*Defizite*", "*schlechter Stand*", "*Schülerwissen sehr gering*", "*sehr bedenklich*", so S. 145, 151, 153) finden. Man muss dann eigentlich nur sagen, an welchem ausgewiesenen Maßstab man denn misst und wie man methodisch überzeugend den Schwellenwert (heute sagt man "Mindeststandard" oder "Regelstandard") festgelegt hat (gab es z.B. eine Vergleichsbefragung unter Erwachsenen, z.B. Akademikern?). Das bleibt stets höchst willkürlich, zumal ja niemand zwischen dem Wissen um den *genauen Tag der bedingungslosen Kapitulation*²⁷ und heutigem Engagement für Frieden, Gerechtigkeit, Demokratie und Menschenrechte eine überzeugende direkte Beziehung herstellen kann. Was ist schlimm daran, wenn man nur das Jahr 1945 oder den Monat Mai 1945 kennt?

Das Ulbricht-Zitat könnte - strukturell - auch von Stalin stammen, was - völlig vernünftig -

²⁴ Die besondere Wichtigkeit und curriculare Relevanz sind allerdings gesondert zu prüfen; sie werden in der Studie nur einfach behauptet.

²⁵ Im Westen wurde allerdings statt 8. Mai (Karlshorst) früher oft der 7. Mai (Reims) genannt, umso erstaunlicher sind die besseren Antworten im Westen gerade - und ausnahmsweise - bei diesem Item.

²⁶ Das sagt noch nichts oder wenig zur Relevanz: Man kann Zusammenhänge differenziert erfasst und abgewogen haben, aber sich im Datum irren - und man kann ein Datum korrekt wiedergeben, aber von Zusammenhängen wenig begriffen haben.

²⁷ Im Fragebogen heißt das: "*Welches Datum steht für das Ende des Zweiten Weltkrieges?*" Ist diese wolkige Formulierung für "Waffenstillstand" zwecks Vermeidung des Wortes "Kapitulation" - allerdings müßte zwingend "in Deutschland" ergänzt werden, denn der Weltkrieg endete ja keineswegs - wirklich besser als "*bedingungslose Kapitulation*" 1945? Bei der Frage "*Wie bewerten Sie das Ende des Zweiten Weltkrieges?*" sprechen - was die Fragebogenkonstrukteure freuen wird - deutlich mehr Lernende von "*Befreiung*", "*Chance für friedlichen Neuanfang*" und "*Beginn der Besatzung und Spaltung*" als von "*Kapitulation*". Wäre doch wenigstens überhaupt noch nach "*Zusammenbruch Deutschlands*" und "*Ende des 'Dritten Reiches' (Großdeutschlands)*" gefragt worden, um irgendwelche Interpretationschancen zu gewinnen und nicht nur jeweils "politisch korrekt" den neusten Stand der "verbindlichen Konvention" abzubilden, die hoffentlich auch genauso im Schulbuch steht!

fast dreimal so viele Befragte vermuten. Die Unkenntnis einiger Personen (muss man Chris Geoffroy und Wilhelm Pieck wirklich zwingend kennen?) und der Parallelfälle (Ungarn, CSSR, Polen) ist ärgerlich und spricht gegen vorhandene Übersicht; aber heute wird ja gerne - vielleicht manchmal übertrieben - mit exemplarischen Tiefenbohrungen gearbeitet. Die Einzeldaten von Herbst und Winter 1989/90 mögen spannend und erhellend sein; das bleibt aber eine Art Ereignisgeschichte. So etwas ist doch hoch willkürlich ausgewählt: An der "Gültigkeit" der Messung (Repräsentation der *wichtigen* Probleme zur Geschichte der DDR in diesen ausgewählten Items) ist entschieden zu zweifeln.

Später werden dann die richtigen Antworten aus Berlin, Alten Ländern und Neuen Ländern verglichen, mit dem Ergebnis, dass das - gemischte - Berlin durchweg am besten abschneidet und dass - meist - die Ostländer vor den Westländern liegen (S. 158f.). Das ist in der Tat auffallend, aber nicht sehr erstaunlich. In Berlin ist - wegen des relativ häufigen Zusammentreffens beider Seiten - vermutlich die Aufmerksamkeit am größten. Das Thema ist die DDR, eben die Vergangenheit des Ostens; bei Fragen zur BRD - mit Erhard wurde ja eine solche gestellt - lagen vermutlich die Westländer vorne, wie es ja bei Erhard auch der Fall ist.

Obwohl dieses Ergebnis plausibel wäre, ist es - hinsichtlich Berlins - methodisch durchaus nicht zuverlässig gesichert; denn in Berlin wurden 504 Schüler(innen) aus fünf Schulen befragt (was bereits ein erhebliches Oversampling von Berlin bedeutet!), aber wir wissen nicht, wie viele Eliteschulen oder Problemschulen, wie viele 9. Klassen oder 13. Klassen darunter waren.²⁸ Merkwürdigerweise fehlt sogar der Versuch einer Skalenbildung: Addition, Durchschnittsbildung und Konsistenz- bzw. statistische Qualitätsprüfung (Cronbachs Alpha) für ein gemeinsames Messinstrument "*Kenntnisse zur DDR-Geschichte*", das - mit immerhin 20 Items - ja u.U. eine hohe Zuverlässigkeit ("Reliabilität") - allerdings nicht Gültigkeit ("Validität", s.o.) - erreichen könnte. Als Tendenz ist die Hypothese einer Berliner Spitzenstellung interessant, als feste Behauptung wäre ihre Feststellung methodisch ungesichert.

Bei den meisten Fragen des Fragebogens aber sind Kenntnisse und Gesinnungen (oder wenigstens Deutungen) so eng miteinander vermengt, dass eine bloße Auswertung der Prozentsätze von (halber und ganzer) Zustimmung wenig Gewinn bringt. Das "*sowjetische Modell des Sozialismus, das in der DDR nachvollzogen wurde*" erhält folgende Einschätzungen: 73,8% sagen

²⁸ Beim Klassenstufenvergleich (S. 169) wird deutlich, dass in Berlin die 12./13. Klassen, die im Mittel (S. 166f.) auch höhere Kenntnisse hatten, deutlich stärker vertreten waren (fast 40%) als in den Ostländern (10%) und den Westländern (22%). Wie gesagt: Umstandslose Ländervergleiche verbieten sich eigentlich methodisch.

"(es) war von Anfang an zum Scheitern verurteilt", nur 37,7% "(es) war ein gutes Konzept, das schlecht ausgeführt wurde". 59,5% befinden: "(es) war eine Verfälschung der Ideen von Marx und Engels" und nur 36,2% "(es) war eigentlich nur für Rußland geeignet".

Was soll daraus eigentlich geschlossen werden? Hat die Mehrheit die richtigen Antworten gegeben und die Lernziele erreicht? Sind die Minderheiten²⁹ historisch schief gewickelt oder antidemokratisch oder politisch gefährlich? Angesichts heute üblicher, verabsolutierter Marktverherrlichung und Marktwirtschaftsgläubigkeit handelt es sich vielleicht um Oppositionelle ("demokratischer Sozialismus") oder Querdenker ("selbstständiges Urteil") oder Törichte ("ans Schulsystem Unangepasste"). Zu jeder der vier Aussagen müsste diskutiert, argumentiert, differenziert, abgewogen und erst dann begründet - und zwar in einer pluralistischen Gesellschaft u.U. je persönlich - entschieden werden. Genau das geschieht - in Befragung wie Auswertung - nicht einmal versuchsweise.

Es gibt nämlich überhaupt keine Interpretation, nur später - im Rahmen einer langen Liste innerhalb einer großen Tabelle - einen Vergleich von Berlin, Neuen Ländern und Alten Ländern. Im Osten stimmen fast 14% (58,8 - 72,7) weniger für "von Anfang an zum Scheitern verurteilt" und fast 17% (51,5 - 34,6) mehr für "gutes Konzept - schlecht ausgeführt" (S. 160f.). Dieser Unterschied ist in der Tat beachtlich. Zwar sind - aus oben dargelegten Gründen (Nicht-Repräsentativität, unklare Klassenstufen, Schul-Klumpungen, kleine Sample in kleinen Ländern) - Signifikanzschwellen nicht genau berechenbar (die Studie verspricht da zu viel, S. 160); aber höchst wahrscheinlich ist die Differenz triftig,³⁰ zumal sie frühere Befunde bestätigt, dass der Wunsch nach stärkerer Staatslenkung und sozialer Sicherheit in den neuen Bundesländern anhält.

Ein zweites Beispiel ist vergleichbar informativ: Der "tatsächliche Stellenwert der Gleichberechtigung der Frau" wird von 63,5% als "reine Propaganda", von 61,7 als "lediglich politisch gewollt" (eine unglückliche, unklare Formulierung!), von 50,9% als "ökonomisch erforderlich" und nur von 36,2% als "gesellschaftliche Realität" eingeschätzt (S. 150). Das ist zweifellos spannend, nur wäre gerade hier ein Vergleich mit der Einschätzung des gleichen Phänomens in

²⁹ Es werden übrigens angesichts der allgemeinen Bejahungstendenz vermutlich vielfach dieselben Personen sein, die auch die anderen "korrekten" Sätze bejaht haben; aber das könnte nur ein Korrelation oder Kreuztabulierung der Items zeigen.

³⁰ Dennoch: Was wird hier eigentlich gemessen? Sicher *Einstellung* bzw. ein Gemenge von Sach- und Werturteil. Aber was sagt eine Zustimmung aus, was eine Ablehnung? Besonders bei Falle letzterer: Bezieht sie sich auf das Sachurteil (gleiche "Idee") oder auf das Werturteil ("schlecht ausgeführt")? An einer gewissen "Doppelköpfigkeit" des Items ist kaum zu zweifeln.

der alten Bundesrepublik, wie wir selbst ihn 1992 und 1994 haben ziehen lassen (vgl. v. Borries u.a. 1995, 80ff., 1999, 323ff., 346ff.) noch wesentlich ertragreicher gewesen. Dann wäre wenigstens ein kleines Stück des Doppelcharakters von objektiv-messbarer *Wirklichkeitswahrnehmung* und subjektiv-gruppenspezifischer *Wirklichkeitswahrnehmung* entschlüsselbar geworden.³¹

Die verschiedene Zustimmungintensität im Osten und im Westen ist nicht erstaunlich; sie wurde - wenn auch viel differenzierter und für mehrere Altersgruppen - schon 1992 gefunden und 1994 abgeschwächt bestätigt (vgl. v. Borries u.a. 1995, 1999). Westdeutsche Lernende betonen bei der Gleichberechtigung in der DDR 2005 um fast 23% häufiger als ostdeutsche (74,0 - 51,3) die "*reine Propaganda*" und um fast 24% seltener (29,9 - 53,7) die "*gesellschaftliche Realität*". Beide Seiten identifizieren sich eben im Nachhinein ganz erheblich stärker mit ihrem Teilstaat. Das war auch vor 15 Jahren schon so, hat sich erstaunlich wenig abgemildert und hat mit historischen Kenntnissen wenig zu tun. Solche Wertungen werden - nach "*gefühlten*" und "*verarbeiteten*" Lebenserfahrungen - "aus dem Bauch" und "aus den Familien- und Nachbarschaftserzählungen" gezogen.

Gelegentlich streift auch Arnswald diese Erkenntnis, wenn er schreibt: "*Es ist zu vermuten, dass diese Entscheidungen aus Gründen der Zugehörigkeit zu diesem Gebiet [DDR bzw. BRD, B.v.B.], und nicht auf der Basis tieferer Sachkenntnisse zur Entnazifizierung gefallen sind.*" (S. 162f.) Dasselbe, also "*Identitätsausfluss*", könnte man bei den meisten (vielleicht allen) verschieden bewerteten Phänomenen, nicht nur bei der "Entnazifizierung" - und eben der "*Gleichberechtigung*" - sagen. Genau das ist auch theoretisch zu erwarten, da Geschichte ("Historie") eben kein einfaches Auftürmen oder Herunterbrechen von Tatsachen ist, sondern ein "Deutungsgeschäft". Dann aber muss es um die Art der persönlichen Verarbeitung gehen, und die wurde nicht angemessen abgetestet.

1.4. Folgerungen und Fazit

In der Zusammenfassung (S. 174-176) verweist der Autor korrekterweise (1.) auf das große geäußerte Interesse der Jugendlichen an DDR-Geschichte, (2.) auf die häufige Nennung von Nutzung geschichtskultureller Angebote außerhalb des Geschichtsunterrichts (in anderen Fächern und von anderen Agenturen) sowie (3.) auf die Bevorzugung einer "integrierten Version" der

³¹ Im Text (S. 150) gibt es keinen einzigen interpretierenden Satz, nur einen kurzen referierenden.

deutschen Geschichte 1945-1990. Der erste Punkt kann freilich erst im Vergleich mit anderen Themen, die nicht angeboten wurden, validiert werden;³² der zweite Punkt ist aus zahlreichen früheren Befragungen bekannt (bleibt aber sehr wichtig); und der dritte Punkt liegt nahe, bedarf aber der Ausgestaltung (früher war in vielen Bundesländern ein "Systemvergleich" vorgeschrieben).

