NBER WORKING PAPER SERIES

USING MAIMONIDES' RULE TO ESTIMATE THE EFFECT OF CLASS SIZE ON SCHOLASTIC ACHIEVEMENT

Joshua D. Angrist
Victor Lavy

Working Paper 5888

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1997

Special thanks to Nora Cohen and the staff at the Chief Scientist's office in the Israeli Ministry of Education for assistance with the data, and to hard-working research assistants Phil Ellis and Jon Guryan. We also benefitted from the helpful comments of Mike Boozer, Aaron Yelowitz, seminar participants at Harvard and seminar participants in the June 1996 "Econometrics in Tel Aviv" conference. The authors bear sole responsibility for the content of this paper. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Using Maimonides' Rule to Estimate the Effect
of Class Size on Scholastic Achievement
Joshua D. Angrist and Victor Lavy
NBER Working Paper No. 5888
January 1997
JEL Nos. I21, J24
Labor Studies

ABSTRACT

The effect of class size on student achievement has long been of concern to educators, parents, and scholars. In Israeli public schools today, class size is partly determined using a rule proposed by Maimonides in the 12th century. This rule induces a nonlinear and non-monotonic relationship between enrollment size and class size. We use this relationship to construct instrumental variables estimates of the effect of class size on the test scores of Israeli 4th and 5th graders in 1991 and 3rd graders in 1992. Because the up-and-down pattern in class size induced by Maimonides' rule matches a similar pattern in test scores, the rule provides a credible source of exogenous variation for investigation of the causal effect of class size on student achievement. Our use of Maimonides' rule can be viewed as an application of Campbell's (1969) regression-discontinuity design to the class size question. The results of this application show that reductions in class size induce a significant and substantial increase in reading and math scores for 5th graders and a smaller increase in reading scores for 4th graders. In contrast, there is little evidence of any association between class size and the test scores of 3rd graders, although this finding may result from problems with the 1992 wave of the testing program. The estimates also suggest that the gains from small classes are largest for students from disadvantaged backgrounds. Besides being of methodological interest and providing new evidence on the class size question, these findings are of immediate policy interest in Israel where legislation to reduce the maximum class size is pending.

Joshua D. Angrist
Department of Economics
Massachusetts Institute of Technology
E52-353
50 Memorial Drive
Cambridge, MA 02139-4307
and NBER
angrist@mit.edu

Victor Lavy
Department of Economics
Hebrew University
Mt. Scopus
Jerusalem 91905
ISRAEL
msvictor@olive.mscc.huji.ac.il

When asked about their views on class size in surveys, parents and teachers generally report that they prefer smaller classes. This may be because they believe that smaller classes promote student learning, or simply because smaller classes offer a more pleasant environment for the pupils and teachers who are in them (Mueller, Chase, and Walden, 1988). Social scientists and school administrators also have a long-standing interest in the class-size question. Because class size can be easier to manipulate than other school inputs, it is a variable at the heart of policy debates on school quality and the allocation of school resources in many countries (see, e.g., Robinson, 1990, for the US; OFSTED, 1995 for the UK; and Moshel-Ravid, 1995, for Israel).

The broad interest in the consequences of changing class size notwithstanding, the causal effect of class size on pupil achievement has proven very difficult to measure. Even though the level of educational inputs differs substantially both between and within schools, these differences are often associated with factors such as remedial training or students' socioeconomic background. Probably for this reason, much of the research on the relationship between class size and achievement is inconclusive. In widely cited meta-analyses of class-size research, Glass and Smith (1979) and Glass, Cahen, Smith, and Filby (1982) conclude that smaller classes raise children's test scores. Card and Krueger (1992a,b) also found that lower pupil-teacher ratios [1] in school are associated with higher adult earnings, while randomized trials in Tennessee and Ontario provide evidence for beneficial effects of randomly assigned reductions in class size (Finn and Achilles, 1990; Wright, et all, 1977). But results from the Glass, et al meta-analyses have been questioned (Slavin, 1989), and Hanushek's (1986, 1996) surveys of research on the effects of school inputs, including pupil-teacher ratios, show mixed results. Recently, Card and Krueger's results have also been challenged (Heckman, Layne-Farrar, and Todd, 1995).[2]

The academic interest and public controversy surrounding the class size question is not only a modern

---

[1]Pupil-teacher ratios and class size are highly correlated but not identical measures of school inputs. Boozer and Rouse (1995) discuss the implications of this distinction for class-size research.

[2]See Card and Krueger (1996) for an updated survey of work on the quality-earnings connection.

phenomenon; the choice of class size has been of concern to scholars and teachers for hundreds of years. An important early example is the Babylonian Talmud, completed around the beginning of the 6th Century, which discusses rules for the determination of class size and pupil-teacher ratios in bible study. The great 12th century Rabbinic scholar, Maimonides, interprets the Talmud's discussion of class size as follows: "Twenty-five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed." (Hyamson, 1937, p. 58b).[3] Although Maimonides' maximum of 40 students was partly derived by interpreting the Talmud, this rule actually leads to smaller classes and lower pupil-teacher ratios than the original Talmudic rule, which allows a maximum class size of 50.[4]

The importance of Maimonides' rule for our purposes is that it is used to determine the division of enrollment cohorts into classes in Israeli public schools today. As we show below, this rule generates a source of variation in class size that can be used to estimate the effects of class size on the scholastic achievement of Israeli pupils. To see how this variation comes about in practice, note that according to Maimonides' rule, class size increases one-for-one with enrollment until 40 pupils are enrolled, but when 41 students are enrolled, there will be a sharp drop in class size, to an average of 20.5 pupils. Similarly, when 80 pupils are enrolled, the average class size will again be 40, but when 81 pupils are enrolled the average class size drops to 27.

Of course, Maimonides' rule is not the only source of variation in Israeli class sizes, and average class size is generally smaller than what would be predicted by a strict application of this rule. But Israeli classes

---

[3]This is from Chapter II of "Laws Concerning the Study of Torah" in Book I of Maimonides *Mishneh Torah*. The same chapter discusses compulsory school attendance (at public expense from the age of 6 or 7 for boys), the penalty for nonenforcement of compulsory attendance laws (excommunication of the entire town), hours of instruction (long), holidays (few), use of corporal punishment (limited), qualifications for teaching positions (strict), competition between schools for students (permitted, desirable), and busing school students between towns to schools of higher quality (permitted only if the towns are not separated by a river).

[4]The Talmudic portion that Maimonides relied on is: "The number of pupils assigned to each teacher is twenty-five. If there are fifty, we appoint two teachers. If there are forty, we appoint an assistant, at the expense of the town." (Quote from chapter II, portion chaf-aleph of the Baba Bathra, english translation on page 214 of Epstein, 1976).

are large by US standards and the ceiling of 40 students per class is a real constraint faced by many school principals. Both the median and the median Israeli class sizes for Jewish students in our data are around 31 pupils, with 25% of classes having more than 35 pupils.[5] A regression of actual class size at mid-year on predicted class-size using beginning-of-the-year enrollment data and Maimonides' rule explains about half the variation in class size in each grade (in a population of about 2,000 classes per grade).

