University of Wisconsin-Madison

IRP Discussion Papers

Glen G. Cain
Arthur S. Goldberger

PUBLIC AND PRIVATE SCHOOLS REVISITED

DP #724-83

Public and Private Schools Revisited

Glen G. Cain
Department of Economics and
Institute for Research on Poverty
University of Wisconsin-Madison

Arthur S. Goldberger
Department of Economics and
Institute for Research on Poverty
University of Wisconsin-Madison

May 1983

Cain's work was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the U.S. Department of Health and Human Services pursuant to the Economic Opportunity Act of 1964. Goldberger's work was supported by the William F. Vilas Trust Estate. The authors are grateful to Robert Crain, Betty Evanson, Sally Kilgore, Peter Mueser, Michael Olneck, Aaron Pallas, Aage Sørensen, and Douglas Willms for help in preparing the paper. The opinions and findings expressed are solely those of the authors.

ABSTRACT

We reconsider the causal analysis of cognitive outcomes by Coleman, Hoffer, and Kilgore in the light of their reply to our earlier criticism, both of which were published in Sociology of Education (April/July 1982). Our previous conclusions are reinforced. The statistical methods, reporting style, and mode of inference of Coleman, Hoffer, and Kilgore fall below the minimum standards for social scientific research. Their conclusions about the virtues of private education and efficacy of educational policies are not warranted by their empirical evidence.

PUBLIC AND PRIVATE SCHOOLS REVISITED

1. Introduction

Our dismay at the quality of Coleman, Hoffer, and Kilgore's rejoinder to our critique of their study compels us to respond. We do so to provide guidance for those who share our concern that quantitative methods be used competently and responsibly in social science.

Our critique (Goldberger and Cain, 1982; henceforth GC) focused on two reports by Coleman, Hoffer, and Kilgore (henceforth CHK): 1981a (the Draft Report, henceforth DR) and 1982a (lead article in Sociology of Education, henceforth SE1). CHK's rejoinder was contained in 1982b (henceforth SE2). We will also refer to the other publications that present their analyses of the 1980 High School and Beyond sample: 1981b (their Harvard Educational Review article, henceforth HER), 1981c (the Final Report, henceforth FR), and 1982c (the book that resulted, henceforth BB).

The discussion that follows covers a range of issues: the reliability and validity of test scores, the appropriateness of the regression models employed and statistical inferences drawn, the extent of selectivity bias in background controls, the validity of the "common school" hypothesis, the use of dropout rates to adjust sophomore-to-senior changes, and, finally, the attribution of causality to what are regarded as "school policies." To make the discussion reasonably self-contained, we will be restating the main elements of our critique and of CHK's rejoinder.

2. Sectors and Test Scores

We argued that the non-Catholic private school subsample was too small, heterogeneous, and erratically collected to permit reliable inferences. CHK express agreement: "Generalizations to sectors as a whole should be limited to comparisons between public and Catholic sectors" (SE2:169). But throughout BB and FR, CHK still treat that "other private" sector on a par with the public and Catholic sectors, despite their initial caveats (FR: 14-15, BB: 12-13). Here, with rare exceptions, we will confine attention to the public and Catholic sectors.

The main dependent variables in CHK's analysis were the scores on short subtests of reading, vocabulary, and mathematics (R, V, and M). Because of the brevity of the subtests, we questioned their reliability; because of their elementary content, we questioned their validity as measures of high school achievement. We also questioned CHK's reliance on the sophomore-senior differences in these scores as a measure of two years of educational achievement in high school.

In their rejoinder, CHK provide no fresh information on the content of the subtests, nor on the translation of sophomore-senior differences into school-year equivalents. They do summarize (SE2: 164-165) their own new computations (FR: 196-198), along with analyses by Heyns and Hilton (1982, henceforth HH), to support their contention that the full tests show the same, or slightly larger, Catholic sector effects than the subtests did. Incidentally, there is an apparent error in CHK's report of HH's estimates of the Catholic sector effects (SE2: 164, Table 1, rows 3 and 11). Instead of reporting HH's estimates for the interactive model when evaluated at public sophomore means, CHK report estimates obtained when zero values are assigned to all 17 background control variables. The latter estimates, which are not comparable to any others reported in CHK's Table 1, tend to be considerably larger than those obtained with mean values.

CHK make no use of tests that are relatively specific to the high school curriculum, such as science, civics, and advanced mathematics, on the grounds that such tests are inappropriate for the sophomores. Other researchers appear to consider them appropriate. For example, HH(101), who use them, find sector effects similar to those advanced by CHK. On the other hand, Fetters et al. (1981) and Willms (1982b) find smaller (sometimes zero) Catholic sector effects for the curriculum-specific tests.

CHK say (SE2: 164) that HH's results on the statistical properties "generally justify" the use of the subtests. A balanced summary of HH's assessment would include the following, less sanguine, comments about the tests:

> Being measures obtained at one point of time, they tell us little about cognitive growth during high school. In fact, they may tell us more about the admission requirements of private schools than about their differential effectiveness (HH:95).
>
> [T]he conclusions regarding differential effects by sector rest on the inherently ambiguous issue of whether the tests measure achievement or ability differences among students (HH:97).