Anderen Aussagen der Zusammenfassung kann man durchaus nicht zustimmen. Unerfindlich bleibt, warum der Autor seinen Fragebogen - bzw. dessen kurze und völlig un-reflexive und nicht-methodenorientierte Kenntnisteile - für eine einigermaßen valide Operationalisierung des wichtigen Wissens über die DDR-Geschichte hält. Vollends methodisch unerlaubt wird es dann, wenn der Autor aus der - übrigens recht relativen - Ähnlichkeit der Einstellungs-Antworten im Osten und im Westen auf "Annäherung" (so gleich zweimal S. 175 und 176, auch schon 166) und ein "Indiz des Zusammenwachsens" (S. 175) schließt.

"Annäherung" ist nicht "Nähe" oder "Ähnlichkeit", sondern - wie "Zusammenwachsen" - ein *Prozess*; man kann sie nur behaupten, wenn man zweimal gemessen hat (am besten zweimal mit dem gleichen Instrument). Genau das hat Arnswald (anders als wir 1992 und 1994, vgl. v. Borries u.a. 1995, 1999) nicht getan. Die Behauptung mag also richtig oder falsch sein (*im Verlaufe der Zeit kann man sich auch - z.B. bei gefühlter "Ungerechtigkeit" - auseinanderleben!*), ist aber methodisch schlicht verboten, also für einen Wissenschaftler erstaunlich. Jedenfalls müsste man mehr Vergleichsmaterial beibringen, im Detail argumentieren und eine intensive Interpretation (auch mit Eingehen auf andere Daten, z.B. die Shell- und IBM-Jugendstudien) vornehmen.

Um zu einem Fazit zu kommen: Arnswalds Studie hat durchaus eine Fragestellung: Geschichtswissen und Geschichtseinstellung deutscher (und bildungs-inländischer) Jugendlicher hinsichtlich der DDR-Geschichte. Diese Problematik könnte genauer präzisiert und sorgfältiger entfaltet werden; aber im Prinzip geht dieser Punkt in Ordnung. Die Hintergrundtheorie wird dagegen nicht recht klar. Hinsichtlich der DDR-Geschichte sind unaufgedeckte, aber feste Positionen eher wahrscheinlich zu machen als "Multiperspektivität", "Kontroversität" und "Pluralität". Über eine elaborierte Theorie des Geschichtslernens und des Geschichtsbewusstseins (ein-

³² Schon gar muss man sagen, was man denn - falls der Stundenanteil des Faches Geschichte nicht erhöht werden soll - streichen will. Dann hat man abzuwägen: Sollen wirklich weitere Teile der ohnehin arg gerupften und meist nicht mehr begriffenen "Vormoderne" (vor 1789) entfernt werden? Gibt es inzwischen ganz viel "Geschlechtergeschichte", die wieder hinausgeworfen werden kann? Darf im Zeitalter der (vierten oder fünften) "Globalisierung" der allerletzte Rest "außer-europäischer" Geschichte beseitigt werden? Kann man - nach dem Ende der "Nachkriegszeit" - guten Gewissens die Anteile für den NS kürzen? Vielleicht sind wir in gesteigerter Gefahr einer - schon etwas provinziell werdenden - "bloß *nationalen* Nabelschau" (wenn auch in guter Absicht).

schließlich Kompetenzansatz) verfügt die Studie nicht. Statt dessen hat sie eine feste Überzeugung von dem, was *wichtig* und was *richtig* ist, und zwar auf der - analytisch nicht unterschiedenen - Sachverhalts-, Sachurteils- und Werturteilsebene.

Die Studie hat ein großes Materialkorpus - in Gestalt eines Datensatzes nach einer Befragung von über 5000 Lernenden - selbst erzeugt. Aber Repräsentativität ist - auch abgesehen von der schwer zu rechtfertigenden Beschränkung auf das Gymnasien - nicht gewährleistet (z.B. schlackernde Klassenstufenanteile). Und Validität des Befragungsinstruments kann - auch mangels Operationalisierung von Theorie - nicht vorausgesetzt werden. Die Auswertungstechnik (SPSS-Analyse) ist zwar angegeben; aber die Auswertungsmethode bleibt schlicht ärmlich, bloß deskriptiv, nicht hypothesenprüfend. Damit werden alle Befunde zum "Nennwert" genommen, können aber weder auf ihre Reliabilität geprüft ("Methodeneffekte") noch überzeugend interpretiert (oft kein Wort) und kausal zugeschrieben werden.

Die Behauptung, dass Berliner Schüler mehr wissen als ostdeutsche und westdeutsche, ist - angesichts der Schwächen von Stichprobe, Instrument und Auswertung - methodisch nicht gut abgesichert, wenngleich ein bescheidener Vorsprung möglich, vielleicht sogar wahrscheinlich ist. Der Befund insgesamt ist - hinsichtlich der Kriterien (auch Schwellenwerte) für "Defizit" oder "Zufriedenheit" - methodisch unausgewiesen, hinsichtlich der explizit behaupteten "Annäherung" von Ost und West (als Prozess) eine glatte Erschleichung ohne Grundlage (keine Replikationsstudie).

Die didaktisch-politische Konsequenz, nämlich Erweiterung des Unterrichts über die DDR-Geschichte - aber integriert als gesamtdeutsche Geschichte - entspricht zwar den geäußerten Wünschen der Schüler (übrigens auch einer Debatte in der Zeithistorie), lässt sich aber aus den Befunden zu Wissen und Einstellungen nicht sauber ableiten (bei starker "Annäherung" in Richtung *innere Wiedervereinigung, die ja behauptet wird*, könnte doch Unterricht über die DDR eher "überflüssig" werden, eben "erledigt" sein!) und muss mit anderen möglichen Konsequenzen abgeglichen werden. Zwischen der empirischen und der normativen Ebene ist eine überzeugende Verbindung nicht hergestellt - und so nicht herstellbar.

Die zugrunde liegende Vermutung (und das geheime Interesse) - "*Ein abgesichertes Wissen zur doppelten deutschen Nachkriegsgeschichte kann letztlich nur die allgemeinbildende Schule vermitteln*" (S. 107)³³ - ist empirisch untriftig und theoretisch naiv. Nach allen bisherigen Stu

³³ Unterstellt wird dabei gewissermaßen das Modell des "Nürnberger Trichters". Geschichtslernen heißt aber von sich aus "Sinn zu bilden", eine "Welt zu konstruieren", "sich zu orientieren", sein "Selbst- und Weltver-

dien ist die Schaffung eines *abgesicherten Wissens* - was immer das sei und wer immer das festzulegen versucht hat - noch stets misslungen (auch und gerade in der Schule). Eben die Befragung selbst hat zudem gezeigt, dass die Jugendlichen ihrerseits mindestens den audiovisuellen Massenmedien Film und Fernsehen - und auch "*Gesprächen mit Eltern und Freunden*" - einen größeren bzw. nahezu gleichen Einfluss wie dem Geschichtsunterricht zusprechen (S. 142), zudem auch einigen benachbarten Schulfächern (S. 143).

Alles, was wir über die Aufarbeitung des NS, des Stalinismus, des Kolonialismus etc. wissen, spricht dafür, dass lebensweltliche Mythen und Legenden durch Geschichtsunterricht nicht unterzukriegen sind (vgl. z.B. Welzer u.a. 1997, 2002, Zülsdorf-Kersting 2007, Hollstein u.a. 2002), was natürlich überhaupt nicht gegen guten Geschichtsunterricht, aber stark gegen platte "Wissenschaftsorientierung" und illusionäre Wunscherfüllungs-Fantasien spricht.

2. Forschungsverbund SED-Staat: "*Das DDR-Bild von Schülern in Berlin*" (Monika Deutz-Schroeder und Klaus Schroeder 2007)

Da die grundsätzlichen Probleme einer quantitativ-empirischen Untersuchung zum Geschichtsbewusstsein mit geschlossenem Fragebogen am vorigen Beispiel (Arnswald 2006) ausführlich erörtert worden sind und da die Schwächen vielfach erneut auftauchen, kann die Analyse der zweiten Studie (Deutz-Schroeder/Schroeder 2007) deutlich knapper gefasst werden.³⁴ Mangels eines Methodenkapitels³⁵ und eines klaren Aufbaus ist hier auch die Aufteilung auf Teilkapitel etwas schwieriger.³⁶

hältnis zu erweitern".

³⁴ Über die Befragung in Brandenburg, Nordrhein-Westfalen und Bayern - sowie den Vergleich der Bundesländer - kann leider vor Vorlage der Datenauswertung, die zunächst für Anfang 2008 versprochen war, sich jetzt aber ins Frühjahr 2008 zu verzögern scheint, nichts - weder kritisch noch zustimmend - Genaues gesagt werden. Nur die Bedingungen möglicher Aussagen überhaupt, soweit sie sich aus der Anlage und Methode der Studie ergeben, lassen sich beschreiben.

³⁵ Die Ausführungen unter "*Methodisches Verständnis der standardisierten Befragung*" (S. 40-46) und "*Zusammensetzung der Stichprobe*" (S. 46-68) können keineswegs als solches anerkannt werden. Im ersten Falle geht es um einen Teil des Fragebogens und die Bildung von Kurzskalen; im zweiten werden schlicht die Fragen zur Sozialisation und politischen Einstellung ausgewertet, was mit der Überschrift nichts zu tun hat.

³⁶ Schon die Beschreibung der Absicht fällt etwas unklar aus: "*Mit den hier vorgelegten Ergebnissen einer Umfrage unter Berliner Schülern zum DDR-Bild soll die Frage beantwortet werden, wie das DDR-Bild von Schülern - im Alter von 15 bis 17 Jahren - in Ostdeutschland im Vergleich zu Westdeutschland ausfällt und worauf die Urteile beruhen. Gleichzeitig sollte der Wissensstand der Schüler ermittelt und eine vergleichende Beurteilung von Bundesrepublik und DDR auf verschiedenen staatlichen und gesellschaftlichen Feldern erfragt werden.*" (S. 3) Wenn man nicht "West-Berlin" zu "Westdeutschland" zählt, ist der Anspruch natürlich logisch unmöglich einzulösen. Aber auch die Vorstellung - und das Verhältnis - von "Kenntnissen" und "Urteilen"/"Beurteilungen" bleiben

2.1. Stichprobe und Erhebung

Die Studie von Deutz-Schroeder/Schroeder stellt keine im technischen und anspruchsvollen Sinne "repräsentative" Studie dar (was sie durchaus noch nicht wissenschaftlich entwerfen muss). Es wird keine umfangreiche (geschichtete) Zufallsstichprobe gezogen. An Schulformen werden nur Gymnasium und Gesamtschule berücksichtigt; Realschulen und Hauptschulen³⁷ fehlen, was - rein kognitiv - die Ergebnisse massiv nach oben verschieben muss.³⁸ Wirklich dramatisch ist allerdings, dass auch in seriöser Presse (z.B. Süddeutsche Zeitung 28.12.2007, S. 7) der Eindruck entstanden ist (wie eigentlich?), die Studie sei repräsentativ (für welche Klassenstufe und Schulform?).

Durchaus argumentativ wird nämlich eine *Konvenienzstichprobe* (keine *Zufallsziehung*) angestrebt, die zunächst sechs möglichst *unterschiedliche* von zwölf Berliner Bezirken (zwei "West", zwei "Ost" zwei gemischte) abbilden will.³⁹ Konvenienzstichproben haben - für bestimmte Fragen - einen guten Sinn; aber sie müssen von repräsentativen Zufallsstichproben unterschieden bleiben. Freilich ist die Verweigerungsquote - besonders im Ostteil - so hoch gewesen (wenn auch nicht im Detail ausgewiesen), dass einerseits ein weiterer Ost-Bezirk dazu genommen wurde (S. 6, 16f.), andererseits ein gemischter Bezirk zum Fragment (in seinem Westteil, vgl. S. 20) schrumpfte. Bei so häufiger Ablehnung bekommt die Verallgemeinerbarkeit einer Studie ein Problem (systematischer "Bias"), wie auch die Autoren gelegentlich einräumen (z.B. S. 4).