In this paper, we use the class-size function induced by Maimonides' rule to construct instrumental variables estimates of class-size effects in regressions of test scores on class size and school characteristics. Although the class-size function and the instruments derived from it are themselves a function of the size of enrollment cohorts, these functions are non-linear and non-monotonic. We can therefore control for a wide range of secular cohort-size effects when using the class-size function as an instrument. It turns out that in practice, there is not always a need to do this. Moreover, in cases where there are secular enrollment-size effects, we can still match the up-and-down pattern in class size induced by Maimonides' rule to a similar pattern in test scores. Because it seems unlikely that secular enrollment-size effects would generate such a pattern, the class-size function derived from Maimonides' rule provides an unusually credible source of exogenous variation that can be used to identify the causal effect of class size on student achievement. Use of Maimonides' rule as an instrument while controlling for secular enrollment effects can be viewed as an application of Campbell's (1969) regression-discontinuity research design to the class size question.

The paper is organized as follows. Following a description of Israeli test score data in Section I, Section II presents a simple graphical analysis. Section III describes our statistical model and framework for inference. Section IV reports the main estimation results, and Section V interprets some of the findings. Section VI concludes. The results suggest that reductions in class size induce a significant and substantial

[5]These figures can be contrasted with an average class size of roughly 25 for US 8th graders in 1988 (Akerhielm, 1995; US Department of Education, 1996), an average class size of roughly 20 for Tennessee first graders in 1985, an average class size of roughly 27 for Ontario 4th graders in 1974 (Wright, et al, 1977), and a median class size of 28 in British primary schools in 1994-95 (OFSTED, 1995).

increase in math and reading achievement for 5th graders, and a modest increase in reading achievement for 4th graders. On the other hand, there is little evidence of an association between class size and achievement of any kind for 3rd graders, although this may be because of the manner in which the 1992 wave of the testing program was implemented. In addition to being of methodological interest and contributing to the international class-size debate, the findings presented here are of immediate policy relevance because the Israeli Parliament is currently considering legislation that would reduce maximum class size from 40 pupils to 30 pupils.

I. Data and descriptive statistics

The test score data used in this study come from a short-lived national testing program in Israeli elementary schools. In June of 1991, near the end of the school year, all 4th and 5th graders were given achievement tests designed to measure mathematics and (Hebrew) reading skills. The tests are described and the results summarized in a pamphlet from the National Center for Education Feedback (1991). The scores used here consist of a composite constructed from some of the basic and all of the more advanced questions in the test, divided by the number of questions in the composite score, so that the score is scaled from 1-100. This composite is commonly used in Israeli discussions of the test results.[6] As part of the same program, similar tests were given to 3rd graders in June 1992. The June 1992 tests are described in another pamphlet [National Center for Education Feedback (1993)].[7] The achievement tests generated considerable public controversy because of lower scores than anticipated, especially in 1991, and because of large regional difference in outcomes. After 1992, the national testing program was abandoned.

---

[6]The 4th grade tests included 45 math questions and 57 reading questions. The 5th grade tests included 48 math questions and 60 reading questions. Among these, 15 questions are considered basic, and the remainder more advanced.

[7]The 1992 exams included 40 math questions, of which 20 were considered basic. The math composite score includes 10 of the basic questions plus 20 of the more advanced questions. The reading exams included 44 questions, of which 20 were considered basic. The reading composite includes 10 of the basic reading questions plus all of the more advanced questions.

Our analysis begins by linking average math and reading scores for each class with data on school characteristics and class size from other sources. The details of this link are described in the data appendix. Briefly, the linked data sets contain information on the population of schools covered by the Central Bureau of Statistics (1991, 1993) Censuses of Schools. These are annual reports on all educational institutions at the beginning of the school year (in September), based on reports from school authorities to the Israeli Ministry of Education and supplemented by Central Bureau of Statistics data collection as needed. Our measure of beginning-of-the-year enrollment is taken directly from the computerized files underlying these reports, and the classes in the schools covered by the reports define our study population. The data on class size are from an administrative source, and were collected between March and June of the school year that began in the previous September. We use the Central Bureau of Statistics files as a reference population because the data from other sources were not collected at exactly the same time and because the reference population tells us how complete the testing program was. In fact, we have test scores for over 96 percent of the classes in the reference population, although not every pupil in every class was tested because of absences on days the tests were given.

The unit of observation in the linked data sets and for our statistical analysis is the class.[8] The linked data sets include information on average test scores in each class, the Spring class size, beginning of the year enrollment in the school for each grade, a town identifier, and a school-level index of students' socioeconomic status (SES).[9] Also included are variables identifying the ethnic character of the school (Jewish, Arab, other minority group) and the administrative affiliation of the school (public-secular, public-religious, private-religious). Schools in Israel are segregated along religious and ethnic lines, and, even within the Jewish public

[8]Micro data on students are available for 3rd graders in 1992. For comparability with the 1991 data, we aggregated the 1992 micro data up to the class level.

[9]The SES index is discussed by Algrabi (1975), and used by the Ministry of Education to allocate supplementary hours of instruction and other school resources. It is a function of pupils' fathers' education and continent of birth, and family size. The index is recorded as the fraction of students in the school who come from what is defined (using index characteristics) to be a disadvantaged background.

school system, there are separate administrative divisions and separate curricula for secular and religious schools.

This study is limited to Jewish pupils in the public school system, including both secular and religious schools. These groups account for the vast majority of school children in Israel. We exclude Arab pupils because they were not given reading tests in 1991 and because no SES index was computed or published for Arab schools until 1994. The SES index is a key control variable in our analysis because it is correlated with both enrollment size and test scores. Also excluded are students in independent religious schools because these schools are associated with orthodox Jewish groups, and have a curriculum that focuses on religious education to a much greater degree than does the public school curriculum.

Table 1 reports descriptive statistics for the population of over 2,000 classes in Jewish public schools in each grade (about 62,000 pupils). The distribution of test scores, also shown in the table, refers to the distribution of *average* scores in each class. The last panel of the table reports statistics weighted by class size, which can be interpreted as averages for pupils as opposed to averages for classes. The weighted and unweighted statistics are generally similar, except for class and enrollment sizes, for which the pupil-weighted averages are naturally larger than the class averages. Note that we do not have a measure of total variation for the micro-level test scores in 1991.

One important feature of the descriptive statistics is that mean scores are much higher and the standard deviations of scores are much lower for 3rd graders than for the 4th and 5th graders. We believe this is because of a systematic preparation effort for the 1992 test on the part of teachers and school officials, in light of the political fallout resulting from what were felt to be were disappointing test results in 1991.

II. Graphical analysis

The class size function derived from Maimonides' rule can be stated formally as follows. Let e_s denote beginning-of-the-year enrollment in school s in a given grade and let f_{sc} denote the class size assigned to class c in school s, for that grade. Assuming cohorts are divided into classes of equal size, we have

$$f_{sc} = e_s/(\text{int}((e_s-1)/40)+1), \tag{1}$$

where, for any positive number, a, int(a) is the largest integer less than or equal to a. Equation (1) captures the fact that Maimonides' rule allows enrollment cohorts of 1-40 to be grouped in a single class, but enrollment cohorts of 41-80 are split into two classes of average size 20.5-40, enrollment cohorts of 81-120 are split into three classes of average size 27-40, and so on. Of course, any single class necessarily includes an integer number of students, so that an enrollment cohort of 121 would likely be divided into 3 classes of 30 pupils and one class of 31 pupils.