3. Statistical Inference

We observed that DR gave no standard errors at all, and that those given in SE1 were merely nominal ones in that they took account neither of the clustered sample design nor of missing-value imputations. As a rough guide, which might serve until proper measures of reliability were reported, we proposed that readers use a ± 3-sigma, or ± 4-sigma, rule for assessing significance in place of the customary ± 2-sigma rule. That is, we advised readers to multiply by 1.5, or by 2.0, to convert a nominal standard error into an approximately correct standard error. (The 1.5 conversion factor had previously been used by Fetters et al., 1981).

CHK respond (SE2: 169) by announcing that FR and BB contain the nominal standard errors "for all estimates," along with corrected standard errors for "one especially critical set of comparisons." Hence, they say, our rule of thumb is "quite unnecessary."

Now, that "especially critical set of comparisons" consists of the 12 estimated increments in test scores (3 tests x 2 grades x 2 sectors) attributable to enrollment in the private sector rather than the public sector. For those 12 estimated increments, the ratios of correct standard errors to nominal standard errors, are, as reported by CHK (FR: A8):

Sector/Grade	Test		
	R	V	M
Catholic Sophomores	1.04	2.07	1.20
Catholic Seniors	2.12	1.39	1.16
Other Private Sophomores	2.14	2.94	3.31
Other Private Seniors	3.21	3.40	2.25

CHK themselves summarize the twelve ratios as averaging 1.5 for the Catholic sector, and 3.0 for the other private sector (FR: 16). Obviously, our 1.5 or 2.0 proposal was a reasonably good guide for those twelve cases. More importantly, except for that dozen of the hundreds of regression-derived point estimates in BB and FR, only nominal standard errors are presented. The need for a conversion rule persists. Since CHK offer no alternative, we stick with our rule of thumb.

Further, FR and BB do not contain standard errors "for all estimates." No measures of reliability have yet been produced for two other sets of comparisons, which are equally critical in CHK's analyses, namely the sophomore-to-senior growth rates by sector, and the contributions of "school policy" variables to sector differences in achievement. (For the latter set, at least, the omission is explained by the fact that CHK do not know how to obtain the measures: see the footnote, FR: 251.)

4. Additive versus Nonadditive Models

In DR, the estimates of sector effects were obtained from regressions of test scores on background variables, fitted separately to each sector. We observed that this insistence on a nonadditive model (in which all the regression coefficients are allowed to differ across the sectors; that is, in which a full set of sector x background interactions is specified) was not accompanied by evidence that an additive model (in which only the intercept is allowed to differ across the sectors) was inadequate to fit the data. (We recognized that the additive model usually showed smaller Catholic sector effects.)

CHK have yet to provide such evidence. In SE2 (166, 169) they purport to do so via a "Chow test." But what they describe is not the Chow test for equality of regression coefficients. It is, instead, a test for equality of residual variances and is simply irrelevant to the issue at hand--namely, are there real differences in regression coefficients across the sectors?

In view of the heavy weight which CHK place on differential background effects throughout their publications, it is astonishing that they left the reliability of those differences up in the air. None of their publications to date provides the information needed to carry out a proper test of the joint null hypothesis that all regression slopes are the same across sectors. As is well-known, the conventional test statistic is calculable from the error sums of squares of one pooled, and two separate, regressions, and is expressible in terms of the increment to R^2. In December 1982, Peter Mueser of the Graduate School of Business at the University of Chicago did the relevant computation, the first to do so, as far as we know. According to the material provided us by Mueser (personal correspondence), the relevant R^2's and associated F-statistics are:

	Sophomores			Seniors		
	Pooled	Separated	F-Statistic	Pooled	Separated	F-Statistic
R	.189	.191	2.8	.188	.190	4.8
V	.218	.219	1.8	.237	.238	2.8
M	.253	.255	4.7	.260	.262	4.4

At the 5% level, the relevant critical value is $F(17, \infty) = 1.6$, so that the increments to R^2, while modest, would indicate significance for CHK's sector interactions. However, an adjustment for clustering and missing value imputations is again required. The easiest way to do that adjustment is to multiply the critical value by $2.25 = (1.5)^2$. When this is done, only three of the six increments retain significance. We conclude that CHK's heavy reliance on differential background effects lacks empirical justification.

McPartland and McDill (1982; henceforth MM), working with a simplified version of the test-score-on-background regressions, did publish a conventional Chow, or F, test of the hypothesis that coefficients are constant across sectors. MM found that in most cases the increment to R^2 was nonsignificant; that is, that the additive model was adequate. (Allowance for the design effect would strengthen that conclusion.) CHK (SE2: 163-164) announce that the MM procedure is wrong because "the private sector constitutes such a small fraction of the total sample." Reference to a treatise on regression analysis or, for that matter, to an elementary textbook on statistics will show that CHK, not MM, are wrong. For it appears that CHK want to cancel the role of small sample size in the determination of statistical significance. Indeed, their reasoning may be paraphrased as "When the sample size is small, any difference is real."

The available evidence, therefore, does not justify CHK's contention that estimates of sector effects obtained from additive models can be dismissed out of hand as "misspecified" (SE2: 160). It would be useful for researchers who work with these data to display results from competing specifications to gauge the robustness of inferences regarding sector effects and to suggest hypotheses deserving further examination.