Über den Prozess der Genehmigung wird nichts berichtet.⁴⁰ Sehr erstaunlich ist, dass offenbar die Schulleiter (S. 20) das letzte Wort haben; das widerspricht älteren Erfahrungen und Rechtsverhältnissen (auch der Praxis von Arnswald 2006!), nach denen - neben Datenschutzbeauftragten - Kultusminister, Schulleiter(innen), Lehrer(innen), Eltern und ggf. Schülerinnen selbst zustimmen mussten. Eine bloße Unterstützung durch die Schulbehörde, wie sie berichtet

unscharf. Deutlicher ist die Motivation; denn im vorausgehenden ersten Absatz wird die schmerzliche Erfahrung erwähnt, dass sich seit 1990 in vielen Köpfen das Bild der "DDR" aufhellt, das der "Vereinigung" eintrübt.

³⁷ Diese Schulformen sind hier relevant, da ein Teil der Befragten zweifellos noch die Sekundarstufe I besucht; in der Sekundarstufe II hätte entsprechend - neben Gymnasium und Gesamtschule - in *beruflichen Schulen* befragt werden müssen/sollen.

³⁸ Besonders deutlich ist, dass die lernenden "Ausländer" mit nur 4,7% krass unterrepräsentiert sind, vermutlich auch die lernenden "Deutschstämmigen" (meist Spätaussiedler) mit 3,1% (S. 47).

³⁹ Wahrscheinlich ist in den drei anderen Bundesländern der Vergleichsstudie (Brandenburg, NRW und Bayern) ähnlich verfahren worden.

⁴⁰ Immerhin wird erstaunlicherweise die Forderung nach schriftlicher Einwilligung der Eltern, die bei Befragungen durchaus üblich ist, als Grund genannt, auf eine Schule zu verzichten (S. 20).

wird (S. 3), genügt an sich nicht für große und datenschutzrelevante Studien. Das lässt die Ablehnung - gerade durch korrekte Ostberliner (und u.U. ostdeutsche) Schulleiter in einem wesentlich günstigeren Lichte erscheinen.

Die fehlende Repräsentativität der Studie ist insofern sehr ernst zu nehmen, als sie zwar anfangs eingeräumt (z.B. S. 4), im weiteren aber, in den Folgeinterpretationen, faktisch nicht mehr beachtet wird. Zudem ist die Verschiebung (nach "oben" und in Richtung "Akzeptanz") gegenüber dem "*Durchschnitt*" nicht quantifizierbar, vor allem aber die Vergleichbarkeit mit den anderen drei Bundesländern (Brandenburg, Bayern, NRW), in denen befragt wurde, nicht gesichert, es sei denn, dort seien genau entsprechende nicht-repräsentative Stichproben befragt worden, was sich aber methodisch praktisch schwer bewirken lässt: Gleiche Schulformen und Schulformanteile genügen - angesichts verschiedener Bildungspolitik und Rekrutierungsquoten - durchaus noch nicht; das würde auch dann noch gelten, wenn es auch in Bayern Gesamtschulen gäbe.

Die Durchführung war offenbar relativ schlank, uneinheitlich und nicht vorgeschrieben (eine "Standardisierung" ist ja nicht nur beim Fragebogen, sondern auch bei der Testsituation selbst nötig!); die Testleiter(innen) haben offenbar ausführlich Fragen der Lernenden beantwortet und Zusatzinformationen gegeben; anders kann der Satz "*Das Vorgehen und die Antwortmöglichkeiten wurden den Schülern von den Mitarbeitern des Projektes, die die standardisierte Befragung vor Ort durchführten, vor dem Ausfüllen des Fragebogens ausführlich erläutert. So konnten Unklarheiten über die Aussagerichtung bestimmter Statements, aber auch Vorbehalte gegen einzelne Items, vor allem bei Fragen nach Ausländern in der DDR, ausgeräumt werden.*" (S. 42) wohl nicht verstanden werden. Dabei dürfte unbewusster Suggestion Tür und Tor geöffnet gewesen sein (geschulte Testleiter in anderen Studien haben ziemlich klare Verhaltensanweisungen und dürfen - außer freundlich zu sein und zu ermutigen - ziemlich wenig eingreifen).

Die Lehrenden wurden nicht selbst gleichzeitig mit einem anderen Fragebogen schriftlich befragt, sondern waren meist in den Klassen - mit vielfach auch angewandten Kontrollchancen und Autoritätsmöglichkeiten gegenüber den Lernenden - anwesend. Die Autoren berichten darüber - vor allem hinsichtlich des Ostteils der Stadt - gelegentlich in geradezu feindseligem Ton. Das macht zusammen mit dem eher lässigen Umgang mit der Anonymisierung (jedenfalls auf Schulebene, vielleicht sogar auf Lehrerebene) ein Problem - auch der Objektivität - aus. Auch Anwesenheit und Beeinflussungsmöglichkeit der Lehrenden weichen von strikten methodischen

Anforderungen ("state of the art") ab, erlauben aber ihrerseits - was auch genutzt wird - Rückschlüsse auf das Schulklima. Auch wenn ich das selbst deutlich anders gemacht hätte (und habe), können die Befunde - nach weiteren Vorsichtsmaßnahmen - als bedingt belastbar angesehen und - mit Vorsicht - interpretiert werden. *Man muss sich ja immer bewusst halten: Keine sozialwissenschaftliche Methode - auch die elaborierteste, die hier aber nicht vorliegt! - ist perfekt, jede hat ihre Kosten und Grenzen. Es gibt zwar Standards, aber kein "allein wissenschaftsmachendes" Verfahren.*

Die Befragung von 2.356 Schüler(inne)n aus 22 Schulen bzw 86 Klassen und Kursen hat 2005/06 stattgefunden.⁴¹ Aber in welchen Klassenstufen? Darüber werden merkwürdigerweise - es klingt kaum glaublich - keinerlei Angaben gemacht. Man muss vielmehr aus dem Alter der Befragten indirekte Schlüsse ziehen. Da etwa 23% unter 16 Jahren, 46% 16 Jahre und 31% ab 17 Jahren alt sind (S. 47),⁴² kann es sich nicht durchgehend um 11. Klassen handeln, wie man zunächst vermuten möchte. Denn in Klasse 11 könnten Fünfzehnjährige, z.B. "Springer", nur ziemlich vereinzelt auftauchen. Wahrscheinlicher ist, dass die Probanden die 10. bis 12. Klasse (vielleicht sogar die 9. bis 13. Klasse) besuchen, allerdings mit einem Schwerpunkt auf der 11. Klasse. Eine verschiedene schulische Klassenstufe ergibt sich auch ziemlich zwingend daraus, dass von den unter 16-jährigen noch 37%, von den ab 17jährigen nur noch 14% angeben (16jährige: 20%), keinen Unterricht über DDR-Geschichte genossen zu haben (S. 53).

Zur Erhebungsmethode kann man auch die *Zusatzzinformationen* rechnen. Die Nachbefragungen einzelner Lernender, die allerdings auf dem Wege über Freiwilligkeit eine - zunächst durchaus gesehene und eingeräumte - sehr hohe Selektivität implizieren, sind positiv zu bewerten. Da wird eine Chance zur Zweistufigkeit, zur Kombination von quantitativen und qualitativen Verfahren sowie zur methodischen "Triangulation" eröffnet. Die Dokumentation⁴³ wie die Auswertung überzeugt allerdings nicht oder weniger, und zwar weil hier nicht die *Logik* der - ja sehr vereinzelt und unrepräsentativen - "Einzelfälle" ("*Fall-Rekonstruktionen*") im Detail, in den Zusammenhängen und den Wirkungsmustern untersucht wird, sondern unternommen wird,

⁴¹ Über die Rücklaufquoten (z.B. Vollständigkeit der Anwesenheit) gibt es keine Auskünfte.

⁴² Für diese drei Altersgruppen werden im Folgenden bei jedem Einzelitem immer getrennte Prozentangaben gemacht, freilich nicht in Kombination zur Ost- oder West-Herkunft, sondern nur alternativ dazu. Die beiden Abhängigkeiten der Aussagen vom Alter und von der Stadthälfte sind also nicht methodisch getrennt. Das könnte eine beachtliche Verfälschung nach sich ziehen.

⁴³ Vielfach heißt es bei durchaus wichtigen Punkten einfach nur "*wie wir aus Gesprächen erfuhren*" (S. 167) oder "*wie wir in den Gesprächen erfuhren*" (S. 171). Das ist - für zentrale Informationen - in sozialwissenschaftlichen Studien eher eine ungewöhnliche Form des Beleges.

Grundzüge eines gemeinsamen typischen und stereotypen Weltverständnisses (DDR-Verständnisses) der gesamten Gruppe aus wenigen Extremfällen herauszuarbeiten, damit also gegen die anfängliche Vorsichtsbemerkung (wegen "Selektivität") verstoßen wird.⁴⁴

Methodisch auffallend ist, dass in den Nachgesprächen die Konfrontation mit den Jugendlichen durchaus nicht gescheut, wenn nicht gar gesucht wurde. Sie wurden mit entgegenstehenden "Informationen"/"Tatsachen" versehen, auf die sie dann natürlich *reagieren* mussten (z.B. betroffen, erschreckt). Das kann man so machen; Piaget ist z.B. so verfahren. Aber man ist damit in einem anderen Paradigma, weil man nicht mehr Bestände oder Performanzen untersucht, sondern Prozesse und Operationen, u.U. in Belastungssituationen, beobachtet. Außerdem stellt sich das Problem des Autoritätsgefälles in der Forschungssituation verschärft. Das ist dann natürlich eine andere Logik als die der "Befragung"; es muss als "Intervention" bezeichnet werden.⁴⁵ Wie knapp und entschieden dürfen solche Interventionen sein? Was mißt man mit ihnen an Lernprozessen und was an Verhaltenskonventionen? Dass sich trotzdem im einzelnen viele interessante Bemerkungen (allerdings fast ohne wörtliche Zitate) finden, braucht wohl nicht besonders betont zu werden.

Auch in den Klassen war die Erhebungssituation offenbar nicht "standardisiert"; die Testleiter(innen) dürften nicht geschult worden sein. Da bei Deutz-Schroeder/Schroeder ein ausgesprochenes Methodenkapitel - ganz gegen die üblichen Regeln in der Sozialwissenschaft - fehlt, lässt sich dazu leider wenig Exaktes sagen. Immer wieder wird aber erwähnt, dass man nach der Befragung diskutiert und auch per Handaufheben - d.h. "offen" - abgestimmt habe (ein wesentlicher Teil der Informationen und auch der Interpretationen des quantitativen Materials wird daraus gezogen). Das lässt nicht mehr auf strikt neutrale "Sozialforscher" schließen, was gerade beim relativ heiklen Thema DDR gewiss keine Stärke ist. Andererseits wird versucht, bei diesen Gelegenheiten auch durch Provokationen und Handzeichen-Abstimmungen Zusatzinformationen zu erheben und z.B. Widersprüche zwischen Lehrer-Unterricht und Schüler-Hintergrund oder Fragebogenankreuzung und Schülerüberzeugung herauszukitzeln. Das ist spannend, entspricht aber eher einer "wild-ethnologischen" als einer "klassisch-sozialwissenschaftlichen" Methode.

⁴⁴ Zudem dürfte die gewählte Interpretation von Schüleräußerungen als Ausdruck von "political correctness" ein theoretisch weniger angemessenes Konzept sein als die geläufigere "soziale Erwünschtheit".

⁴⁵ Was in einer Befragungsstudie ein unerwünschter "Methodeneffekt" ist, kann in einer "Interventionsstudie" zum zu messenden "Lerneffekt" werden. Aber da braucht man eine Menge theoretischer und praktischer Absicherungen. *Alles ins Eins* zusammenwerfen sollte man nicht.

Interessant ist der Bericht über Lehrergespräche (S. 33-39); freilich bleibt der methodische und methodologische Status offen. Verwendet sind offenbar Aufzeichnungen, die im Anschluss an ausgesprochen informelle Gespräche in den Schulen (Pause, Freistunde) gemacht worden sind. Dieser wiederum eher ethnologische als sozialwissenschaftliche Zugriff erbringt zum Teil spannende Ergebnisse (auch recht elaborierte und selbstkritische Äußerungen einzelner Befragter), ist aber methodologisch und forschungsethisch nicht ganz unbedenklich. Wie steht es zum Beispiel mit einer wirklich perfekten, auch für Insider nicht zu knackenden Anonymisierung, wenn z.B. von zwei Geschichtslehrern (aus dem Osten und dem Westen) an einem Gymnasium in einem ausdrücklich genannten Stadtteil die Rede ist? Wir selbst haben uns da stets höhere Hürden aufgebaut.

2.2. Instrument und Auswertung

Der relativ kurze Fragebogen enthält zunächst sozialstatistische Angaben und Aussagen zur Sozialisation (z.B. zu historischen Gesprächen in Familie und Freundeskreis), zur Parteipräferenz, zu Bekanntschaften im anderen Teil der Stadt (und mit Ausländern), aber auch zur Behandlung des Themas DDR im Unterricht.⁴⁶ Das ist im Prinzip vernünftig, wenn auch manches präziser - oder weniger suggestiv - formuliert sein könnte.