The actual relationship between class size and enrollment size involves many factors, but in Israel it clearly has a lot to do with the class-size function, f_{sc}. This can be seen in figures 1 and 2, which plot the average class size by enrollment size for 4th and 5th grade pupils, along with the class size function. The horizontal lines in the figures mark the class sizes where the class-size function has corners. The figures show that at enrollment values where class size is increasing, it increases approximately linearly with enrollment size. But average class size also drops sharply at the corners, as predicted by the class size function.[10]

Other noteworthy features of the figures include the fact that average class size never reaches 40 when enrollment is less than 120, even though the class size function predicts a class size of 40 when enrollment is either 40 or 80. This is because schools can sometimes afford to add extra classes before reaching the maximum class size.[11] Note also that average class sizes do not drop as much at the corners of the class size function as f_{sc} predicts. This is because beginning-of-the-year enrollment, e_s, is not necessarily the enrollment at the time the class size data were collected (if enrollment has fallen, then an initially large cohort will not

[10]The averages in these and the following graphs are unweighted. Averages weighted by class size, which show expected values per-pupil instead of per class, show very similar patterns.

[11]Schools' ability to support more classes is partly a function of the SES index. The Israeli Ministry of Education funnels special discretionary resources, including hours of instruction and teachers, to those schools where the SES index is low (Lavy, 1995).

necessarily have been split) and because a few classes are reported to include more than 40 pupils.[12] Finally, it is clear from the figures that when correlations are calculated over the entire support of the enrollment data, average class size is positively correlated with enrollment size.

In addition to exhibiting a strong association with average class size, the class-size function is also correlated with average test scores. This is easiest to see in the relationship between test scores and enrollment size. First, test scores are generally higher in schools with larger enrollments and hence larger predicted class sizes. Second, and most importantly, average scores by enrollment size exhibit an up-and-down pattern that is, in large part, the mirror image of the class size function. These patterns are apparent in figures 3a, 3b, and 3c, which plot average test scores and average values of f_{sc} by enrollment size, in enrollment intervals of 10, for the reading and math scores of 5th graders and for the reading scores of 4th graders.[13]

The overall positive correlation between scores and enrollment is partly attributable to that fact that larger schools in Israel are more likely to be located in relatively prosperous big cities, while smaller schools are more likely to be located in relatively poor "development towns" outside of major urban centers. In fact, enrollment size and the SES index measuring the proportion of students who come from a disadvantaged background are highly negatively correlated. It is also possible that better schools (as measured by pupil test scores) have high enrollments because they attract more students.

After controlling for the trend association between test scores and enrollment size, and between test scores and pupil's SES, there is a negative association between f_{sc} and scores. This can be seen in figures 4a, 4b, and 4c, which plot residuals from regressions of average scores and the average of f_{sc} on average enrollment

---

[12]The empirical analysis is restricted to schools with at least 5 pupils reported enrolled in the relevant grade and to classes with less than 45 pupils. There are only a handful of classes reported to have more than 45 pupils. These appear to be coding errors.

[13]Intervals of 10 were used to construct the figures instead of the single-value intervals in figures 1-2 because the test score data have more idiosyncratic variation than the class size data. The enrollment axes in the figures record interval midpoints. Averages were computed for schools with enrollments between 9 and 190. This accounts for over 98 percent of classes. The last interval (165 on the x-axis) includes enrollments from 160-190.

in each interval and the average SES index for each interval. Again, the x-axis is enrollment size. Although the mirror-image relationship between detrended average scores and detrended f_{sc} is clearly not deterministic, it stands out in all three figures. If we regress the detrended average scores on the detrended average f_{sc}, the slopes are roughly -.22 for 5th graders and -.11 for 4th graders. Thus, the estimates for 5th graders imply that a reduction in *predicted* class size of 10 students is associated with a 2.2 point increase in average test scores, a little more than one-quarter of a standard deviation in the distribution of class averages.

III. Measurement framework

The figures suggest a clear link between the variation in class size induced by Maimonides' rule and pupil achievement, but they do not provide a framework for formal statistical inference. A statistical model for individual test scores is used to describe the causal relationships to be estimated. For the *i*th student in class *c* and school *s*, we can write

$$y_{isc} = X_s'\beta + n_{sc}\alpha + \mu_c + \eta_s + \epsilon_{isc} \tag{2}$$

where y_{isc} is pupil I's score, X_s is a vector of school characteristics, and n_{sc} is the size of class c in school s. The term μ_c is an i.i.d. random class component and the term η_s is an i.i.d. random school component. The remaining error component, ϵ_{isc}, is specific to pupils. The three error components, assumed to be mutually orthogonal, are introduced to parameterize possible within-school and within-class correlation in scores. The class-size coefficient α is the parameter of primary interest.

Our interpretation of equation (2) is that it describes the potential outcomes of students under alternative assignments of n_{sc}, when X_s is held fixed. Although equation (2) is linear, linearity of the true causal response function is not necessary for estimates of α to have a valid causal interpretation. For example, if n_{sc} were randomly assigned conditional on X_s, then α would be a weighted average response along the length of the true causal response function connecting class size and pupil scores (Angrist and Imbens, 1995). But since n_{sc} is not randomly assigned, in practice it is likely to be correlated with potential outcomes (in this case, the error components in [2]). Thus, OLS estimates of (2) do not have a causal interpretation, although IV estimates still might. The causal interpretation of IV estimates turns on whether it is reasonable to assume that, conditional on X_s, the only reason for any association between instruments and test scores is the association between instruments and class size.

Equation (2) is cast at the individual level because it is pupils who are affected by class size. In practice, however, much of the literature on class size treats the class as the unit of analysis and not the pupil (see, e.g., Finn and Achilles, 1990; Wright, et al, 1977). Since class size is naturally fixed within classes, and student test scores are correlated within classes, little is likely to be lost in statistical precision from this aggregation. Moreover, in the case of the test scores for 4th and 5th graders, we have no option other than a class-level analysis because the micro-level data are unavailable. To make the analyses from different years comparable, we also aggregated the 1992 data on 3rd graders to the class level. The class-level estimating equations have the form,

$$\bar{y}_{sc} = X_s'\beta + n_{sc}\alpha + \eta_s + [\mu_c + \bar{\epsilon}_{sc}], \tag{3}$$

where overbars denote averages. The term $[\mu_c + \bar{\epsilon}_{sc}]$ is the class-level error term, while the random school component, η_s, captures correlation between class averages within schools.[14]

Because of the random-effects error structure, the variance of the class-level error term is not proportional to the reciprocal of class size even if the micro-level error terms are homoscedastic. Consequently, weighted least squares estimation using cell (class) sizes as weights is not the efficient generalized least squares (GLS) estimator for this problem. Moreover, as Deaton (1995) and Pfefferman and Smith (1985) have emphasized, unless one believes that the behavioral relationship of interest is truly linear with constant coefficients, statistical theory provides little guidance as to what the best weighting scheme would be. On the other hand, it seems reasonable to require that any inference procedure reflect the correlation

[14]In their analysis of the Tennessee Project STAR experiment, Finn and Achilles (1990) also used a model with school random effects in an analysis of class-level averages. Consistency of parameter estimates in models with this kind of group structure requires that the number of groups get large (Pakes, 1983).