For the nonadditive model, we recommended (GC: 109) that sector effects be evaluated at several reference points. We illustrated the recommendation by calculating sector effects with an average private school sophomore as reference point, contrasting the numbers with those in DR, which used an average public school sophomore as reference point. (Our calculation was relevant to the question: If students in the public sector, having the same background characteristics as the average student in the private sector, switch to the private sector, what will their test scores become? Presumably, public sector students who are on the margin of transferring are of this type. See Olneck, 1981, for further discussion.) CHK admonish (SE2: 169-170) that our illustration is misleading because it does not incorporate a third, "symmetric," reference point. However, when they face the same problem with a nonadditive model elsewhere in their work, they now use the two, rather than the three, reference points (FR: 256-260).

5. Background Controls and Selectivity Bias

We argued at some length that CHK's background-controlled estimates of sector effects are contaminated by selectivity bias. Since unmeasured determinants of sector choice are presumptively correlated with unmeasured determinants of test performance, the attribution of causal effects to the sectors is suspect. Specifically, the evidence now in hand shows that the background control variables that have a net positive relation to test scores (e.g., parents' education) also tend to have a net positive effect on enrollment (selection) into private schools. Thus, a more complete specification of the control variables that represent an advantaged background will tend to reduce the estimated private-sector effect.

Consider the following points:

(1) When CHK use their "long list" of 17 background-control variables instead of a "short list" of 5, the Catholic sector effects are reduced by about one-half (calculations from FR: 192, A-32).

(2) When Noell (1982) adds four background variables--region, sex, handicap, and college expectations in the 8th grade--to CHK's 17, he substantially reduces the Catholic-sector effect. CHK show a modest reduction when they add two of Noell's variables (SE2: 164, Table 1, rows 1 and 2). Incidentally, being a male and being in a small-sized family are exceptional cases of background characteristics that are apparently

positively related to enrollment in the public sector (relative to enrollment in the Catholic sector) and positively related to test scores. (We say "apparently" because we have access only to the simple correlation matrix.)

(3) Other investigators have found that using school-level averages of the background variables substantially reduces the Catholic sector effect (Crain and Ferrer, 1982; Murnane, 1982). Their interpretations are that the school-level averages control for the background context of the student's peers (see also MM, 1982) and that the school-level averages correct for measurement errors in the individual student's background variables--points we had mentioned in passing (GC: 120). CHK (SE2: 164) dismiss the Crain and Ferrer study because its means, standard deviations, and regression results are different from those in Kilgore's dissertation (1982), which also uses school-level averages. Now, Crain and Ferrer used the "formula score" for the full tests, which accounts for the differences in means and standard deviations. We conjecture that when Kilgore's school-level calculations are redone with the same 17 background control variables as Crain and Ferrer used, they will show Catholic sector effects about half the size of CHK's individual-level estimates. (Crain and Ferrer, using an additive model, show an even larger reduction.)

(4) We noted that curricular track, a variable ignored in DR, serves as a proxy for the background academic orientation of the student. We conjectured (GC: 110-111) that among students of comparable measured background, (i) the academically oriented would be enrolled in the private sectors and (ii) those on the academic track would do well on cognitive tests. Consequently, we conjectured that (iii) adding track to the

list of explanatory variables would reduce the CHK estimates of private sector effects (as Willms' empirical study (1982a) suggested).

On our reading, all three of our conjectures have been verified by CHK. With respect to (i), they say "Controlling for family background, seniors in a Catholic school were 25 percent more likely to be in an academic program than were public school seniors" (SE2: 171); for sophomores, the corresponding figure is 21 percent (FR: 264). With respect to (ii), readers can made the relevant calculation from the table in HER (532, n.8). With respect to (iii), CHK write that "carrying out separate analyses for achievement in academic and general programs ... in public and Catholic schools ... show(s) generally a reduction in estimated effects ... We find the effects in the academic program to be rather sharply reduced ... but we find little reduction of the effect for students in the general program" (SE2: 171).

To be sure, CHK's interpretation of these empirical relationships between track and achievement is different from ours. Believing that sector determines track, they assert that "the academic program in Catholic schools contains in it students who would be in a general or vocational program in a public school," and that "Catholic schools ... place more students into academic programs who would be, if they attended public schools, in a general or vocational program" (SE2: 171). But these assertions are quite uncalled for: CHK have no independent information on the process by which students choose tracks.

CHK's striking new finding (SE2: 171, FR: 261-268), that Catholic schools are particularly effective for general-track students, might best be put on hold. While 32% of the students in Catholic schools report that they are in the general track, the administrators in the same schools report that only 19% are. Some attention to classification error is required. Whether or not the self-reported classification is correct, close attention should be given to the initial achievement levels of the minority of Catholic school students in the general track relative to the 42% of public school students in the general track. As we have indicated, there is a presumption that the private sector tends to select more academically oriented students.

(5) As one device to decontaminate the sector effects, we proposed that James Heckman's econometric approach to selectivity bias adjustment, unmentioned in DR, be implemented. CHK now tell us (SE2:172) that they have "done some modeling using Heckman's technique," referring to their other publications for the details. Pursuing the references, we learn that the econometric approach was applied to the sophomore mathematics tests, that it produced results that were "not reasonable," and that it raised rather than lowered the estimate of the Catholic sector effect (FR: 214-217; BB: 213-214).