Der erste Hauptteil besteht in Vergleichsaussagen über BRD und DDR ("**Deutschland vor 1989**: Was glauben Sie war besser oder gleich in der DDR oder in der Bundesrepublik Deutschland?"). Schon diese Fassung - mit dem doppelten "oder" ist nicht gerade glücklich (trotz "Pilotierung"!); Wichtiger: Solche - extrem simplen und vielfach logisch keineswegs eindeutigen - Fragen (z.B. "Schulausbildung", "Hilfsbereitschaft" oder "Theater und Kino") kann man natürlich stellen und (technisch) auswerten.

Man muss aber wissen, dass der Status völlig unklar ist, weil dabei (quantitative und qualitative!) Tatsachenbehauptungen, generalisierende Deutungen und Bewertungen von Konsequenzen ununterscheidbar ineinander fließen. Wie soll man die Befunde einigermaßen konsistent und argumentativ interpretieren? Zudem - und das haben die Forscher im Nachhinein selbst erkannt und betonen es mehrfach - werden die (meist extrem vagen) Formulierungen von vielen Befragten nicht oder nicht halbwegs dem "Gemeintem" ähnlich verstanden.⁴⁷ Die Ergebnisse

⁴⁶ Diese Fragen von Deutz-Schroeder/Schroeder (2007) lassen sich mit der Studie von Arnswald (2006) besonders gut vergleichen, jedenfalls weit besser als die meisten anderen Items.

⁴⁷ Insofern scheint die Pilotierung ("Pretest") in zwei Spandauer Schulen (S. 7) den erwünschten und beschriebenen Zweck der Verbesserung des Instruments nur sehr teilweise erfüllt zu haben.

sind daher nur recht wenig wert, weil unklar ist, was sie messen.

Gewiß ist nur - und deshalb fallen die Ost-West-Unterschiede in der Stadt Berlin dann vielfach so groß aus -, dass bei Vagheit und Unverständlichkeit alle Items (mal mehr, mal weniger) zu bloßen Sympathieprojektionsflächen zugunsten der DDR oder zugunsten der BRD zu mutieren drohen (also Abhängigkeit der Antworten von "Identifikationen"). Das verwandelt sich dann bei der Auswertung leicht in ein Verfahren des - nicht nur wertlosen, sondern gefährlichen - "wissenschaftlichen" Zirkelschlusses. Die bloße Bevorzugung einer Seite (DDR oder BRD) - und nach eben der wird bei Deutz-Schroeder/Schroeder gefragt - ist, wegen unpräziser Maßstäbe, schlicht eine schlechte Frageform.⁴⁸

Im zweiten Hauptteil des Fragebogens von Deutz-Schroeder/Schroeder werden zahlreiche wertende Statements aufgelistet, die auf fünfstufigen Likert-Skalen einzuschätzen sind und von denen viele - aber nicht alle - bei der Auswertung später in sieben Subskalen einer DDR-Wertschätzung sowie eine Skala DDR-BRD-Äquidistanz eingehen werden. Ein solches Vorgehen der Einstellungsmessung ist, auch wenn die Differenz zur vorherigen Fragegruppe nicht ganz deutlich wird, im Prinzip sinnvoll.⁴⁹ Das heißt nicht, dass die Einzelformulierungen besonders glücklich gewählt wären. Vor allem ist der Status der Antworten zwischen historischer Sachanalyse, historischem Sachurteil bzw. Werturteil, politischer Einstellung usw. ziemlich unklar.

Der letzte Hauptteil besteht - angeblich - in Kenntnisfragen zur DDR-Geschichte. Dabei sind verschiedene Typen zu unterscheiden:

- Es gibt drei schlichte Kenntnisfragen nach "Mauerbau" 1961 (Jahr und Akteur) und "Volksaufstand" 1953 (die Terminologieentscheidung, die auch eine Deutungsfrage ist, sei hier ausgeklammert). Wie immer in solchen Fällen ist ein enttäuschendes Ergebnis voraussehbar, aber auch der diagnostische Wert muss begrenzt bleiben.
- Einige strukturelle Fragen ("Todesstrafe", "Lebenserwartung", "Umwelt", "Arbeitslager", "Einkommensgleichheit" in der DDR) sind interessant, aber aus verschiedenen Gründen hoch problematisch. Zunächst sind sie schlicht nicht curriculumgerecht, da gerade diese Gesichtspunkte in Lehrplänen und Unterrichtspraxis marginalisiert werden (oder auch fachlich marginal sind). Die Sozialforscher setzen gewissermaßen ihre fachliche Position - in Einzel-

⁴⁸ Weit überlegen ist es, beide Seiten getrennt voneinander jeweils im einzelnen (z.B. auf Likert-Skalen) einschätzen zu lassen und dann - zusätzlich zu den beiden Bewertungen des gleichen Phänomens im Osten und Westen den Vergleich, also die "Bevorzugung" auswerten zu können. Genauso wurde 1992 und 1994 verfahren (vgl. v. Borries u.a. 1995, 1999) - und die Ergebnisse und Instrumente waren seit 1995 publiziert.

⁴⁹ Genau so sind wir selbst mehrfach in eigenen Studien verfahren (vgl. v. Borries u.a. 1995, 1999, 2005).

heiten wie in der Gewichtung - absolut. Erschwerend kommt hinzu, dass - mangels öffentlich-geschichtskultureller oder schulisch-unterrichtlicher Erwähnung dieser Fragen (seien sie so wichtig oder so unwichtig sie wollen) - die Antworten nur aus der Tendenz bzw. dem Urteil über die DDR insgesamt erschlossen werden können. Mit anderen Worten: Auf einem wenig raffinierten Umweg handelt es sich *erneut* um Gesinnungs- und Bewertungsfragen (wiederum "*Identifikationen*"), nicht eigentlich um Wissensfragen oder Einsichten. Das wird natürlich hoch relevant, wenn man später mit den Einstellungen korrelieren will. Ein Beispiel ist besonders schlagend: In der Frage der Einkommensgleichheit/Sozialgerechtigkeit ("*In der DDR hatten alle etwa das gleiche Einkommen und Vermögen.*", Fr.S. 5) enthält das Item selbst ein - ziemlich unbestimmtes - "*in etwa*". Dennoch werden alle Antworten, die der DDR - relative - Gleichheit bescheinigen, als falsch abgelehnt. Denn die Forscher empfinden einen Faktor von ca. Eins zu Fünf oder Eins zu Zehn eben nicht mehr als "*in etwa*" gleich. Das ist ziemlich kühn, weil ja die Lernenden unter der Hand unvermeidlich immer vergleichen müssen und in einer Bundesrepublik leben, in der - inzwischen als "Gerechtigkeitslücke" kritisiert - Unterschiede des Einkommens - von Vermögen zu schweigen - von Eins zu Hundert selbstverständlich sind. In solchen Teilen arbeiten die Forscher einseitig und suggestiv.

- Danach werden Spitzenpolitiker beider Seiten der DDR wie der BRD zugeordnet. Das ist ein sehr einfaches und effektives Verfahren, es setzt aber leider einen bestimmten (eher veralteten) Geschichtsbegriff schon voraus, nach dem es vor allem auf die Leistungen - und Fehlleistungen - der "großen Männer" (neuerdings auch seltener "großer Frauen") ankomme. Persönlich würde ich auch diese Fragen nicht stellen und für diagnostisch nicht besonders geeignet halten. Im sogenannten Wissenstest besteht aber - ähnlich Arnswald (2006) - die Mehrheit aus solchen relativ einfallslosen Namensfragen.

Auf die Curricula der Länder wird im - übrigens kaum begründeten - Fragebogen explizit kein Bezug genommen. Man hat eher den Eindruck, dass es darum geht, von einer festen eigenen Meinung über Bedeutsames aus das - voraussehbar - Defizitäre an Curricula und Praxis zu beleuchten. Die Validität (Gültigkeit) des gewählten Instruments für "Kenntnisse und Einstellungen zur DDR-Geschichte" kann durchaus nicht bestätigt werden (weniger noch als im Falle Arnswalds!), was nicht heißt, dass nicht einzelne Befunde - auch viele Befunde - ausgesprochen spannend sein können.

Über die Art der Auswertung verlautet so gut wie nichts. Auch hier fällt das Fehlen eines Methodenkapitels unangenehm auf. Man kann nicht wissen, wieweit die Forscher schlichteste fachliche Standards beachtet haben.⁵⁰ Nicht einmal Mittelwerte und Standardabweichungen werden mitgeteilt. Die beiden unteren und die beiden oberen Stufen der Angaben werden stets schon zusammengezogen abgedruckt, so dass faktisch eine Schrumpfung auf drei Niveaus ("Ja", "Enthaltung", "Nein") erfolgt ist und ein großer Teil der Originalinformation verschenkt wird.⁵¹ Soweit ersichtlich, ist keine Wissensskala erstellt worden (damit auch keine Konsistenzprüfung). Außerdem scheinen nicht einmal Korrelationen (statistische Zusammenhänge) berechnet worden zu sein. Gleichwohl werden Kausalitäten ("Auswirkungen") angedeutet und behauptet, so ganz zum Schluss von "konkretem Wissen über die DDR" auf Kenntnis der "Trennlinien zwischen Demokratie und Diktatur" (S. 319). Das ist natürlich - wenn auch inhaltlich nicht unplausibel - methodisch schlicht unzulässig.

Auf verschiedenen Ebenen und nach verschiedenen Kriterien muss man feststellen, dass es sich nicht um eine methodisch herausragende Studie handelt. Die handwerklichen Mängel sind unverkennbar; zum "state of the art" besteht ein großer Abstand. Die Vorstellung von Geschichtsunterricht und seiner Wirkung sowie Geschichtskultur und Geschichtswissenschaft (keinerlei "Kontroversität") ist eher schlicht. Das heißt natürlich nicht, dass eine kluge Lektüre des Vorliegenden (von einer elaborierteren Auswertung des Datensatzes ganz abgesehen!) nicht durchaus ertragreich und relevant sein könnte, erst recht natürlich im - noch nicht zugänglichen (aber für Anfang 2008 versprochenen) - Vergleich mit anderen deutschen Bundesländern (*der allerdings unbedingt vorsichtig verfahren muss*).

Die Auswertung ist erstaunlich bescheiden, ja ärmlich, da zwischen den einzelnen Items keinerlei Beziehungen hergestellt werden ("multivariate Analyse"), sondern alle Aussagen im Bereich der Abhängigkeit von "sozialstatistischen Variablen" (Ost-Berlin - Westberlin, Jungen - Mädchen, Deutsche - Ausländer - "Deutschstämmige" [d.h. Spätaussiedler], Jüngere - Mittlere - Ältere) bleiben. Selbst die inneren Differenzierungen Berlins zwischen reichen begünstigten und armen belasteten Vierteln, die bei der Herstellung des Konvenienz-Samples so wichtig zu sein schienen, werden überhaupt nicht wieder aufgenommen. Das ist erstaunlich.

⁵⁰ Die (S. 42, 43) erwähnten Faktorenanalysen dienen nur der Konsistenzprüfung von Kurzskalen, ihre Ergebnisse sowie die Reliabilität der Skalen und das Zusammenstimmen der Teilskalen, mithin auch die Konsistenz der Gesamtskala (S. 242-246), werden nicht erwähnt, geschweige denn ausgebreitet.

⁵¹ Die Überlegungen zur Willkür gesetzter Schwellenwerte und zur künstlichen Herstellung von drei Hauptgruppen auch bei Skalen (S. 42f.), nicht nur bei Einzelitems, sind ehrenwert, bleiben aber im weiteren Verlauf ziemlich folgenlos.

Bei den zahlreichen "Bevorzugungsitems" (S. 69-164 "*Der bewertende Blick auf das geteilte Deutschland*") wird zunächst jeweils in oft ziemlich ausführlichen Abhandlungen (z.B. 82-85, 94-98, 102-105, 144-146) dargelegt, welcher Staat tatsächlich "besser" - also zu bevorzugen - war. Dann werden die Antworten der Jugendlichen tabellarisch vorgestellt und anschließend verbal umschrieben. Eigentliche Interpretationen fehlen; doch werden meist - oft interessante, aber natürlich nicht verifizierbare und quantifizierbare - Informationen aus den "Nachdiskussionen" angeführt, oft genug solche über ein völlig anderes Verständnis der Frage bei den Probanden als bei den Forschern: recht viele Items waren ja hoch abstrakt oder ausgesprochen "doppelköpfig", d.h. uneindeutig. In manchen Fällen gibt es, wie man anerkennen muss, auch "Selbstkritik" der Forscher (z.B. S. 156). Man wünschte nur, das bliebe auch in anderen Teilen der Darstellung - und bei der öffentlichen Vermarktung - stets bewusst; sonst handelt es sich um eine unverbindliche Bescheidenheitsfloskel.