across observations implied by the grouped error structure. We therefore report unweighted ordinary least squares (OLS) and instrumental variables (IV) estimates of (3), along with corrected standard errors. In practice, allowing for a heteroscedastic grouped error term had very little impact on inferences, so that the grouped errors were treated as homoscedastic. But the correlation of class averages within schools does have a noticeable impact on the estimated standard errors, leading to 10-15 percent larger standard errors than conventional formulas. The standard errors were corrected for intra-school correlation using the formulas in Moulton (1986). Corrected standard errors are reported in square brackets below the uncorrected standard errors and OLS or IV coefficient estimates.[15]

IV and regression-discontinuity designs

In an influential overview of non-experimental methods in evaluation research, Campbell (1969) considered the problem of how to identify the causal effect of a treatment that is assigned as a (possibly deterministic) function of an observed covariate. Campbell used the example of estimating the effect of National Merit scholarships on scholarship applicants' later academic achievement. National Merit scholarships are awarded largely on the basis of the scholastic achievement of applicants at the time of application. Campbell suggested that the assignment rule for National Merit scholarships be modified to include sharp discontinuities (or non-linearities) in the relationship between previous scores and awards. Examples include a discrete cutoff point for eligibility, or an abrupt change in the slope of the relationship between awards and pre-treatment scores. Since there is typically no reason for a regression relating outcomes to the pre-treatment covariate to share this pattern, any evidence of such a pattern is evidence in favor of a non-

[15]The fact that the data set used here includes the entire population of 3rd, 4th, and 5th graders raises the question of what measures of sampling variance represent. One view is that even in finite populations standard errors measure the extent to which each parameter contributes to the fit of the model, as reflected in the curvature of the estimation minimand in the neighborhood of the parameter estimates. See, e.g., Angrist (1991). Another view, called "analytical inference" by Pfefferman and Smith (1985), is that inferences are being drawn from the observed finite populations about the relationships of interest in other years, or under alternative hypothetical realizations from the same behavioral process.

zero treatment effect.

The graphs discussed in the previous section can be seen as applying Campbell's suggestion to the class-size question (see, especially, Campbell's figures 12-14).[16] We interpret the up and down pattern in the conditional expectation of test scores given enrollment as reflecting the causal effect of changes in class size that are induced by changes in enrollment. This interpretation is plausible because the class-size function is known to share this pattern, while it seems likely that any other mechanism linking enrollment and test scores will be much smoother.

Something missing from Campbell's original analysis and his later work (e.g., Cook and Campbell, 1979), is how to use a regression-discontinuity design to go from a general finding of evidence in favor of the existence of an effect to a precise estimate of the size of the effect.[17] This gap is filled in here by the discussion of equations 1-3. IV estimates of equation (3) use discontinuities or non-linearities in the relationship between enrollment and class-size (captured by f_{sc}) to identify and estimate the causal effect of class size, at the same time that any other relationship between enrollment and test scores is controlled by including smooth functions of enrollment in the vector of covariates, X_s. The usual IV formulas serve to scale the reduced-form association between f_{sc} and outcomes into the units measured by the regressor of interest.

IV. Estimation results

OLS estimates for 1991

OLS estimates with no control variables show a strong positive correlation between class size and achievement. Controlling for pupil's SES, however, this association largely disappears and, in some cases, becomes negative. These findings can be seen in Table II, which reports coefficients from regressions of the

---

[16]We thank Guido Imbens for bringing this to our attention.

[17]Goldberger (1972) discusses statistical properties of a regression-discontinuity design for the special case where treatment and the discontinuity are perfectly correlated.

math and reading scores of 4th and 5th graders on class-size, the SES index, and enrollment size. In a regression of the average reading scores of 5th graders on class size alone, the class-size effect is a precisely estimated .221, but when the SES index is added as a control variable, the estimated class-size effect falls to -.031 with a standard error of .022.

Lavy (1995) previously observed that the positive association between class size and test scores in Israel is largely accounted for by the association between larger classes and higher SES among pupils. A similar point was made earlier for the US in the Coleman (1966) report regarding the allocation of school resources in general, and has been emphasized again recently in the meta-analysis by Hedges, Laine and Greenwald (1994). We note, however, that controlling for SES in the Israeli data does not completely eliminate the positive association between class size and math scores. Also, the negative OLS estimates of effects of class size on reading scores are small and, at best, marginally significant. Another result of interest in Table II is the lack of an association between enrollment size and reading scores, once SES is included in the regression. The enrollment coefficients for math scores are also small, once SES is added to the equations.

Reduced form estimates for 1991

The reduced form relationship between predicted class size (f_{sc}) and actual class size, reported in Table III for a variety of specifications, shows that higher predicted class sizes are associated with larger classes and lower test scores. Reduced form estimates in the first panel of Table III report the results of regressions on f_{sc} with controls for SES only and with controls for both SES and enrollment size.[18] The association between f_{sc} and increasing class size and decreasing test scores is strongest for 5th graders. There is a precisely estimated negative association between 4th grade reading scores and f_{sc} as well. It is also noteworthy that the reduced form relationships between f_{sc} and reading scores are not sensitive to the inclusion of a control for enrollment size. On the other hand, there is no evidence of a relationship between math scores and predicted

[18]Additional variations on these specifications are discussed below.

class size for 4th graders.

Another way of capturing the reduced-form relationships is to use a set of dummy variables indicating values of f_{sc} as multiple instruments. The dummy-variable parameterization provides a more flexible first stage and a simple check on the causal interpretation of the link from f_{sc} to class size and test scores. To implement this approach, we divided f_{sc} into 4 groups of 8 values each. The smallest enrollment size, and hence the smallest value of f_{sc} is 8. The first dummy is therefore $f_1 = 1(f_{sc} \leq 16)$, the second dummy is $f_2 = 1(16 < f_{sc} \leq 24)$, and the 3rd dummy is $f_3 = 1(24 < f_{sc} \leq 32)$. The reference group contains values of f_{sc} greater than 32. This grouping scheme has the desirable feature that the second two groups straddle some of the corners in the class-size function, so that the dummy-variable parameterization is not equivalent to a monotone step function of enrollment size.

The dummy reduced form for class size, reported in the lower half of Table III, shows an average of 15-20 fewer pupils in the first f_{sc}-group, 7-11 fewer pupils in the second group, and 2-5 fewer pupils in the 3rd group, depending on grade and whether enrollment controls are included in the model. The reduced form estimates of equations for the test scores of 5th graders also show a clear association between f_{sc}-groups and higher test scores. For example, the estimates in column (4) show that for 5th graders, reading test scores are 3.1 points (s.e.=1.1) higher in the first group (smallest classes), 2.4 points (s.e.=.6) higher in the second group, and .73 points (s.e.=.4) higher in the 3rd group. The estimates for 5th grade math scores show a similar pattern.

These reduced form estimates in Table III can be used to produced simple Wald-type IV estimates of the effect of class size on test scores. For example, dividing the test score effects in column 4 by the class size effects in column 1 leads to an estimated class-size effect on 5th grade reading scores of 3.14/-15.2 = -.21 when the dummy for the smallest group is used as an instrument. Using the second dummy as an instrument leads to an estimate of 2.4/-7.1 = -.33, while using the 3rd dummy as an instrument leads to an estimate of .73/2.3 = -.32. The IV estimate based on the linear reduced forms in this column is -.149/.542=-.275. The fact that all of these estimates are similar (given the underlying sampling variance) provides a weak specification check, in the sense that f_{sc} is always associated with larger class sizes and lower test scores, regardless of how the reduced-form relationship is parameterized. Assuming that the alternate Wald estimates are in fact the same, and that the residual in the grouped equation is homoscedastic, the efficient linear combination of all three Wald estimates can be generated by 2SLS estimation using the 3 dummies as instruments at the same time (Angrist, 1991b).