There is no indication that CHK applied Heckman's technique to the reading and vocabulary tests, and/or to seniors. For the sophomore mathematics tests, they report numerical results only for the first (probit) step of the econometric approach, not for the second (regression) step. Now, CHK's discussion of selectivity (FR: 214-217) is extremely unsettling. First, they suggest that selectivity bias arises only when a nonrandom subsample is observed and is unlikely to occur when the full sample is observed. In fact, many scholars have recognized the selectivity associated with nonrandom assignment to treatment and control groups, and have used the econometric approach to isolate treatment effects when both groups are fully observed. Second, they announce that selectivity bias may arise when the full sample is observed, if separate equations are fitted to the two groups, but not if a single equation is fitted to both groups. In fact, the issue of selectivity bias is entirely distinct from the issue of nonadditive effects. Third, they say that the additional explanatory variable introduced in the second step of Heckman's technique is "a term representing the probability of the private sector," explaining that "the inverse of this quantity is technically known as Mill's ratio." In fact, while the additional variable may be the inverse of Mill's ratio, it is a mean and not a probability. Frankly, we find CHK's discussion so garbled and their reporting of results so inadequate as to leave us doubting that they have carried out the correct arithmetic for Heckman's technique.

6. The Common School

We were skeptical, to put it mildly, of CHK's conviction that the causal linkage between socioeconomic status (SES) and test scores is weaker in the Catholic sector than in the public sector. On our reading of the Draft Report, that "common school" conclusion emerged from regression analyses in which (i) twelve of the measured background variables had been omitted (thus distorting the coefficients on the remaining five variables which CHK had picked to represent SES), and (ii) no measures of reliability had been obtained (thus leaving open the possibility that the observed differences were due to mere chance). We also suggested (iii) that selection into the Catholic sector was partly determined by academic ability (thus raising the possibility that SES effects in that sector were artifactually attenuated).

CHK's rejoinder is to state (i) that their common school conclusion holds up when the full set of measured background variables is used; (ii) that standard errors became available and, they implied, showed statistical significance for their hypothesis; and (iii) that our selection-attenuation argument "fails the one empirical test to which it has been put."

Only a limited amount of information about these issues is published. CHK have never published their long (17-variable) regression for the Catholic sector. (DR, FR, and BB all report such regressions for the private--that is, Catholic + other-private--sector). They have given (HER:534) selected public-Catholic comparisons for three variables (parental education, race and ethnicity) obtained from 19-variable regressions (their standard 17 augmented by two curricular track variables). Mueser has provided us with the 17-variable fully interactive regressions for the public + Catholic sectors. To evaluate CHK's rejoinder, we will refer to all these sources.

In HER (534) we see that of the 18 contrasts CHK offer (3 tests x 2 grades x 3 SES variables), 17 show a smaller sector difference in the long regression than they did in the short regression (DR: 178; BB: 145). The average reduction is about one-third. Only 4 of those 18 sector differences are statistically significant by the 3-sigma rule. This examination supports our arguments about (i) "distortion" and (ii) "mere chance."

Careful readers will have noticed the slippage between 5 (the number of variables which CHK had picked to represent SES) and 3 (the number of variables for which they present and discuss sector differences in coefficients). When we examine all 5 variables and the 30, rather than

18, coefficient differences, our arguments are again supported. Based on the unpublished long regressions for public-Catholic comparisons, we see that 25 of the 30 comparisons show smaller sector differences than are shown with the short regression (FR: A32; BB: Table A-10), and only 4 of the 30 coefficient differences are significant by the 3-sigma rule.

In fact a broader view of the "common school" hypothesis is called for now that CHK explicitly declare that all 17 variables qualify as measures of "SES background" (SE2: 172-173, n. 7). Using the unpublished material, we see that of the 102 available comparisons (3 tests x 2 grades x 17 variables), 43 show coefficient differences _contrary in sign_ to the broadened common school hypothesis. Of the 102, we estimate that perhaps 15 are significantly different from zero by the 3-sigma rule. Readers may determine for themselves that very similar results would be obtained if the published regression results for the public and private sectors are examined (FR: A28-A31; BB: Tables A-6 to A-9). We note that perhaps 3 of the 12 coefficient comparisons from the two race-ethnicity variables are significant, and 11 of the 12 are in the egalitarian direction that CHK claim for the Catholic sector. Perhaps, therefore, the sector difference in the effects of race and ethnicity are not attributable to chance. By the same token, the other 15 background variables offer scant support for the broadened common school hypothesis: Of the remaining 90 coefficient differences between the Catholic and public sectors, 42 show the wrong sign. It is these wrong signs, and not the lack of statistical significance for individual coefficients, that provides the strongest evidence against the common school hypothesis.

To illustrate our argument that selectivity-attenuation, rather than egalitarianism, might have led to the smaller slope estimates in the

Catholic sector, we remarked that a similar situation should obtain within curricular tracks, where selection is also partly determined by academic ability. Coefficients on the SES variables, we predicted, would be smaller in regressions run within curricular tracks of the public sector than they were in those run across the entire public sector. (Actually we wrote "within academic track," an ambiguous term for which we apologize; our intent was "within each of the academic (that is, curricular) tracks," as the logic of our argument required.)