Es ist deutlich, dass die Verfasser ein recht schlichtes Verständnis vom Geschichtslernen haben: Die Befunde der Wissenschaft, wie sie sie verstehen und erläutern, sollten in einer Art "Eins-zu-Eins-Umsetzung" in die Köpfe der Schüler übertragen worden sein. In naiv-aufklärerischer und auch etwas eifernder Weise wird ein intellektuell und moralisch korrektes Modell-Abbild der Vergangenheit verlangt. Dass die Jugendlichen jeweils eine - in Ost und West verschiedene - krude Kombination von "Unwissenheit", "Konsumismus" und "DDR-Nostalgie" ("Ostalgie") präsentieren, erschreckt die Forscher. Sie ahnen allerdings, dass es dabei mehr um den Ausdruck von *Identifikationen* geht als um *historische Analysen*.

Bei den folgenden Einstellungs-Statements mit Likert-Skalen ("*Die Skalen und ihre Items*", S. 165-246) werden jeweils (meist) mehrere angebliche verwandte bzw. bedeutungsähnliche Items nacheinander abgehandelt (sieben Teilskalen, S. 178-241) und dann zu einer "Bild der Skala" (S. 242ff.) addiert.⁵² Bei Skalen ist es üblich, ja zwingend, eine Analyse auf die Konsistenz und damit Zuverlässigkeit ("Reliabilität", "Konsistenz") vorzunehmen (übliches Maß: Cronbachs Alpha). Besonders wichtig wäre natürlich auch eine Kontrolle der Übereinstimmung bzw. Überschneidung zwischen den Teilskalen (Korrelation) und damit eine Analyse der Dimensionalität. Alles das wird leider nicht geleistet, mindestens nicht vorgeführt; man weiß also nicht, ob man sich methodisch in einem einigermaßen gesicherten Gelände bewegt.

Aber was kommt heraus? Zunächst ist von der *Einzelkala* (deshalb nicht als "Teilskala" zu

⁵² Positiv kann man nur sagen, dass - zwecks Minimierung des Stellenwerts von etwaiger "Bejahungstendenz" - möglichst jedesmal etwa gleich viele normale und "re-codierte" (umgepolte) Items benutzt sind (S. 41).

bezeichnen) "Äquidistanz" zu berichten. Der Befund einer in Ost-Berlin wesentlich stärker verbreiteten - aber keineswegs dominanten - "Äquidistanz" zu DDR und BRD (S. 165-178) ist nicht gerade überraschend für jemanden, der gelegentlich auf Wahlergebnisse geschaut hat.

Bei den folgenden sieben (Teil-)Skalen "Bild der DDR" (178-240) sieht es ähnlich aus.⁵³ In einer Gesamtskala (S. 241, 242) wird wirklich einmal ein relevanter Durchschnittswert berichtet. Das "(günstige) DDR-Bild" kriegt im Osten einen Mittelwert von 3,34, im Westen einen von 3,69 (wobei 1,0 der günstigste mögliche und 5,0 der ungünstigste mögliche Wert ist). Hier haben wir einmal nicht bloße Prozentzahlen, sondern eine Mittelwertdifferenz von einem Drittel eines Punktes, was leider wieder dadurch entwertet wird, dass die Standardabweichung nicht mitgeteilt wird, die ja zur Abschätzung der Effektstärke die nächstliegende Möglichkeit bietet.⁵⁴ Man kann aber sagen, dass die Westberliner doppelt so weit vom Neutralitätspunkt entfernt sind wie die Ostberliner. Die Gesamtschüler sind auch DDR-freundlicher als die Gymnasiasten (Differenz 0,25 Punkte) und die Mädchen etwas DDR-freundlicher als die Jungen (Differenz 0,07 Punkte).

Die "*Wissensfragen zur DDR/BRD*" (S. 269-305) sind - wie oben beschrieben - besonders problematisch. Die Zuordnung der Staatsmänner zu Osten und Westen kann als relativ einfach und wirkliche - wenn auch nicht besonders wichtige - Kenntnisangelegenheit anerkannt werden (S. 288-302). Auch die drei Einzel-Items (Datum des Mauerbaus, Verantwortliche des Mauerbaus und Ort des Volksaufstands) können als - wenig erleuchtete - Wissensitems gelten. Die Fragen der Todesstrafe, der Einkommens-Gleichheit, der Arbeitslager, der Lebenserwartung und der Umweltqualität messen in Wahrheit Sympathie - nicht historische Kenntnis - und (sie hätten locker auch in der ersten Gruppe der "Bevorzugungen" stehen können).

2.3. Befunde und Interpretation

Eine gedankenreiche und problembewusste Auswertung hat - noch? - nicht stattgefunden. Auffregend sind z.B. ziemlich massive Geschlechtseffekte, wobei - was nicht ganz neu ist - die Jungen oft wesentlich besser informiert, die Mädchen oft reifer im Urteil (aber auch merkwürdig

⁵³ Die kleinste Differenz zwischen Ostberlinern und Westberlinern ergibt sich bei "*Verharmlosung der Diktatur*" und "*Verharmlosung der Unterdrückung*", die größte bei "*Außenpolitik*", "*Sozialpolitik*" und "*Wirtschaftspolitik*" (S. 242). Die DDR-Nostalgie setzt eben nicht bei einer wirklichen Verteidigung des "*politischen Systems*" an, sondern bei der - rückblickend verklärten - Sozial- und Wirtschaftspolitik.

⁵⁴ Nach der Erfahrung meiner Studien pflegen Standardabweichungen bei fünfstufigen Likertskalen etwa zwischen 0,8 und 1,2 zu liegen. Dann würde es sich - nach Cohen 1988 - bei der Ost-West-Differenz um eine (knapp) mittlere Effektstärke handeln (29% bis 44% der Standardabweichung).

DDR-freundlicher) sind.⁵⁵ Wenn man diesen Hinweisen in Detailuntersuchungen nachgehen würde, wären spannende Befunde möglich. Tatsächlich unterliegen beide Geschlechter ja demselben Geschichtsunterricht und - weithin - auch denselben massenmedialen Einflüssen. Die drastischen Geschlechtsdifferenzen sind also ein deutliches Anzeichen dafür, dass es auf sozialpsychologische Verarbeitungsprozesse (auf Bedürfnisse, Kommunikationen, Konstruktionen, Identitätsentwürfe) ankommt und eben nicht auf unterrichtlich abgespeicherte Fakten und unterrichtlich übernommene Deutungen. Mit einer solchen Betrachtungsweise würde man an die wichtigen Fragen heranrücken, die die Untersuchung bisher nur streift. Der Geschlechtseffekt ist nur ein Beispiel für eine bisher verschenkte Interpretationschance

Das alles, Erhebungskritik wie Auswertungskritik, heißt nicht, dass die Studie ohne jeden Wert wäre. Man hätte sie nur - bei gleichem Geld-Aufwand, aber wesentlich mehr Denk-Aufwand, vor allem bei der Entwicklung des Instruments (Fragebogen) - deutlich besser machen können. Gleichwohl belegt sie - für den Augenblick - einige wichtige Punkte, die zwar nicht grundsätzlich neu sind, die aber - trotz Arnswald 2006 - nicht unbedingt für die allerletzten Jahre nachgewiesen waren, z.B.:

- Das Wissen von Jugendlichen über "DDR-Geschichte" und "BRD-Geschichte" ist ziemlich gering (nur bei einem valideren Fragebogen könnte man das näher spezifizieren). Das wäre allerdings bei anderen Themen (selbst NS, erst recht Kolonialgeschichte und Dekolonisation, Französische Revolution oder USA als Weltmacht) höchst wahrscheinlich nicht besser. Jedenfalls legen das alle älteren Befragungen nahe.
- Die Einstellungen - zur DDR und zur Alt-BRD - sind im Osten und Westen Berlins (und sicherlich auch ganz Deutschlands) recht verschieden. Das ist verständlich, weil - auch ohne exakte oder umfangreiche Informationen - die gefühlte *Bestätigung der eigenen Biografien* im "kommunikativem Gedächtnis" eine mächtige Determinante für die Verarbeitung von Zeitgeschichte darstellt.
- Sowohl das Konzept "Demokratie" ist vielen Jugendlichen unklar; als auch die Bedeutung von "Demokratie" wird bei vielen ausgesprochen gering eingeschätzt (z.B. S. 159). Das ist auch in vielen älteren Studien so gefunden worden (z.B. sogar die Bevorzugung von Meinungsfreiheit vor Demokratie). *Sicherheit* geht relativ vielen Menschen vor *Freiheit*. Aber

⁵⁵ Auch das ist nicht neu, sondern war seit 1992 so (vgl. v. Borries u.a. 1995). Vermutlich ist es, auch wenn Deutz-Schroeder/Schroeder das zu widerlegen suchen, doch nicht ganz falsch, wenn die weibliche Hälfte sich eher als "Verlierer(innen) der Wiedervereinigung" empfindet.

auch das wurde wohl schon genauer und besser gemessen.

Die ausgesprochen mindere methodische Qualität der Studie darf nicht ohne weiteres dazu führen, die Befunde, die ja so neu nicht sind, nicht ernst zu nehmen. Dass aus Parteienverdrossenheit, die allerdings bei Deutz-Schroeder/Schroeder nicht thematisiert wird, eine Systemkrise werden kann, dürfte unbestritten sein. Auch formalisierte demokratische Verfahren gelten, soweit man weiß, Jugendlichen weniger als ad-hoc-Initiativen (auch diese Frage wird bei Deutz-Schroeder/Schroeder nicht angegangen).

Wie leicht oder schwer man den beklagten Zustand historisch-politischer Bildung durch Instruktion (also "Intervention") ändern kann, lässt sich aus der Studie, auch wenn sie in etwas "wilden" Verfahren jeweils Nachdiskussionen in den Klassen geführt hat, nicht erschließen. Die Versuche von Kausalattributionen sind methodische Erschleichungen und müssen daher als unzulässig zurückgewiesen werden. Wer dazu etwas sagen will, muss ernsthafte Unterrichtsexperimente veranstalten, die sich aber methodisch als ausgesprochen schwierig erwiesen haben. Locker von oben herab - und angeblich im Besitz der alleinigen historischen und sozialwissenschaftlichen Wahrheit - curriculare Forderungen aufzustellen und bildungsmäßige Erfolge zu versprechen, wäre ausgesprochen unseriös.

2.4. Folgerungen und Fazit

Als Fazit kann man feststellen: Die Studie von Deutz-Schroeder/Schroeder befindet sich hinsichtlich ihrer theoretischen Grundlagen und ihrer methodischen Standards nicht auf einem besonders hohen Niveau. Sie verletzt einige Regeln, die man - als "state of the art" - beherzigen sollte. An manchen Stellen ist sie übermäßig lang und umständlich.

Die *Fragestellung* wird nicht ausführlich entfaltet. Die *Hintergrundtheorien* sind nicht expliziert. Doch wird klar, dass ein einigermaßen elaborierter kulturwissenschaftlicher Ansatz zu "Geschichtsbewusstsein" und "Geschichtskultur" nicht vorliegt. Das geschichtswissenschaftliche Verständnis ist eher positivistisch (allerdings in Bezug auf DDR-Geschichte vorzüglich informiert).

Das *Materialkorpus* (die selbst durchgeführte Befragung mit ihren Nachinterviews, Klassendiskussionen und Lehrergesprächen) leidet an verschiedenen Schwächen, besonders Nicht-Repräsentativität der Stichprobe, Nicht-Dokumentation - oder jedenfalls Nicht-Zitation - der Zusatzbefragungen, Klassengespräche und Lehrerunterhaltungen, Nicht-Validität des Fragebo-

gens).

Die *Auswertungsmethode* bleibt schlicht deskriptiv ohne irgendwelche raffinierteren statistischen Analysen, mit denen sich andere Schwächen - wie die ungleiche Verteilung der Klassenstufen - kompensieren ließen, und deshalb auch mit bescheidenen Interpretationschancen, die aber nicht einmal genutzt werden. Es ist jedoch durchaus möglich, dass Deutz-Schroeder/Schroeder in der Buchfassung bzw. beim Ländervergleich noch Analysen nachlegen wollen und können. Auch zu *Schlussfolgerungen* äußert sich - abgesehen von der allgemeinen Warnung vor Fehlentwicklungen und der nicht einmal expliziten Forderung von mehr Unterricht über DDR-Geschichte⁵⁶ - das Skript von Deutz-Schroeder/Schroeder praktisch nicht.