IV Estimates for 1991

In practice, use of f_{sc} itself as an instrument appears to provide a good parameterization of the relationship between predicted class size and actual class size. This can be seen in Figures 5a and 5b, which show scatterplots of class size against f_{sc} for 5th and 4th graders, after regression-adjusting for SES and enrollment size. Most of the points appear to be clustered on a 45-degree line. On the other hand, the collection of points below and to the right of the 45-degree line tends to pull the actual first-stage relationship down, so that allowing for some nonlinearity might have a payoff in the form of lower second-stage standard errors. In any case, we begin by reporting IV estimates using f_{sc} as an instrument, saving IV results from a flexible dummy-variable parameterization for later.

The IV estimates of class size effects using f_{sc} as an instrument in equations for 5th graders, reported in Table IV, consistently show a large negative effect of larger classes on math and reading scores. This conclusion is robust to a range of changes in the model used to control for secular enrollment-size effects. The alternate controls include a quadratic function of enrollment and the log of enrollment, in addition to the model with linear enrollment controls used to produce the reduced form estimates.

The IV estimates for reading scores of 5th graders in a model without controls for enrollment size is -.16 with a standard error of .04. The results from models including various controls for enrollment size, reported in columns 3-5, range from -.26 to -.31. These estimates are reasonably precise, with adjusted standard errors ranging from about .07 to .09. Without enrollment controls, the IV estimate for 5th grade math scores is virtually zero. But in models with enrollment controls, the IV estimates for the math scores of 5th graders are similar to the estimates in the corresponding models for reading scores. For example, the estimated class-size effect in column 9 is -.23.

The IV estimates for 4th graders, reported in Table V, also show a robust negative association between class size and reading achievement, although the effects for 4th graders are smaller than the effects for 5th graders. The estimate in a model without enrollment controls is -.11 (s.e.=.040), and with a linear enrollment control, the estimate is -.13 (s.e.=.059). Other estimates are not significantly different from zero, although all are negative. Although all of the IV estimates for 4th grade math scores in models with some sort of enrollment control are negative (columns 9-12), but none of them are significantly different from zero.

Additional Results for 1991

To further probe the main findings, Table VI reports estimation results from a number of variations on the specifications reported in Tables IV and V. Columns 1 and 7 report the results of adding nearly 300 town dummies to the model with quadratic enrollment controls. These estimates are generally smaller and less precise than those without town dummies, but they are not statistically distinguishable from the earlier specifications. The fact that these results are weak is not surprising, since the inclusion of town dummies naturally eliminates any small towns with only one school, and generally reduces the useful variation in f_{sc}.

An alternative to models with a full set of town dummies is partial control using 5 dummies for region of residence, 7 dummies for major cities, a dummy for the religious affiliation of the school, and a quadratic function of SES. Results from this variation, reported in columns 2 and 8, show strong effects for 5th graders, but not for 4th graders. Columns 3-5 and 9-11 report the results of replacing the instrument f_{sc} with the three dummies for groups of eight that were used to compute the reduced form estimates in Table III. Estimates for 4th grade reading scores using three dummies as instruments tend to be somewhat larger than those using f_{sc}

itself.[19] Finally, columns 6 and 12 report the results from models where the effect of class size on test scores is interacted with SES. The instruments in this case are f_{sc} and SES*f_{sc}. The estimates imply that the benefits of small classes are larger in schools where there is a higher proportion of pupils who come from a disadvantaged background.

Results for 1992 (3rd graders)

The OLS estimates for 3rd graders, reported in columns 2 and 6 of Table VII, show essentially no relationship between class size and test scores. Turning to the reduced form and IV estimates, the reduced form effect of f_{sc} on 3rd grade class size, reported in column 1, is much the same as the effect of f_{sc} on 4th and 5th grade class size. But estimates of a regression of 3rd grade test scores on f_{sc}, SES, and enrollment size, reported in columns 3 and 7, offer little evidence of a relationship between f_{sc} and scores. Finally, while the IV estimates for 3rd graders, reported in columns 4 and 5, are all negative, they are smaller than the estimates for 4th and 5th graders. None of the estimates are precise enough to be statistically distinguishable from zero.

One possible explanation for the weak findings for 3rd graders is that the effects of class size may be cumulative. Since enrollment cohorts tend to progress through elementary school together, 5th graders who happen to be in enrollment cohorts that generate small class sizes may have been grouped into small classes since first grade. The accumulation of years of experience in small classes may be required before any benefits are detectable. This sort of cumulative effect would also explain why the effects for 4th graders are smaller than those for 5th graders.

Another likely explanation for the absence of effects on 3rd graders is the fact that testing conditions were different in 1992, with a variety of (non-educational) activities directed towards increasing test scores and reducing the variation in scores across schools. As noted at the beginning of the paper, the descriptive statistics in Table I do show higher scores and lower variance in 1992. The official report of the 1992 test

---

[19]This is also true for models with town dummies.

results (National Center for Education Feedback, 1993) highlights major differences between the 1992 and 1991 waves of the testing program. For example, regular class teachers (as well as an outside exam proctor) were present when tests were taken in 1992 but not in 1991. Preparation for the 1992 tests is described as follows (page 3): "During the past year there was an intense and purposeful remedial effort on the part of the elementary school division [in the Ministry of Education] in a large number of schools with high failure rates in 1991. Similarly, in light of last year's scores [in 1991], and because of the anticipated new tests [in 1992], there was an intensive remedial effort on the part of schools, district supervisors, counselors, and others." It is also worth noting that our 1992 estimation results show much smaller effects of SES on test scores than were observed in 1991. Since the distribution of the SES index over towns and cities was essentially unchanged, this finding is consistent with the hypothesis that test preparation compressed the distribution of test scores.

V. Interpretations and comparisons

The literature discussing estimates of the relationship between class size and scholastic achievement often reports a summary statistic known as "effect size". This is the test score change that would result from a given change in class size, divided by the standard deviation of the empirical test-score distribution. Finn and Achilles (1990) discuss two variations on this measure, one using the standard deviation of test scores among pupils and one using the standard deviation of class means. Since the overall variance is naturally larger than the between-class variance, measures based on the first standard deviation are always smaller than measures based on the second. Here, we are limited to reporting results based on the second measure because we do not have the micro test score data for 4th and 5th graders. Note, however, that since class size is a class-level intervention, it seems reasonable to measure impacts relative to the distribution of average class outcomes.[20]

[20]In a world of approximately homoscedastic residual variation, changes in class size would have little effect on the within-class distribution of scores.