That was the prediction which, according to CHK (SE2: 173), failed its only empirical test. They say that when they ran test-score-on-background regressions within the public academic track, all 18 coefficients (3 tests x 2 grades x 3 SES variables) were larger than they were in the regressions run across the entire public sector. We cannot verify this because CHK have not published their within-track regressions. On the other hand, we do have access to Willms' results, which give a different picture. Willms uses the five variables originally picked by CHK to represent SES, two tests, and only one grade level. As shown in Table 1 below, all 18 of his income and parental education coefficients are smaller in within-track than in all-track regressions, in accordance with our prediction. But only six of the twelve race and ethnicity coefficients are in accord with our prediction.

The logic of our selectivity-attenuation principle is so compelling that it is useful to examine why the prediction we drew from it is partially disconfirmed with the race and ethnicity variables. The logic is as follows. A typical background variable, say father's education, will be positively associated with test scores via two paths: the higher the father's education, (a) the more likely the student is to be in the aca-

Table 1

Regression Coefficients from Analyses of All Tracks
and Within Tracks (Five Variable Regression)
(Public School Sophomores: Reading and Mathematics Tests)

Independent Variable	Reading				Mathematics			
	All Tracks	Acad.	Gen.	Voc.	All Tracks	Acad.	Gen.	Voc.
Income	.15	.03	.12	.10	.45	.28	.31	.31
Father's Ed.	.28	.19	.20	.14	.54	.35	.34	.28
Mother's Ed.	.21	.15	.10	.09	.42	.23	.30	.22
Black	-2.40	-2.86*	-1.96	-2.44*	-5.18	-6.41*	-4.21	-4.87
Hispanic	2.03	2.40*	1.30	-2.14*	-3.91	-4.93*	-2.52	-3.61

*"Within-track" coefficient is larger (in absolute value) than "all-track" coefficient.

Source For "all track" coefficients, see Willms (1982a: Appendix 1 and 2). The "within-track" coefficients are from unpublished results which were sent to us by Willms.

demic track, and (b) within any curricular track, the more likely the student is to achieve a higher academic performance. Both paths are manifest in the overall regression, whereas only the second is manifest in the within-track regressions. Under fairly general conditions, the overall regression should show larger slopes when, as here, both paths are positive. We knew that a more advantaged background was positively associated with Catholic sector enrollment. We presumed that it was also associated with academic track enrollment, a presumption which, we now learn, is correct for income and parental education, but not for race and ethnicity. In the public sector, academic track correlates about .20 with each of the income and parental education variables, but only about .05 with the race-and-ethnicity variables. Looking at all three tracks, we calculate that the mean of the absolute values of the correlations between the race-and-ethnicity variables and the three tracks is .04; the corresponding mean of the absolute values of the three SES-track correlations is .14 (FR: A48, A60). (Being simple correlations, these are only suggestive of the relevant partial correlations.) Thus the empirical presumption which led us from our selectivity-attenuation principle to our within-track prediction was too weak as far as race and ethnicity are concerned. We conjecture that most of the remaining twelve background variables do influence track status, and we predict that their coefficients will be attenuated within tracks.

In summary, our examination of CHK's rejoinder to our criticism of the common school hypothesis leaves us more skeptical than before. We find that CHK did exaggerate the sector differences in the narrow set of coefficients they had picked to represent SES, and that when the representation is broadened, their hypothesis is disconfirmed, with the possible exceptions of race and Hispanic ethnicity. We find that they implied statistical significance for their results when the use of (approximately) correct standard errors shows no significance. And we find that their empirical test does not serve to reject our hypothesis of selectivity-attenuation.

The only part of their common school discussion that appears valid is that blacks and Hispanics enrolled in the Catholic sector have higher test scores, relative to whites, than do those enrolled in the public sector. Are the smaller (negative) coefficients for the race-ethnicity variables in the Catholic sector attributable to selection-attenuation? An equivalent question is whether academic selection in the Catholic sector, relative to the public sector, is stronger by blacks and Hispanics than by whites. The issue deserves further investigation.

7. Senior Increments and the Sophomore-to-Senior Change

To obtain an alternate measure of sector effects, CHK used the raw difference between senior and sophomore average scores for each sector, after adjusting senior scores for dropouts. The adjustment, which drew on CHK's own estimates of dropout rates, employed an arbitrary assumption to assign hypothetical test scores to the missing seniors, an assumption said by CHK to be tilted in favor of the public schools. We pointed out that their dropout-rate estimates were too high, being out of line with official estimates from the National Center for Education Statistics (NCES). Explicating (or so we thought) the mechanics of their assignment device, we saw no pro-public tilt. CHK (SE2: 174) respond that both our arithmetic in the rate comparison and our description of the assignment device were wrong.

There was nothing wrong with our arithmetic. Expressed as a percentage of sophomores, the NCES national dropout rate is 17%, which is well below CHK's 23%. Expressed as a percentage of seniors, the NCES national dropout rate is 20%, which, as we said, is well below CHK's 29%. Rather than check our report of this statistic, which is crucial to their calculation of senior increments, CHK assumed, quite gratuitously, that our 20% was a percentage of sophomores. CHK have since conceded that their dropout rates were too high (FR: 203-204), but they have not changed their estimated increments accordingly.

Our explication of the assignment device was dead wrong, as foonote 8 in SE2 (174) makes clear. We apologize for that blunder. However, the pro-public tilt, which we could not see, is no longer visible to CHK either. The passage we quoted from the Draft Report was:

> This assumption probably errs on the side of being favorable to those schools with high proportions of dropouts (in this case, the public schools) because dropouts are probably concentrated more toward the bottom of the distribution than is assumed (DR: 193).