3. Vergleich und Beurteilung

3.1. Einordnung in Hintergrundtheorien

Die Befunde von Arnswald (2006) wie von Deutz-Schroeder/Schroeder (2007) sind viel weniger spektakulär, als scheinbar der Eindruck entstanden ist (oder erweckt werden soll). Bei noch jeder Wissensbefragung von Jugendlichen zu historischen Themen hat sich kognitiv "*eine blanke Katastrophe*" herausgestellt; noch bei jeder Einstellungsuntersuchung sind bei großen Minderheiten "*höchst bedenkliche Positionen*" herausgekommen. Das ist übrigens bei der Befragung von Erwachsenen überhaupt nicht anders. Wer, von hoch normativen Vorstellungen - oder von einer *reinen* Fachwissenschaft - herkommend, soziale Wirklichkeit genau anschaut, pflegt tief enttäuscht zu werden.

Speziell zu den Ost-West-Unterschieden liegen auch frühere Ergebnisse voll repräsentativer Studien (1992, 1994) mit krassen Unterschieden in der Einschätzung der DDR und der BRD - und ziemlich geringen Unterschieden in den meisten anderen Fragen - vor (v. Borries u.a. 1995, v. Borries u.a. 1999). Diese Befunde, die ausführlich interpretiert worden sind, sind den Veranstaltern der neuen Befragungen offenbar unbekannt geblieben (bei Arnswald findet sich die Erwähnung eines einzigen kurzen Aufsatzes). Auch die - zentral einschlägigen - Shell-Jugendstudien (z.B. Jugend '97, Jugend 2000, Jugend 2002, Jugend 2006) werden nicht herangezogen.

Es ist mehr als naiv, zu glauben, die Kinder der Menschen, die in Ostdeutschland andere

⁵⁶ Hier genügen - scheinbar empirische - Sätze wie "*Mehr als 80% aller Schüler bekennen, nichts oder wenig über die DDR zu wissen - ein beschämendes Fazit! Allerdings bekundet ein Großteil starkes Interesse an der DDR und möchte mehr über sie und das Leben dort erfahren (...)*" (S. 67).

biografische Erfahrungen und mentale Verarbeitungsprozesse (einschließlich Um-Interpretationen und "Ostalgie") hinter sich haben als im Westen (übrigens seit 1990 auch ein erheblich anderes Wahlverhalten an den Tag legen), würden - nach ein paar Stunden von Lehrplan, Schulbuch und Lehrperson in einem eindeutig steuerbarem Geschichts- oder Politik-Unterricht das selbe DDR- und (Alt-)BRD-Bild haben können wie Kinder aus westdeutschen Bundesländern.

- Das ist - mehr implizit als explizit - eine grobe Verkennung von "Lernen". Lernen ist nicht eine anzuordnende und automatisierbare Informationsaufnahme (Modell "*Nürnberger Trichter*"), sondern ein eigener Konstruktionsprozess ("*Selbst- und Weltverhältnis*"), bei dem vor allem die Anknüpfung an vorherige mentale Strukturen entscheidend ist ("Vernetzung"). Eben das führt aber auch zu Auswahl und Um-Interpretation ("*selektive Aufsuchung, selektive Verarbeitung, selektive Abspeicherung, selektive Erinnerung*"). Zudem müssen die tatsächlich aufgewendeten Zeiträume bedacht werden: wenige Lektionen bleiben gegenüber den Hauptbeschäftigungen auf der Strecke, wenn es sich nicht (ausnahmsweise) um "Durchbrüche", "Aha-Erlebnisse" oder "fruchtbare Momente" handelt. Und die kann man nicht verordnen (auch nicht mit dem Pflichtbesuch einer Gedenkstätte für Stasi-Opfer).
- Ebenso grob wird - auch das mehr subkutan - die Logik des Geschichtsbewusstseins und der Verarbeitung von Geschichte im Leben verkannt. "Geschichte" ist nicht das Wichtigste aus der Vergangenheit (ein "maßstabsgetreues und fotografisch korrektes Modell" der Vergangenheit) und nicht ein bloß kognitiver Abspeicherungsprozess; statt dessen enthält sie massive ästhetische, moralische und emotionale Anteile (auch unbewusste Triebstreben). Geschichte ist einerseits die sinnbildende und identitätssichernde (aber höchst selektive und perspektivische!) Erzählung von Erfahrungen und Veränderungen im Verlaufe der Vergangenheit und bis heute hin. Sie ist andererseits der Umgang mit, die Deutung von, die Schlußfolgerung aus diesen Wandlungen (bleibende Relevanz). Deshalb wird Geschichte - gerade Zeitgeschichte, um die es sich bei der DDR noch immer handelt - wesentlich im "kommunikativen Gedächtnis" gebildet. Die kulturelle Überlieferung - Assmann (1997) sagt "kulturelles Gedächtnis" - als "offizielle und verpflichtende Version" kann zwar ebenfalls schon eingespeichert werden (sie pflegt jedoch umstritten zu sein), vor allem aber kann eine solche Version - gegen das Alltagswissen der Schüler, zu dem auch ein beredtes kommunikatives Beschweigen zählen kann - kaum durchgesetzt werden.
- Vermutlich - Buch und Manuskript geben dazu keine direkte Auskunft - ist auch die Vor-

stellung von Geschichtswissenschaft und ihrer Funktion im Leben und in der Gesellschaft eher schlicht. Historiker glauben oft, die Gegenwart werde durch die Vergangenheit, die man deshalb möglichst genau kennen müsse, geprägt, aber zusätzlich auch durch das Geschichtsverständnis, also den bewussten Umgang mit Geschichte ("Vergangenheitsdeutung"). Das ist zwar nicht falsch, aber positivistisch-illusionär und sozialpsychologisch-naiv verkürzt. Sozial- und kulturwissenschaftliche Studien haben nämlich schon mehrfach ergeben, dass es so einfach nicht ist: Umgekehrt gilt auch (wenn nicht vorwiegend): Gegenwart bestimmt die "Geschichte", nicht umgekehrt. Aus gegenwärtigen Wahrnehmungen, Begriffen, Interessenlagen (vor allem *gefühlten* mentalen Interessen, nicht objektiv gegebenen *Vorteilen/Nachteilen*) wird die Vergangenheit und ihre Verwandlung in die Gegenwart zusammengereimt, wobei sehr schiefe und irrige Konzepte entstehen können. Das gilt merkwürdigerweise auch für die jüngere Zeitgeschichte: Mythen und Legenden entstehen sehr rasch und verfestigen sich durchaus. Jugendliche *wissen* über die Zeit ihrer Eltern verblüffend wenig und haben trotzdem feste Meinungen; jedenfalls wurde das europaweit schon 1995 für die Gorbatschow-Zeit festgestellt (v. Borries u.a. 1999).

- Eine ausgesprochen optimistische Machbarkeit-Illusion hinsichtlich des DDR-Verständnisses und der BRD-Bewertung - bei Arnswald ausdrücklich vertreten - verkennt zudem völlig die Erfahrungen in anderen Fällen. Auch 63 Jahre nach dem Ende der nationalsozialistischen Gewaltherrschaft (oder soll es, weil schon die Terminologie eine Deutung impliziert, lieber "Drittes Reich", "Deutscher Faschismus", "Hitlerismus" oder wie sonst heißen?) zeigt jede neue empirische Studie (zuletzt Zülsdorf-Kersting 2007, Meseth u.a. 2004, Hollstein u.a. 2002, Welzer u.a. 1997, 2002, Lutz 2000), dass Geschichtsbewusstsein völlig anders läuft und nicht politisch zentralistisch, aber auch nicht wissenschaftlich *anzuordnen* ist. Wen das Beispiel NS noch nicht überzeugt, der muss nur nach Rußland schauen, wo der Staat - längst aber auch viele Bürger - den Stalinismus (interessanterweise neben dem Zarenreich!) inzwischen wieder zur positiven Epoche aufgewertet und projiziert haben. Um ältere Beispiele zu nennen: Mehr als hundert Jahre hat nach dem Amerikanischen Bürgerkrieg (Secessionskrieg 1861-1865) der Süden seine Legenden und Mythen gepflegt, die sich als verhältnismäßig lernimmun erwiesen haben (vgl. Schivelbusch 2003). Für Nord- und Süditalien nach 1859/60 dürfte, obwohl - wie seitens der DDR 1989/90 - der "Anschluss" zunächst freiwillig und begeistert erfolgte, dasselbe gelten. Aufklärende Historie hat es auch

dann schwer, wenn sich - was selten ist - alle intellektuell redlichen und moralisch reifen Experten ausnahmsweise einmal weitgehend einigen können sollten.

3.2. Fehlende Einbeziehung des Forschungsstandes

Mit einer gewissen Empörung nimmt die Studie von Deutz-Schroeder/Schroeder zur Kenntnis, dass Demokratie und Diktatur keine wirklich geläufigen und gefüllten Begriffe sind (Arnswald hat das kaum gefragt). Der bisherige Forschungsstand hätte - bei näherer Kenntnis (hier muss sehr abgekürzt werden) - genau das hergegeben. Seit den sechziger Jahren (Habermas u.a. 1961, Teschner 1968) ist bekannt, dass von gefestigter Demokratie-Überzeugung nicht die Rede sein kann. Gerade bei Hauptschülern fehlten die elementarsten Kenntnisse und Konzepte (Marz u.a. 1977).

Noch in jüngster Zeit wurde festgestellt, dass - auch international - nur etwa die Hälfte der Jugendlichen "Opposition" als Merkmal von Demokratie kennt und anerkennt (Torney-Purta e.a. 2001). Schon 1992 war klar, dass von Jugendlichen der "Demokratie" als Wert nur ein sehr bescheidener Rang zuerkannt wird, übrigens kommt "Meinungsfreiheit für alle" viel besser weg (v. Borries u.a. 1995). Das gilt zudem - bei bezeichnenden Unterschieden im Detail - nicht nur in Deutschland, sondern in ganz Europa (vgl. v. Borries u.a. 1999). Wer in solchen Fällen entsetzt ist, hat über die Lebenswirklichkeit von Jugendlichen - keineswegs nur der Benachteiligten - nicht genügend nachgedacht und Empirie nicht sorgfältig verarbeitet.

Die *vorausgehenden Parallelbefunde* sind reichlich. Ironisch kann man sagen, dass Lektüre bisheriger Studien Arnswald (2006) und Deutz-Schroeder/Schroeder (2007) in vielen Fällen vor der Illusion von "Neuentdeckungen" geschützt hätte. Gerade in der empirischen Sozialforschung ist es ausgesprochen wichtig, sich im Rahmen bisheriger Fragestellungen und Befunde zu verorten und sehr genau zu präzisieren, wo man ältere Skalen wiederverwenden und Befunde replizieren will ("Stabilitätsuntersuchungen") und wo man wirklich neue Felder erstmalig untersuchen will ("Innovationen"). Nun behaupten Deutz-Schroeder/Schroeder (2007, 40), sie wollten das Geschichtsbewusstsein erstmals mit Skalen untersuchen (ähnlich Arnswald 2007, 108). Das trifft jedoch keineswegs zu.

Es ist schon erstaunlich, im Literaturverzeichnis keine einzige der berühmten und nun wirklich allgemein bekannten Shell-Jugendstudien (Jugend '97, Jugend 2000, Jugend 2002, Jugend 2006) zu finden, die sich - seit vielen Jahren - ja gerade auf den Ost-West-Vergleich konzen-

triert haben und durch die man sehr wohl vorher wissen konnte, dass Ost-Herkunft bzw. West-Herkunft nach wie vor in vielen - nicht allen! - Lebensgewohnheiten und Einstellungen durchschlagen. Auch in der Wertewelt finden sich weiterhin beachtliche Unterschiede, zudem gerade im Demokratieverständnis, im Sicherheitsstreben usw. Aus den Shell-Studien (gerade auch der neuesten: "Jugend 2006") ist aber auch die wenig ausländerfreundliche (ja zuwanderungsfeindliche) Position einer starken (wachsenden) Minderheit der deutschen Jugendlichen bekannt, die Deutz-Schroeder/Schroeder zu recht beklagen, aber leider hätten schon vorher erwarten müssen.