The Tennessee STAR experiment described by Finn and Achilles (1990, Table 5) yielded effect sizes of about .13σ-.25σ among pupils, and about .32σ to .66σ in the distribution of class means. We compare our results to the Tennessee experiment by calculating the effect size for a reduction in class size at the same rate as was done in the experiment (about 35 percent). A reduction of 35 percent from the mean of 30 pupils in Israeli 5th grade classes would mean 10.5 fewer pupils per class. Multiplying this times the IV estimate from column 3 in the table for 5th graders, (an estimate of -.275 in a model with enrollment controls) we obtain an effect size of about .36σ (= 2.9 points) in the distribution of class means. The effect size probably ranges from about 0.10σ to 0.20σ among pupils.[21] Thus, our estimates of effect size for 5th graders are in the range of those found in the Tennessee experiment.

The effect sizes based on our estimates for 4th grade reading scores are only about half as large as those for 5th graders, equal to roughly .17σ, or about 1.4 score points. As noted by Mosteller (1995), however, even apparently small effect sizes can translate into large movements through the score distribution. For example, the gap between the lower and upper quartiles of the distribution of class mean reading scores is just 10 points, and the gap between the median and the 3rd quartiles is only 4-5 points.

We can also compare the results reported here to the IV estimates reported by Akerhielm (1995), Boozer and Rouse (1995), and Hoxby (1996). Using school-level data on total 8th grade enrollment and the average 8th grade class size as instruments, Akerhielm (page 235) finds statistically significant effects on science and history achievement on the order of .15σ (in the pupil score distribution) for a 10-pupil reduction in class size.[22] Using the same National Education Longitudinal Study (NELS) data set, Boozer and Rouse (1995) report IV estimates on the order of .29σ in equations that control for base-year test scores. Their

---

[21]This inference is based on the 3rd grade micro data, where the standard deviation among students is about two and half times larger than that among classes.

[22]A problem with the use of average class size as an instrument is that average class size is equal to total enrollment divided by the number of classes. The number of classes may be endogenous (i.e., correlated with students ability) for the same reason that class size is potentially endogenous.

instruments are indicators of state maximum class sizes.

The results from the Akerhielm and Boozer and Rouse studies are similar to those reported here for 4th and 5th graders. On the other hand, using district-level population variation to construct instruments for class size in a panel data set with information on Connecticut school districts, Hoxby (1996) finds no evidence of a relationship between class size and test scores. This may be because Connecticut classes are already very small, even by US standards (the average class size in Hoxby's data is 19). Also, our findings suggest that population size alone (or enrollment size, as in the Akerhielm study) is not necessarily a good instrument. In the Israeli data at least, there is a clear "trend" association between enrollment size and test scores that is unlikely to be caused by changes in class size. An advantage of the Israeli data is that Maimonides' rule can be used as an instrument for class size while controlling for such trends.[23]

VI. Conclusions

This paper presents a variety of OLS and IV estimates of the effect of class size on the reading and math scores of elementary school children in Israel. The raw positive correlation between achievement and class size is clearly an artifact of the association between larger classes and higher SES among pupils. IV estimates constructed by using functions of Maimonides' rule as instruments for class size consistently show a negative association between larger classes and student achievement. These effects are largest for the math and reading scores of 5th graders, with smaller effects for the reading scores of 4th graders. The fact that the estimates are largest for the oldest students may reflect cumulative effects, as well as problems with the testing program for 3rd graders.

Even though the largest effects we find are smaller than those reported in the Tennessee STAR experiment, they may nevertheless represent important gains relative to the distribution of Israeli math and

---

[23]We note that Hoxby (1996) controls for secular enrollment effects that are not time-varying using panel data and fixed effects.

reading scores. The Israeli Parliament recently began debating a bill that would lower the maximum legal class size to 30. Using the cohort size distributions in our data, we estimate that the new law will reduce average elementary-school class sizes from 31 to about 25 and reduce the upper quartile from 35 to 27. These changes will clearly be expensive to implement, requiring something like 600 additional classes per grade. But the findings reported here imply that the resulting reductions in class size could lead to a change equivalent to moving two or more deciles in the 1991 score distributions.

It is also worth considering whether results for Israel are likely to be relevant for the US or other developed countries. Israel has a lower standard of living and spends less on education per pupil than the US and some OECD countries (Klinov, 1992; OECD, 1993). As noted in the introduction, Israel also has larger class sizes than the US, UK, and Canada. So the results presented here may be showing evidence of a marginal return for reductions in class size over a range of sizes that are not characteristic of most American schools. On the other hand, while Israeli class sizes are above the US average, some US states, including California, have average class sizes almost as large as those in Israel (US Department of Education, 1996, Figure 16).

Finally, our study serves to highlight a methodological point. Hanushek's (1995) survey of research on school inputs in developing countries shows the same pattern of weak effects reported in his surveys of research results for the US. The findings presented here suggest that this pattern may be at least partly due to a failure to isolate a credible source of exogenous variation in school inputs. The particular source of variation used here for Israel, a highly nonlinear and non-monotonic association between enrollment size and class size that comes about because of binding resource constraints, leads to a regression-discontinuity design that could be used to identify the causal effects of class size in the US and other countries as well. As far as we know, however, US researchers have yet to exploit the discontinuous relationship between class size and school enrollment that class-size ceilings can generate.

References

Akerhielm, Karen, "Does Class Size Matter?," *Economics of Education Review* 14 (1995), 229-241.

Algrabi, M., "Indices to Characterize a School's Social Composition and a Method for Allocating Additional Funds Between Schools," (Hebrew), *Megamot* 21 (1975), 219-227.

Angrist, J.D., "The Draft Lottery and Voluntary Enlistment in the Vietnam Era, *Journal of the American Statistical Association* 86, (1991a), September.

Angrist, J.D., "Grouped-Data Estimation and Testing in Simple Labor Supply Models," *Journal of Econometrics* 47 (1991b), 243-266.

Angrist, J.D., and Guido Imbens, "Average Causal Response in Models with Variable Treatment Intensity," *Journal of the American Statistical Association* 90 (1995), June.

Boozer, Michael, and Cecilia Rouse, "Intraschool Variation in Class Size: Patterns and Implications," NBER Working Paper No. 5144, June 1995.

Campbell, Donald T., “Reforms as Experiments,” *American Psychologist* 24 (1969), 409-429.

Cook, Thomas D., and D.T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Boston: Houghton-Mifflin Company, 1979.

Card, David, and Alan Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100 (1992a, 1-40.

Card, David, and Alan Krueger, "School Quality and Black-White Relative Earnings: A Direct Assessment," *Quarterly Journal of Economics* 107 (February 1992b).

Card, David, and Alan Krueger, “School Resources and Student Outcomes: An Overview of the Literature amd New Evidence from North and South Carolina,” *The Journal of Economic Perspectives* 10 (Fall 1996), 31-50.

Central Bureau of Statistics, *Educational Institutions: Kindergartens, Schools, Post-Secondary and Higher Education 1990/91*, Series of Education and Culture Statistics, No. 198, reprinted from the *Supplement to the Monthly Bulletin of Statistics*, No. 9, 1991.

Central Bureau of Statistics, *Educational Institutions: Kindergartens, Schools, Post-Secondary and Higher Education 1991/92*, in the *Supplement to the Monthly Bulletin of Statistics*, No. 11, November 1993 (Hebrew only).

Coleman, J.S., et al, *Equality of Educational Opportunity*, Washington, DC: US GPO, 1966.

Deaton, Angus, “Data and Econometric Tools for Development Analysis,” Chapter 33 in *The Handbook of Development Economics, Volume III*, edited by J. Behrman and T.N. Srinivasan, Amsterdam: Elsevier Science B.V., 1995.