The corresponding passage in their book is:

> The assumption about where the dropouts came from in the test score distribution may be unfavorable to those schools with high proportions of dropouts (in this case, the public schools) because dropouts may be less fully drawn from the lower part of the test score distribution than assumed (BB: 150).

We made the rather obvious point that CHK's treatment of the important dropout problem was wrong in principle in that it ruled out a direct role for background as a determinant of dropping out. CHK (SE2: 174, n. 9) announce that our argument from principle is wrong because they "do not accept it," a magisterial pronouncement whose force is somewhat diluted by the fact that they had, as we noted (GC: 115), made the same argument themselves (DR: 173).

Because longitudinal information is lacking in the HSB sample, we offered two suggestions for handling the dropout-induced selectivity in senior test scores: school-level regressions could be run with measured background and sophomore test scores as the explanatory variables (also suggested by Olneck, 1981); individual-level regressions could be adjusted by the econometric approach. CHK like our suggestion for a school-level analysis (SE2: 173, 176). Indeed, Kilgore (1982) has adopted it. Unfortunately, her new results are not directly comparable with their original results because, whereas CHK had used 17 background variables, Kilgore uses fewer.

CHK are mystified (SE2: 174) by our reference to an econometric approach when the probit step is unavailable. Evidently, they are unaware of the one-step version of the econometric adjustment procedure, designed for just such "truncated" data situations (see Barnow, Cain, and Goldberger, 1980, and references therein).

CHK converted their dropout-adjusted senior increments into growth rates by a formula (percentage decrease in the mean number of incorrect answers in each sector) that was intended to correct for "ceiling effects" in the tests. Recognizing that the formula had the empirical consequence of discounting gains in the public sector, we observed that the choice of growth rate formulas was nontrivial, and proposed that a "more conventional" formula be used (percentage increase in the mean number of correct answers). We did the relevant calculations and tabulated our alternative dropout-adjusted growth rates along with those we calculated from background-controlled increments.

CHK strenuously object to our label "conventional," construct an extreme example in which our formula produces an absurd result, and dismiss the "whole" of our table as "meaningless" (SE2: 174-175).

It was indeed presumptuous of us to label conventional what was merely familiar to us as economists, and we accept CHK's instruction that their formula is a familiar one in psychology. The substantive issue here is the extent to which "ceiling effects" prevailed in the CHK data set. Heyns and Hilton provide guidance on the issue:

> Ceiling effects are typically assessed by computing the difference between the maximum score possible and the observed mean, divided by the standard deviation. If this score is less than 1, ... ceiling effects are problematic (1982:96).

If this is conventional usage in psychology, then there is no "ceiling effect" problem in the short tests: for all twelve public and Catholic sector means, the ratio is greater than 1.5, not less than 1 (SE1: 69).

In FR (195-198) CHK provide data on the full tests, which are evidently immune from ceiling effects (HH: 96). CHK could have recalculated sector comparisons in growth rates using the full-test scores. They did not, but claim that "inferences would not be changed if the full tests had been used"(FR: 198). The claim is untrue: From the numbers, we can deduce that for two of the three long tests, background-controlled estimates of the "extra senior increment for the Catholic sector" were negative. We do not conclude that we have demonstrated a true negative effect for the Catholic sector; see our previously expressed caveats about estimating sophomore-to-senior gains (GC: 114, 116). We do conclude that the following claim by CHK is invalid: "...the overall evidence from calculations of ranges of learning rates strongly confirms the inference of somewhat greater achievement in the private sector for vocabulary and mathematics..." (BB: 151).

As to our "meaningless" table of alternative growth rates, we note that CHK (FR: 205; BB: 150) have since adopted the style, introduced by us in that table, of displaying background-controlled growth rates along with their favored dropout-adjusted ones. However, they do not include the alternative growth rate based on the gain in the mean number of correct answers, nor do they use the long tests to address the sector differences in sophomore-to-senior changes. Finally, we are bemused by the contortions they go through in a footnote to convert a calculated growth rate of .09 for reading in the Catholic sector into a .10 (FR: 205; BB: 151, n.3).

7. School Policies

We proceed to the mêlange of schoolwork, attitudinal, and behavioral measures which served CHK as "school policies." From a complex series of calculations, they inferred, in the words of the senior author, that those variables

> are in fact those which make a difference in achievement in all American high schools no matter what sector they are in. Schools which impose strong academic demands, schools which make demands on attendance and behavior...are, according to these results, schools which bring about higher achievement (Coleman, 1981: 24-25).

CHK also claimed that their policy analysis

> provides a strong reinforcement to the inference that the average Catholic or other-private school does bring about higher achievement for comparable students than does the average public high school (SE1: 73).

Our critique was directed at the mechanics of the calculations and, more fundamentally, at the attributions of causality.

The calculations relied on regressions of test score on background and "school policies." We observed that DR contained no evidence on the statistical significance of the "school policies" in the public sector (over which such regressions had been run), and no evidence at all on the role of those variables in the private sector (over which, evidently, such regressions had not been run). CHK combined the regression coefficients with the sector differences in levels of the "school policy" variables to provide an accounting of the contributions of those variables to sector differences in test scores. We observed that CHK's interpretation of the accounting was incompatible with their numbers, which showed the sum of the contributions to be out of line with the total being accounted for.