Speziell im Bereich des Geschichtsbewusstseins sind mehrere Befragungsstudien (aus den Jahren 1990, 1992, 1994, 2002) vorgelegt worden, zwei davon (1992, 1994) mit repräsentativen Schülerstichproben aus Bayern/Baden-Württemberg, NRW und den "neuen Bundesländern" (genauer "Neufünfeinhalbland"). Aus diesen Untersuchungen, die in umfangreichen Monografien veröffentlicht sind (v. Borries u.a. 1992, 1995, 1999, 2005) und regelhaft ost- und westdeutsche Probanden vergleichen, sind - und das teilweise seit 15 Jahren - einige Punkte empirisch gesichert und theoretisch interpretiert:

- Die Unterschiede zwischen ost- und westdeutschen Lehrer(inne)n erwiesen sich - wegen der längeren Biografie und der höheren mentalen Involviertheit ("Lebenszeitinvestition") - als deutlich kräftiger als die zwischen den ost- und westdeutschen Jugendlichen.
- Gerade beim Thema "DDR" und "BRD" fielen die Differenzen - anders als bei den meisten anderen Themen - sehr krass aus; denn hier war die "Identität" beider Seiten im Spiel. Nach einigen Jahren hatten sich die Kontraste allerdings etwas abgeschwächt.
- Es erweist sich als fair und fruchtbar, DDR-Geschichte und (Alt-)BRD-Geschichte *gemeinsam* und *symmetrisch* abzufragen, schon um die "Ossi"-Jugendlichen nicht in eine inferiore Position zu versetzen. Es ist aber auch wissenschaftlich erforderlich, weil insgeheim ohnehin immer verglichen wird (das ist eine zwingende Logik des Denkens!) und deshalb für die Interpretation auch genau diese "doppelseitigen" und "kreuzweisen" Daten verfügbar sein müssen.⁵⁷ Das nicht gesehen zu haben, ist eine elementare methodische Schwäche beider neuen Studien und entwertet sie analytisch ("Suggestivität") wie forschungsstrategisch ("Kolonialisierung der Lebenswelt"). Das ist - weil es um die Selbstachtung der ostdeutschen Bürger(innen) geht - ein sehr wichtiges Argument.

⁵⁷ Wenn man die Logik erst einmal begriffen hat, kommt man mit viel weniger Einzelfragen aus, spart also im Umfang auch wieder ein.

3.3. Detailvergleich und Länderrangreihe

Generell scheint der Eindruck entstanden zu sein, bei Arnswald (2006) würden Kenntnisse und Einstellungen hinsichtlich der DDR-Geschichte als zufriedenstellend, als in Berlin besonders günstig und als im ganzen zwischen Ost- und Westdeutschland sich angleichend dargestellt bzw. erwiesen und bei Deutz-Schroeder/Schroeder (2007) sei in allen drei Punkten das Gegenteil der Fall: das Wissen und die Wertungen seien miserabel, Berlin liege am Schluss, und Ost und West klafften weit auseinander.

Es dürfte schon deutlich geworden sein, dass eine nähere Betrachtung diese oberflächlichen Eindrücke nicht bestätigen kann. Beide Studien haben methodisch (unrepräsentative Stichprobe ohne Haupt-, Real- und berufliche Schulen, unvalides Instrument mit bloßen Kenntnis- und Gesinnungsfragen, nur univariate Auswertung) ein eher bescheidenes Niveau. Während Deutz-Schroeder/Schroeder stets Lernende mit Ostberliner und Westberliner Herkunft trennen, weist Arnswald sie niemals gesondert aus, obwohl das nach dem Urmaterial wahrscheinlich möglich wäre. In vielen Fällen untersuchen beide Studien einfach verschiedene Problemkreise bzw. gleiche Problemkreise mit recht verschiedenen - und, besonders bei Deutz-Schroeder/Schroeder, wenig glücklichen - Operationalisierungen.

- Zu den Fragen zum "politischen Neubeginn" nach 1945 (Fragen 1 bis 8) sowie zur "friedlichen Revolution" (Fragen 34 bis 41) bei Arnswald (2006, 116-119, 134-138) fehlen bei Deutz-Schroeder/Schroeder völlig die Entsprechungen.
- Umgekehrt sieht man bei Arnswald (2006) keinerlei Äquivalente zu den - mißglückten - zahlreichen, sektoralen "Bevorzugungsfragen" "**Deutschland vor 1989: Was glauben Sie war besser oder gleich in der DDR oder in der Bundesrepublik Deutschland?**" (Deutz-Schroeder/Schroeder 2007, Fr. S. 2) sowie der eigenen politischen Selbstverortung und Parteipräferenz der Jugendlichen (Deutz-Schroeder/Schroeder 2007, 49ff., Fr. S. 1).

Es gibt aber durchaus auch direkt vergleichbare Bereiche, bei denen man allerdings bevorzugt auf die Ebene von Einzelitems gehen muss, die eher zufällig ganz besonders ähnlich formuliert sind.

- Bei der Erkundung der *Sozialisationsfaktoren* wird z.B. nach der Häufigkeit von "*Gesprächen in der Familie*" (2007) bzw. dem Gewicht von "*Gesprächen mit Eltern und Freunden*" (2006) gefragt. Solche werden von 54,7% (Deutz-Schroeder/Schroeder 2007, 59) und von 56,4% (Arnswald 2006, 142) angegeben. Dieser letzte Wert gilt allerdings für Befragte aus

allen einbezogenen Ländern, nicht nur - wie der erste - für Berlin.

- Beide Studien beziehen die Behandlung der DDR-Geschichte im *Schulunterricht* ein. Bei Arnswald (2006, 115) haben 69,0% das Thema in der 10. Klasse und 11,9% in der Sekundarstufe II behandelt. Fast 80% finden die Sache wichtig, wollen mehr Unterrichtszeit dafür und eine "*Verbindung mit der BRD-Geschichte*" (Arnswald 2006, 139f.). Bei Deutz-Schroeder/Schroeder (2007, 53) haben nur 21,8% DDR-Geschichte "überhaupt nicht" behandelt (bei den unter Sechzehnjährigen noch 36,6%, bei denen ab Siebzehn nur 13,7%). Fast die Hälfte glaubt allerdings, dass das zu wenig war, und fast zwei Drittel wollen "*gerne mehr über das Leben der Menschen in der DDR erfahren*" (Deutz-Schroeder/Schroeder 2007, 53, 56). Das ist eine gute - wenn auch ziemlich allgemeine - Übereinstimmung, vor allen auch hinsichtlich offenbar "*immuner*" Minderheiten um 20 bis 30%.
- Bei den *Einstellungen* wird u.a. nach der Legitimität der Grenz-Absperrung und des Schießbefehls gefragt: Dass die Flüchtlinge sich ihre Gefährdung eigentlich selbst zuzuschreiben hatten ("*Wer die DDR verlassen wollte, hatte selbst Schuld, wenn an der Grenze auf ihn geschossen wurde*"), unterstellen 20% bei 72% Ablehnungen und 9% Enthaltungen (Deutz-Schroeder/Schroeder 2007, 236). In ziemlich anderen Formulierungen (Arnswald 2006, 123, 135) zählen knapp 30% "*Mauer und Stacheldraht*" bzw. Ausreiseverbot *nicht* zu den Hauptpunkten des "*Kalten Krieges*" und *nicht* zu den Hauptgründen der "*Friedlichen Revolution*". Mit anderen Worten: In beiden Studien ist für etwa drei Viertel die Sache "Mauer, Stacheldraht und Schießbefehl" ausgesprochen *wichtig* und *verurteilenswert* (eine ziemlich vage Übereinstimmung).
- Deutlich ähnlicher sind die Formulierungen beim "*Sozialismus als guter Idee, die nur schlecht ausgeführt wurde*". Bei Deutz-Schroeder/Schroeder (2007, 256) stimmen (in Berlin) 37,6% zu, bei 30,4% Ablehnungen und 32,0% Enthaltungen. Bei Arnswald (2006, 123) äußern sich insgesamt ebenfalls 37,7% positiv und 55,5% negativ (hier gibt es - wegen Vierstufigkeit - keine Enthaltungen und nur 6,7% Auslassungen). In Berlin ist die Zustimmung allerdings mit 54,1% viel höher (Arnswald 2006, 161). Der höhere Wert (für Berlin) ist umso erstaunlicher, als der - vermutlich überhaupt nicht attraktive - Zusatz "*sowjetisches Modell*" nur bei Arnswald steht. Hier dürfte das *Verbot* der Enthaltung durchgeschlagen haben: Entdecken die Unentschiedenen vielleicht ein leises Prä für "Sozialismus - als soziale Gerechtigkeit und Sicherheit"?

- Wichtig ist natürlich auch die Stellungnahme zur *Wiedervereinigung*: Bei Deutz-Schroeder/Schroeder (2007, 260, 261) kreuzen 66,1% an: "*Ich finde es gut, dass die DDR durch eine friedliche Revolution ihr Ende fand*",⁵⁸ und 78,7% finden: "*Auch wenn der Aufbau Ost viel Geld kostet, bin ich froh, dass Deutschland wieder vereint ist*" (jeweils etwas höhere Werte im Ostteil der Stadt!). Eine "*Wiederrichtung der Mauer*" sehen bei Arnswald (2006, 137) nur 20,4% der Befragten als "*gut für die deutsche Entwicklung*" an; nur 28,9% halten dadurch "*die Interessen der Menschen im Osten und Westen*" für gefördert (Angaben speziell über Berlin fehlen). Das sieht zwar einigermaßen ähnlich aus, sagt aber nichts über Feinstrukturen und ist aus anderen Befragungen ohnehin bekannt.⁵⁹ Auch die Wiedervereinigung war halt - für die meisten - "*eine gute Sache, die nur (teilweise!) schlecht durchgeführt wurde*".
- Als Beispiel für die *Kenntnisse* sind in beiden Fällen *Wilhelm Pieck* und *Ludwig Erhard* gewählt. Allerdings genügt es einmal (Deutz-Schroeder/Schroeder 2007), sie in DDR oder BRD einzusortieren, das andere Mal (Arnswald 2006) sind ihnen ihre besonderen Funktionen zuzuordnen. Die erste Operation gelingt in Berlin bei Pieck 32,0% und bei Erhard 26,1% der Befragten richtig (Deutz-Schroeder/Schroeder 2007, 296, 290), die zweite bei Pieck 33,4% und bei Erhard 66,5% (Arnswald 2006, 158). Die Werte für alle Befragten weichen bei Pieck etwas nach unten (27,2), bei Erhard etwas nach oben (70,0) ab (Arnswald 2006, 158). Für die Bewertung der Zuverlässigkeit der Studien wie für die Beschreibung tatsächlicher Verhältnisse im Lande ist das nicht sehr hilfreich.

Genug der Einzelbeispiele, die zureichend belegen, dass die Ansätze zu verschieden sind, um im Detail Bestätigungen zu erwarten. Statt dessen kommt es auf eine Gesamteinschätzung an, die hier nur für die Kenntnisdimension - und danach die Einstellungsdifferenzen - noch einmal kurz erläutert werden soll. Beide Studien beschwerten sich zwar lebhaft über beschämende Kenntnisse und unzureichende unterrichtliche Behandlung, machen aber nicht einmal einen ernsthaften Versuch der generellen Quantifizierung und der gemeinsamen Skalierung (von einem Vergleich mit den Erwachsenen ganz abgesehen).

Natürlich kann man das an ihrer Stelle nachholen, *wenn auch leider nur ohne - an sich uner-*

⁵⁸ Es ist sehr bezeichnend für den affirmativen Charakter beider Studien, dass an eine Unterscheidung von "DDR-Wende" und "Wiedervereinigung" kein Gedanke verschwendet wird. Über die Kerngruppen der Bürgerbewegung, die "eine andere DDR" und "einen anderen Sozialismus" wollten, ist die Entwicklung - wegen Mehrheitsentscheidung - bekanntlich hinweggegangen; die Historie aber darf das eigentlich nicht.

⁵⁹ Dass es sich hier nicht um "historische Kenntnisse, Analysen und Einsichten" - und auch nur bedingt um "Geschichtsbewusstsein" - handelt, ist offenkundig.

lässliche - Konsistenzprüfung (z.B. selbst bei validen Aufgaben wichtig für die Frage der Ein- oder Mehrdimensionalität). Die 18 Fragen bei Deutz-Schroeder/Schroeder (2007, 269-302) werden im Mittel in Berlin zu 44,6% "richtig", *d.h. im Sinne der Fragebogenkonstrukteure*, beantwortet (Westberliner 46,5% und Ostberliner 44,1%). Die 20 Fragen bei Arnswald (2006, 158f.) werden im Mittel von 55,3% der Berliner "richtig" beantwortet, aber nur von 44,4% der Westdeutschen (NRW, Bayern) und 48,8% der Ostdeutschen (Brandenburg).

Das sagt über die Perfektion des Lernens zur DDR-Geschichte irgendwo überhaupt nichts Zuverlässiges und Gültiges aus, denn man kann ja - zufällig oder absichtlich - schwere oder leichte Fragen gewählt haben. Im Ost-West-Vergleich zeigt sich, dass bei Deutz-Schroeder/Schroeder der Osten (jedenfalls Berlins), bei Arnswald der Westen (nach Bundesländern) besser abschneidet. Bei Arnswald sind die Berliner besonders stark, bei Deutz-Schroeder/Schroeder - das deuten sie nur an (Pressekonferenz, vgl. Süddeutsche Zeitung, 28.12.2007, S. 7), führen es aber noch nicht aus - besonders schwach (wenn auch etwas besser als Brandenburger).