Epstein, I., *Hebrew-English Translation of the Babylonian Talmud, Baba Bathra, Volume I*, London: Soncino Press, 1976.

Finn, Jeremy D., and Charles M. Achilles, "Answers and Questions About Class Size: A Statewide Experiment," *American Educational Research Journal* 27 (Fall 1990), 557-577.

Glass, Gene V., and M.L. Smith, "Meta-Analysis of Research on Class Size and Achievement, *Educational Evaluation and Policy Analysis* 1 (1979), 2-16.

Glass, G.V., Cahen, L.S., Smith,, M.L., and Filby, N.N., *School Class Size: Research and Policy*, Beverly Hills: Sage, 1982.

Goldberger, A.S., “Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations,” University of Wisconsin, Institute for Research on Poverty, Discussion Paper 123-72, April 1972.

Hanushek, Eric, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24 (September 1986), 1141-1177.

Hanushek, Eric, “Interpreting Recent Research on Schooling in Developing Countries,” *The World Bank Research Observer* 10 (August 1995), 227-246.

Hanushek, Eric, "School Resources and Student Performance," in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by Gary Burtless, Brookings Institution: Washington, DC, 1996.

Heckman, J.J., Anne Layne-Farrar, and Petra Todd, "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship," NBER Working Paper No. 5274, September 1995.

Hedges, L.V., R.D., Laine, and R. Greenwald, "Does Money Matter? A Meta-analysis of Studies of the effects of Differential School Inputs on Student Outcomes," *Educational Researcher* 23 (1994), 5-14.

Hoxby, Caroline Minter, "The Effects of Class Size and Composition on Student Achievement : New Evidence from Natural Population Variation," Harvard Department of Economics, manuscript, July 1996.

Hyamson, Moses, Annotated english translation of Maimonides' *Mishneh Torah, Book I (The Book of Knowledge)*, New York: Jewish Theological Seminary, 1937.

Klinov, Ruth, "Priorities in Public Resource Allocation for Education," The Center for Social Policy Studies in Israel, Jerusalem: October, 1992.

Lavy, Victor, "Endogenous School Resources and Cognitive Achievement in Primary Schools in Israel," Hebrew University, Falk Institute Discussion Paper No. 95.03, 1995.

Loeb, Susanna, and John Bound, "The Effect of Measured School Inputs on Academic Achievement: Evidence from the 1920s, 1930s, and 1940s Birth Cohorts," NBER Working Paper No. 5331, November 1995.

Moshel-Ravid, *Learning, Teaching, Education, and Class Size: A Review of the Literature*, (Hebrew Pamphlet), Jerusalem: The Henrietta Sczold Institute, the National Institute for Research in the Behavioral Sciences, 1995.

Mosteller, Frederick, "The Tennessee Study of Class Size in the Early School Grades," *The Future of Children: Critical Issues for Children and Youths* 5 (Summer/Fall 1995), 113-127.

Moulton, B.R., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics* 32 (1986), 385-397.

Mueller, D., Chase, C.I., and J.D. Walden, "Effects of Reduced Class Size in Primary Classes," *Educational Leadership* 45 (1988), 48-50.

National Center for Education Feedback, "*The Results of Reading and Mathematics Achievement Tests Given to Fourth and Fifth Graders in June 1991*, (Hebrew pamphlet), Jerusalem: Ministry of Education, October 1991.

National Center for Education Feedback, *Selected Analyses of the Results of the June 1992 Third Grade Achievement Tests*, (Hebrew Pamphlet), Jerusalem: Ministry of Education, March 1993.

OFSTED, *Class Size and the Quality of Education: A Report from the Office of Her Majesty's Chief Inspector of Schools*, London: Office for Standards in Education, November 1995.

OECD, Centre for Educational Research and Innovation, *Education at a Glance: OECD Indicators*, Paris: 1993.

Pakes, A., "On Group Effects and Errors in Variables in Regression," *Review of Economics and Statistics* 65 (1983), 168-173.

Pfefferman, D., and T.M.F. Smith, "Regression Models for Grouped Populations in Cross-section Surveys," *International Statistical Review* 53 (1985), 37-59.

Robinson, G.E., "Synthesis of Research on the Effects of Class Size," *Educational Leadership* 47 (1990), 80-90.

Slavin, R.E., "Class Size and Student Achievement: Small Effects of Small Classes," *Educational Psychology* 24 (1989), 99-110.

US Department of Education, National Center for Education Statistics, *Education in States and Nations* (2nd ed.), NCES-96-160, by Richard Phelps, Thomas M. Smith, and Nabeel Alsalam, Washington, DC: 1996.

Wright, E.N., Shapson, S.M., Eason, G., and J. Fitzgerald, *Effects of Class Size in the Junior Grades*, Toronto: Ontario Ministry of Education, 1977.

Data Appendix

Data used in this study come from Ministry of Education files containing administrative records and test results. Because these files are not maintained for research purposes, we had to link information from a variety of sources, and to impute information that was incomplete in one source with information from other sources.

1991 data (4th and 5th graders)

1. 5th graders

A computerized data file from the Central Bureau of Statistics (1991) survey of schools includes 1,027 Jewish public (secular and religious) schools with 5th grade pupils, in 2,073 (non-special education) classes.[24] These data, containing information collected in September, were given to us by the Central Bureau of Statistics. Data on class size collected between March and June, provided by the Ministry of Education, contained records for 2,052 of these classes, with information on class size for 2,029 of them.

Data on average test scores were provided in two forms. Ministry of Education programmers provided one file with information on average test scores and numbers of test takers for 1,733 of the classes (about 85 percent). We also obtained a file that contained average test scores and numbers of test takers for each grade in each school for 1,978 of the classes. Among the 296 classes missing class-level average scores, school-level averages were available for all but 5. Since there was never more than one class missing a class-level score, and we know the number of test takers in each school and in each class with non-missing scores, we were able to impute the missing class-level average for all but the 5 classes missing both class-level and school-level averages. Finally, the SES index and town ID were added to the linked and imputed class/school data set from a separate Ministry of Education file on schools. The SES index was available for every school in the data base.

[24]The relevant Central Bureau of Statistics (1991, p. 67) report indicates that there were 1,081 Jewish public elementary schools in 1990/91, although not all of these have regular (non-special education) classes and not all have enrollment in all grades.

2. 4th graders

The construction of the 4th graders' data set follows that of the 5th graders. A computerized file from the Central Bureau of Statistics (1993) survey of schools includes 1,039 Jewish public schools with 4th grade pupils, in 2,106 (non-special education) classes. Data on class size, provided by the Ministry of Education, contained records for 2,082 of these classes, with information on class size for 2,059 classes.

We were provided with class-level average scores in 1,769 of the 2,059 classes and school-level averages in 2,025 of the 2,059 classes. Among the 290 classes missing class-level average scores, school-level averages were available for all but 4. Since there was never more than one class missing a class-level score, and we know the number of test takers in each school and in each class with non-missing scores, we were able to impute the missing class-level average for all but 4 the classes missing both class-level and school-level averages. The SES index and town ID were then added as with the 5th graders.