To these rather elementary observations, CHK make no response in SE2. Their other publications are more informative. Our reading of the public-sector regressions (BB: Tables A6, A8, A18, A19) is that the increment to R^2 associated with the inclusion of the set of "school policy" variables is sufficiently large as to indicate significance by a conventional F-test. (Perhaps the most striking single item here is the strong net positive association between "having taken an advanced mathematics course"--one of CHK's "school policies"--and the math test score for seniors.) While CHK still fail to tabulate regressions of test score on background and "policies" for the private sector, it is evident that they have run them. For BB (172) gives an alternative accounting, derived from private-sector regressions, of the contributions of the policy variables to sector differences in test scores. To the naked eye, the new figures are wildly different from those derived from the public-sector regressions (DR: 213; BB: 171). For CHK, on the other hand, the picture is neater:

> Without going into details, the results are generally consistent with those of the public analysis.... However, ... achievement in private schools is considerably more sensitive to the school's functioning (BB: 172).

Actually, in the new accounting, "student behavior" has a perverse effect on senior math scores (on balance, the more fighting, threatening, absenteeism, and class cutting, the higher the test score). This detail escapes CHK, who summarize their results by saying: "student behavior in a school has strong and consistent effects on student achievement" (BB: 178).

In the original accounting, the "disciplinary climate" component of the "school policies" was estimated to have a perverse negative effect. CHK (DR: 216-219) rationalized that away, announcing that the effect is really positive because the "disciplinary climate" coefficient became positive when the "student behavior" variables were discarded. We remarked that whatever the merits of that ex post argument might be, it would be improper to simultaneously credit both "disciplinary climate" and "student behavior" with positive effects on achievement. In SE2, CHK make no response to our remark and persist in crediting both sets of variables with positive effects. We await their rationalization of the new perversity found in the private sector.

We called attention to the startling magnitude of some effects of the "policy variables" reported by CHK: <u>four days</u> of additional attendance per semester, we calculated, is as efficacious as <u>two years</u> of high school in the production of mathematics test scores. CHK dismiss our calculation: it is wrong, they say, because it is based on one of their regressions which is "misspecified" by virtue of containing only the short, 5-variable, list of background variables. They do not take the opportunity to redo our calculation for a better-specified regression. We do so here. From their regressions (FR: A40), which include 29 explanatory variables (17 background + 12 "policy"), we now find that the four days of additional attendance per semester are as efficacious as one year of high school. The magnitude is smaller, but still startling. As we stated (GC: 118), magnitudes such as these are implausible as measures of the causal effect of a difference in school policy, but quite plausible as measures of the effect of a difference in personal attributes.

This brings us to our final theme, the causal interpretation of CHK's "school policies." Examining the student questionnaire items from which CHK had assembled their "policy variables," we argued that those schoolwork, behavioral, and attitudinal measures were mainly reflections of otherwise uncontrolled predispositions of the students, rather than being policy instruments controlled by the school administrators. The very same issue, we noted, had been raised by CHK themselves:

> One might argue that...the kind of students who tend to be lower achievers are those who are absent or cut classes, and it is not the absences themselves that reduce achievement (DR: 200).

Upon raising that issue, we noted, CHK had immediately dismissed it with the remark that policy regression coefficients were similar in the several sectors. This remark is repeated in BB (204-205). Its relevance escaped us entirely, and still does.

In SE2, CHK make no attempt to explain their remark. Instead they dismiss our argument on two grounds (SE2: 175-176). First, they say, our argument is an a priori one. Second, they say, our argument is falsified by reference to twelve high-performance public schools in HSB: those twelve schools have high levels of background and test scores, but low levels of homework, attendance, and discipline as compared with the private sector. The relevance of this conjunction of facts eludes us. Those twelve schools are larger; they offer (and their students take) more advanced courses; they offer (and their students take) more vocational courses; their students and administrators report about the same levels of disciplinary problems as are reported in the public sector generally (DR: Chapter 5). These and many other facts can also be conjoined. In what way does any of this demonstrate that CHK's "school policies" are really school policies?

In their book, CHK raise the issue again with a new slant:

> There is also the possibility that variations in absence within a school are symptoms of individual factors that affect achievement, and it is these factors, rather than absence itself, which are responsible for the achievement differences (BB: 160, n.17).

This time they say that the possibility can be dismissed if the within-school regression coefficients are not much larger that the between-school ones. Once again their logic eludes us. It is conceivable that they have in mind their earlier notion (DR: 215-216; SE1:74) that calculating school-level averages suffices to transform student-specific attributes into school policies. But that notion is a misguided one, as we showed (GC: 120). For further discussion of how school-level means might be interpreted properly, see Crain and Ferrer (1982) and Murnane (1982).

The issue remains. We have not denied that policy instruments at the disposal of school authorities might account in part for the observed values of the "policy variables" (see GC: 119). Our point is that the numbers CHK produce as estimated effects of school policies are so grossly exaggerated as to be worthless.