Da beide Itemlisten vorwiegend ganz verschiedene - teilweise zudem einseitige - Fragen enthalten, auch - jedenfalls in meinem Verständnis - *weder als valide noch als reliabel* anzusehen sind, braucht man sich speziell über die - zur Zeit ungeklärte - Rangreihung der Länder "keinen Kopf" zu machen. Wenn man die wirklich unbedingt zuverlässig wissen möchte, könnte man höchstens einmal eine wirklich repräsentative (Hauptschule!) professionelle (Theoriegrundlage und Operationalisierung!) und elaborierte (Analyseverfahren!) Studie anfertigen lassen; die allerdings braucht einen Vorlauf an Zeit, Theorie und Geld.

Für die Einstellungsdimension ließe sich Ähnliches zeigen. Es ist offenkundig, dass Arnswald mehr Items erwischt (vielleicht auch schon gesucht?) hat, die eine generelle Ähnlichkeit - fälschlich mit einem Prozessbegriff "Annäherung" genannt - im Osten und im Westen - wenn auch auf bescheidenem Niveau - belegen. Darüber werden einzelne Bereiche mit ziemlich krasen Unterschieden aber durchaus nicht verkannt bzw. weggeleugnet. Wirklich spannend wäre auch die Bildung einer Typologie (z.B. durch Clusteranalysen).

Deutz-Schroeder/Schroeder üben sich eher im "whistle-blowing", im Alarmruf wegen einer für sie drohenden, sehr spezifischen Bildungskatastrophe. Ihre Items - großenteils zugegebenermaßen mehr zu Sympathie-Projektionsflächen geeignet als mittels Kenntnisständen und Deutungsleistungen beantwortbar - legen mehr spezifische Ost-West-Unterschiede bloß, wiederum bei großen Enttäuschungen über Wissen und Meinungen beider Seiten. Ein wirklicher Wider-

spruch zwischen beiden Untersuchungen besteht nicht, so wenig wie eine professionelle und elaborierte Methode.

3.4. Schlussbemerkung zur "DDR-Aufarbeitung"

Damit soll nicht bestritten werden, dass die Aufarbeitung der DDR-Vergangenheit, mehr noch in der Gesellschaft als in der Schule, sicherlich sehr im Argen liegt. Es gibt - wie es das auch schon 1992 und 1994 gab (vgl. v. Borries u.a. 1995, 1999) - in Teilen der Bevölkerung im Osten eine beachtliche "*DDR-Nostalgie*", wenn sie auch vermutlich bei weitem nicht das Ausmaß der "*NS-Nostalgie*" weiter Kreise in den Fünfzigern erreicht, das bekanntlich damals ebenfalls durch die frühe Demoskopie immer wieder belegt wurde (und nicht wirklich überwunden, aber ausgehalten wurde). Diese *nachgetragene Liebe* (viele möchten die DDR nicht "schlechtgemacht" sehen, aber praktisch niemand will sie "zurück"; das ist günstiger als in einigen Nachbarländern!) ist ein kulturell-mentales Problem.

Übrigens gibt es auch in weiten Kreisen im Westen ein anhaltendes Desinteresse am Osten (belegt z.B. durch die asymmetrischen Reisezahlen) und einen anhaltenden Überlegenheitsdünkel über die ostdeutschen Mitbürger(innen). Eben deshalb muss man die gerade getroffene Feststellung zunächst ergänzen, dass gleichzeitig mit der DDR-Aufarbeitung - in der Grundanlage, nicht in den Durchführungsdetails - *symmetrisch* dazu auch die Aufarbeitung der BRD-Vergangenheit zur Debatte steht. Das ist zudem eine überaus sensible Frage der Selbstachtung im Osten. Noch einmal: Schon die absolut einseitige Art der beiden untersuchten Befragungen halte ich persönlich für pragmatisch unklug (s.o.) und moralisch unsensibel. Das hat übrigens nichts mit etwaiger "Äquidistanz" zu tun.

Es kann sein, dass man - auch, nicht nur - im Schulunterricht (Politik, Geschichte, Ethik) nachzusteuern versuchen sollte.⁶⁰ Es ist aber eine arge Illusion und eine völlige Überforderung, eine solche Aufgabe dem Geschichtsunterricht (mit seinen Nachbarfächern) allein oder vorwiegend übertragen zu wollen oder von neuen Curricula Wunderdinge zu erwarten. Alle Belege sprechen dafür, dass man im Geschichtsunterricht weder "*Identitäten stiften*" noch "*Mythen/Legenden destruieren*" kann. Man kann nur - und das gerade vorwiegend an Beispielen, die nicht zu direkt die Bequemlichkeit und Sicherheit der Lebenswelt der Lernenden stören - im Gelingenfall die Kompetenz zum "historischen Denken" herausfordern und fördern.

⁶⁰ Das verlangt eine eigene Debatte, die hier nicht im einzelnen zu führen ist.

Erwähnte Literatur

- Angvik**, Magne and **Borries**, Bodo v. (Eds.): YOUTH and HISTORY. A Comparative European Survey on Historical Consciousness and Political Attitudes among Adolescents. Volume A: Description, Volume B: Documentation (containing the Database on CD-ROM); Hamburg (edition Körber-Stiftung) 1997.
- Arnswald**, Ulrich: Schülerbefragung 2005 zur DDR-Geschichte, in: Arnswald, Ulrich u.a. (Hrsg.): DDR-Geschichte im Unterricht. Schulbuchanalyse - Schülerbefragung - Modellcurriculum; Berlin (Metropol) 2006, 107-176 (mit CD-ROM).
- Assmann**, Jan: Das kulturelle Gedächtnis. Schrift, Erinnerung und politische Identität in frühen Hochkulturen; München (Beck) 2. Aufl. 1997.
- Borries**, Bodo v. (unter Mitarbeit von **Dähn**, Susanne, **Körber**, Andreas und **Lehmann**, Rainer H.): Kindlich-jugendliche Geschichtsverarbeitung in West- und Ostdeutschland 1990. Ein empirischer Vergleich; Pfaffenweiler (Centaurus) 1992 (= Geschichtsdidaktik. Studien, Materialien. Neue Folge 8).
- Borries**, Bodo v. (unter Mitarbeit von **Weidemann**, Sigrid, **Baeck**, Oliver, **Grzeskowiak**, Sylwia und **Körber**, Andreas): Das Geschichtsbewußtsein Jugendlicher. Erste repräsentative Untersuchung über Vergangenheitsdeutungen, Gegenwartswahrnehmungen und Zukunftserwartungen in Ost- und Westdeutschland; Weinheim/München (Juventa) 1995 (= Jugendforschung).
- Borries**, Bodo v. (unter Mitarbeit von **Körber**, Andreas, **Baeck**, Oliver und **Kindervater**, Angela): Jugend und Geschichte. Ein europäischer Kulturvergleich aus deutscher Sicht; Opladen (Leske & Budrich) 1999 (= Schule und Gesellschaft 21).
- Borries**, Bodo v., **Fischer**, Claudia, **Leutner-Ramme**, Sibylla und **Meyer-Hamme**, Johannes: Schulbuchverständnis, Richtlinienbenutzung und Reflexionsprozesse im Geschichtsunterricht. Eine qualitativ-quantitative Schüler- und Lehrerbefragung im Deutschsprachigen Bildungswesen 2002; Neuried (ars una) 2005 (= Bayerische Studien zur Geschichtsdidaktik 9).
- Cohen**, Jacob: Statistical Power Analysis for the Behavioral Sciences; Hillsdale, New York (Earlbaum) 1988.
- Deutz-Schroeder**, Monika und **Schroeder**, Klaus: Abschlussbericht Das DDR-Bild von Schülern in Berlin; Berlin (Arbeitspapiere des Forschungsverbundes SED-Staat Nr. 38/2007) November 2007.
- Habermas**, Jürgen u.a.: Student und Politik. Eine soziologische Untersuchung zum politischen Bewußtsein Frankfurter Studenten; Neuwied und Berlin (Luchterhand) 1961.
- Hollstein**, Oliver u.a.: Nationalsozialismus im Geschichtsunterricht. Beobachtungen unterrichtlicher Kommunikation. Bericht zu einer Pilotstudie; Frankfurt/M. (Fachbereich Erziehungswissenschaft der Johann Wolfgang Goethe-Universität) 2002 (= Frankfurter Beiträge zur Erziehungswissenschaft 3).
- IBM-Jugendstudie '92**. Die selbstbewußte Jugend. Orientierungen und Perspektiven zwei Jahre nach der Wiedervereinigung, hrsg. vom Institut für Empirische Psychologie; Köln (Bund-Verlag) 1992.
- IBM Jugendstudie**. "Wir sind o.k.!" Stimmungen, Einstellungen, Orientierungen der Jugend in den 90er Jahren. hrsg. vom Institut für empirische Psychologie; Köln (Bund-Verlag) 1995.
- Jeismann**, Karl-Ernst: Geschichte als Horizont der Gegenwart; Paderborn (Schöningh) 1985.
- Jeismann**, Karl-Ernst: Geschichte und Bildung. Beiträge zur Geschichtsdidaktik und zur Historischen Bildungsforschung; Paderborn (Schöningh) 2000.
- Jugend '97**. Zukunftsperspektiven, gesellschaftliches Engagement, politische Orientierungen, hrsg. vom Jugendwerk der Deutschen Shell; Opladen (Leske) 1997.
- Jugend 2000**. 13. Shell Jugendstudie, 2 Bde.; Opladen (Leske & Budrich) 2000.
- Jugend 2002**. Zwischen pragmatischem Idealismus und robustem Materialismus, 14. Shell Jugendstudie; Frankfurt/M. (Fischer) 2002.
- Jugend 2006**. Eine pragmatische Generation unter Druck (15. Shell Jugendstudie), hrsg. von Shell Deutschland Holding; Frankfurt/M. (Fischer) 2006.
- Körber**, Andreas, **Schreiber**, Waltraud und **Schöner**, Alexander (Hrsg.): Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik; Neuried (ars una) 2007 (= Kompetenzen: Grundlagen - Entwicklung - Förderung, Bd. 2).
- Lucas**, Friedrich: Ein Fingerhut voll Erkenntnis, in: Wiesemüller, Gerhard: Unbewältigte Vergangenheit - überwältigende Gegenwart?; Stuttgart (Klett) 1972, 9-25.
- Lutz**, Felix Philipp: Das Geschichtsbewußtsein der Deutschen. Grundlagen der politischen Kultur in Ost und West; Köln & Wien (Böhlau) 2000.
- Marz**, Fritz u.a.: Bestandsaufnahme Sozialkundeunterricht; Stuttgart (Metzler) 1977.
- Meseth**, Wolfgang u.a. (Hrsg.): Schule und Nationalsozialismus. Ansprüche und Grenzen des Geschichtsunterrichts; Frankfurt/M. u.a. (Campus) 2004.

- Rüsen, Jörn:** Historische Orientierung. Über die Arbeit des Geschichtsbewußtseins, sich in der Zeit zurechtzufinden; Köln (Böhlau) 1994.
- Rüsen, Jörn:** Historisches Lernen. Grundlagen und Paradigmen; Köln (Böhlau) 1994.
- Schivelbusch, Wolfgang:** Die Kultur der Niederlage. Der amerikanische Süden 1865 Frankreich 1871 Deutschland 1918; Frankfurt/M. (Fischer) 2003.
- Schreiber, Waltraud (Moderat.):** FUER Geschichtsbewusstsein. Ein internationales geschichtsdidaktisches Forschungsprojekt zum Geschichtsunterricht (= Zeitschrift für Geschichtsdidaktik, 2. Jg. 2003, Jahresband).
- Schreiber, Waltraud, Körber, Andreas, Borries, Bodo v., Krammer, Reinhard, Leutner-Ramme, Sibylla, Mebus, Sylvia, Schöner, Alexander und Ziegler, Béatrice:** Historisches Denken. Ein Kompetenz-Strukturmodell; Neuried (ars una) 2006 (= Kompetenzen: Grundlagen - Entwicklung - Förderung, Bd. 1).
- Teschner, Manfred:** Politik und Gesellschaft im Unterricht; Frankfurt/M. (EVA) 1968.
- Torney-Purta, Judith, Lehmann, Rainer, Oswald, Hans and Schulz, Wolfram:** Citizenship and Education in Twenty-eight Countries. Civic Knowledge and Engagement at Age Fourteen; Amsterdam (IEA) 2001.
- Welzer, Harald u.a.:** "Was wir für böse Menschen sind!" Der Nationalsozialismus im Gespräch zwischen den Generationen; Tübingen (edition diskord) 1997.
- Welzer, Harald u.a.:** "Opa war kein Nazi". Nationalsozialismus und Holocaust im Familiengedächtnis; Frankfurt/M. (Fischer) 2002 (3. Aufl. 2002).
- Zülsdorf-Kersting, Meik:** Sechzig Jahre danach: Jugendliche und der Holocaust. Eine Studie zur geschichtskulturellen Sozialisation; Berlin (Lit) 2007.