In this vein, we wrote that the logic of CHK's attribution of causality to their "policy variables" would compel them to conclude that shifting students from remedial mathematics classes into advanced mathematics classes, and from vocational into academic curricula, would be an effective way to raise their scholastic achievement. We intended that as a reductio ad absurdum of the CHK position. But evidently the irony was lost on CHK (SE2: 176, n. 10). Perhaps some numerical examples will make the point more clearly. (i) Among public school sophomores, the effect of "taking an honors mathematics course" (one of CHK's "policy variables") is to increase the mathematics test score by an amount that is equivalent to almost _five years_ of high school achievement in mathematics (calculated from FR: A40; SE1: 70-71). The regression equation producing this effect contains CHK's long list of 17 background variables along with their 12 "policy variables." Common sense tells us that the large coefficient on "taking an honors mathematics course" arises because that variable is proxying for the otherwise uncontrolled background achievement and aptitude of students who take such courses as compared to those who do not. (ii) Shifting public school students from the general track to the academic track has an effect on test scores that is equivalent to _two to five_ years of high school achievement (calculated from HER: 532, n. 8; SE1: 70-71). This effect is vastly larger than the gain (approximately one year) that CHK claim for the shift from a public school to a Catholic school. Here too only a proxy interpretation of the track variable is compatible with common sense.

CHK's statistical analysis and the implied "policy" results would, in our view, be comical except that they are so stoutly defended by the authors and so "relevant" to the current debates about educational policy. We find it sad, rather than comical, when some reviewers of CHK's work proclaim the importance of CHK's findings on school policies (Ravitch 1981, 1982; Bane, 1982). To avoid misunderstanding, we had better confess that we too believe that homework, attendance, and fair discipline are good things. It would be comforting to have that belief verified, but foolhardy to rely on CHK's study for its verification.

8. Conclusion

Our reconsideration of Coleman, Hoffer, and Kilgore's causal analyses of cognitive outcomes has reinforced our previous conclusions (GC: 103, 121). Their statistical methods, reporting style, and mode of inference fall below the minimum standards for social-scientific research. Their conclusions about the virtues of private education and efficacy of educational policies are not warranted by their empirical evidence.

References

Bane, M.J. 1982. "Families count and schools do too," The New York Times Book Review, Sunday, Nov. 28: 11+.

Barnow, B.S., G.G. Cain and A.S. Goldberger. 1980. "Issues in the analysis of selectivity bias." Pp. 43-59 in E.W. Stromsdorfer and G. Farkas (eds.), Evaluation Studies Review Annual, vol. 5. Beverly Hills: Sage Publications.

Coleman, J.S. 1981. "Private schools, public schools and the public interest." The Public Interest, 64: 19-30.

Coleman, J.S., T. Hoffer and S. Kilgore. 1981a. Public and private schools. Draft Report submitted to the National Center for Education Statistics, Washington, D.C.

______. 1981b. "Questions and answers: our response. Harvard Educational Review, 51: 526-545.

______. 1981c. Public and private schools. Contractor Report submitted to the National Center for Education Statistics, Washington, D.C.

______. 1982a. "Cognitive outcomes in public and private schools." Sociology of Education, 55: 65-76.

______. 1982b. "Achievement and segregation in secondary schools: a further look at public and private school differences." Sociology of Education, 55: 162-182.

______. 1982c. High School Achievement: Public, Catholic, and Private Schools Compared. New York: Basic Books.

Crain, R.L. and R.L. Ferrer. 1982. "Achievement prediction with school level equations: a non-technical example using the public and private schools data." Center for Social Organization of Schools, The Johns Hopkins University, Report No. 323, March.

Fetters, B., J. Owings, S. Peng and R. Takai. 1981. "Review of NORC report, 'Public and Private Schools'". Washington, D.C.: National Center for Education Statistics.

Goldberger, A.S. and G.G. Cain. 1982. "The causal analysis of cognitive outcomes in the Coleman, Hoffer and Kilgore report." Sociology of Education, 55: 103-122.

Heyns, B. and T.L. Hilton. 1982. "The cognitive tests for High School and Beyond: an assessment." Sociology of Education, 55: 89-102.

Kilgore, S. 1982. School policy and cognitive growth in public and Catholic secondary schools. Ph.D. dissertation, Department of Sociology, University of Chicago.

McPartland, J.M. and E.L. McDill. 1982. "Control and differentiation in the structure of American education." Sociology of Education, 55: 77-88.

Murnane, R. 1982. "The uncertain consequences of tuition tax credits: an analysis of student achievement and economic incentives." Program on Non-Profit Organizations. Institute for Social and Policy Studies, Yale University, Working Paper 46.

Noell, J. 1982. "Public and Catholic schools: a reanalysis of 'Public and Private Schools'." Sociology of Education, 55: 123-132.

Olneck, M. 1981. "Are private schools better than public schools?: A critique of the Coleman Report." Institute for Research on Poverty, University of Wisconsin, Focus, 5: 1-4.

Ravitch, D. 1981. "The meaning of the new Coleman report." Phi Delta Kappan, June: 718-720.

_____. 1982. "What makes a good school." Society, Jan./Feb., 19: 10-11.

Willms, J.D. 1982a. "Achievement outcomes in public and private schools: A closer look at the 'High School and Beyond' data." Stanford Institute for Research on Educational Finance and Governance.

_____. 1982b. "Public and private schools: The evaluation of choice in education." Unpublished paper, Faculty of Education, The University of British Columbia.