We checked the imputation of class-level averages from school averages by comparing the school and class averages in schools with one class and by comparing the imputed and non-imputed data. School and class-level averages matched almost perfectly in schools with one class. We were unable to detect any systematic differences between schools that were missing some class-level data and the schools that were not. The empirical findings are not sensitive to the exclusion of the imputed class-level averages.

1992 data (3rd graders)

Construction of the 3rd graders data set differs from the construction of the 4th and 5th graders data sets because we were provided with micro data on the test scores of 3rd grade pupils. As with the 4th and 5th graders, we began with the Central Bureau of Statistics (1993) survey of schools. This includes 1,042 Jewish public schools with 3rd grade pupils, in 2,193 (non-special education) classes.[25] Data on class size, provided

[25]The published Central Bureau of Statistic (1993, p. 125) report indicates that there were a total of 1,080 Jewish public elementary schools in 1991/2.

by the Ministry of Education, contained records with information on class size for 2,162 of these classes.

We used micro data on the test scores of 3rd graders to compute average math and reading scores for each class. Score data were available for 2,144 of the 2,162 classes with class size information in the CBS survey of schools. Finally, we added information on the SES index and town ID from a Ministry of Education file containing information on schools. There was no information on the SES index for 34 of the 2,144 classes with data on size and test scores, so that the 3rd grade sample size is 2,111. This is probably because new schools would not have had an SES index assigned at the time data in our school-level file was entered into the record-keeping system.

Table I: Descriptive Statistics

Variable	Mean	S.D.	Quantiles 0.10	0.25	0.50	0.75	0.90
A. Unweighted							
5th grade (2024 classes, 1027 schools)							
Classize	29.93	6.54	21	26	31	35	38
Enrollment	77.70	38.78	31	50	72	100	129
SES index	14.10	13.49	2	4	10	19.5	35
Verbsize	27.32	6.58	19	23	28	32	36
Mathsize	27.72	6.64	19	23	28	33	36
Avgverb	74.44	8.08	64.15	69.85	75.41	79.85	83.34
Avgmath	67.32	10.32	54.84	61.1	67.8	74.09	79.41
4th grade (2053 classes, 1039 schools)							
Classize	30.32	6.39	22	26	31	35	38
Enrollment	78.36	37.72	30	51	74	102	128
Verbsize	27.65	6.55	19	24	28	32	36
Mathsize	28.07	6.56	19	24	29	33	36
Avgverb	72.48	7.98	62.14	67.67	73.33	78.21	82.00
Avgmath	68.86	8.77	57.50	63.59	69.33	74.97	79.42
3rd grade (2162 classes, 1042 schools)							
Classize	30.33	6.43	22	26	31	35	38
Enrollment	79.37	37.33	34	52	74	104	129
Verbsize	24.31	5.80	17	21	25	29	31
Mathsize	24.49	5.80	17	21	25	29	31
Avgverb	86.34	6.11	78.40	83.06	87.24	90.73	93.10
Avgmath	84.11	6.82	75.00	80.24	84.77	89.02	91.92

Table I (cont.)

B. Weighted by class size

	5th grade		4th grade		3rd grade	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Classize	31.40	6.057	31.67	5.887	31.70	5.779
Enrollment	82.99	38.79	82.95	37.53	83.32	36.89
SES index	13.05	12.57				
Verbsize	28.59	6.245	28.78	6.225	25.33	5.333
Mathsize	28.99	6.288	29.20	6.221	25.52	5.328
Avgverb	74.77	7.778	72.67	7.714	86.45	5.942
Avgmath	67.79	9.753	69.14	8.532	84.25	6.667

Notes:

Variable names

class size	number of students in class in the Spring
enrollment	September grade enrollment
SES index	percent of students in the school from "disadvantaged backgrounds"
Verbsize	number of students who took the reading test
Mathsize	number of students who took the math test
Avgverb	average composite verbal score in the class
Avgmath	average composite math score in the class

Table II: OLS estimates for 1991

	5th Grade						4th Grade					
	Reading comprehension			Math			Reading comprehension			Math		
Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean Score		74.3			67.3			72.5			69.9	
(S.D.)		(8.1)			(9.9)			(8.0)			(8.8)	
Class-size	.221	-.031	-.025	.322	.076	.019	0.141	-.053	-.040	.221	.055	.009
	(.026)	(.022)	(.027)	(.032)	(.030)	(.037)	(.028)	(.024)	(.029)	(.030)	(.028)	(.034)
	[.031]	[.026]	[.031]	[.039]	[.036]	[.044]	[.033]	[.028]	[.033]	[.036]	[.033]	[.039]
SES Index		-.350	-.351		-.340	-.332		-.339	-.341		-.289	-.281
		(.011)	(.011)		(.015)	(0.15)		(.011)	(.012)		(.013)	(.014)
		[.012]	[.013]		[.018]	[.018]		[.013]	[.014]		[.016]	[.016]
Enrollment in grade			-.002			.017			-.004			.014
			(.005)			(.006)			(.005)			(.006)
			[.006]			[.009]			[.007]			[.008]
N		2,019			2,018			2,049			2,049	

Notes: Unit of observation is the average score in the class, standard errors in parenthesis.

Table III: Reduced-form Estimates

	5th Graders						4th Graders					
	Classize		Reading comprehension		Math		Classize		Reading comprehension		Math	
Regressor	(1	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean	29.9		74.4		67.3		30.3		72.5		68.9	
(S.D.)	(6.5)		(8.1)		(9.9)		(6.4)		(8.0)		(8.8)	
1. Linear Instrument												
f_{sc}	0.704	0.542	-0.111	-0.149	-0.009	-0.124	0.772	0.670	-0.085	-0.089	0.038	-0.033
	(0.017)	(0.021)	(0.023)	(0.029)	(0.032)	(0.040)	(0.016)	(0.020)	(0.026)	(0.032)	(0.030)	(0.038)
	[.022]	[0.027]	[0.028]	[0.035]	[0.039]	[0.049]	[0.020]	[0.025]	[0.031]	[0.040]	[0.037]	[0.047]
SES Index	-0.076	-0.053	-0.360	-0.355	-0.354	-0.338	-0.054	-0.039	-0.340	-0.340	-0.292	-0.282
	(0.008)	(0.008)	(0.010)	(0.011)	(0.014)	(0.015)	(0.007)	(0.007)	(0.011)	(0.012)	(0.013)	(0.014)
	[0.010]	[0.009]	[0.012]	[0.006]	[0.017]	[0.018]	[0.008]	[0.009]	[0.013]	[0.014]	[0.016]	[0.016]
Enrollment		0.043		0.010		0.031		0.027		0.001		0.019
		(0.003)		(0.005)		(0.006)		(0.003)		(0.005)		(0.006)
		[0.005]		[0.006]		[0.009]		[0.005]		[0.007]		[0.009]
2. Dummy Instrument												
Group 1	-19.487	-15.189	2.444	3.14	-.119	2.199	-19.583	-16.685	3.470	3.604	1.430	2.940
	(.796)	(.799)	(1.023)	(1.088)	(1.404)	(1.487)	(.716)	(.745)	(1.085)	(1.161)	(1.280)	(1.366)
	[.813]	[.848]	[1.039]	[1.138]	[1.431]	[1.571]	[.782]	[.784]	[1.101]	[1.219]	[1.301]	[1.441]
Group 2	-10.073	-7.053	1.920	2.408	.044	1.670	-11.049	-8.995	1.169	1.264	-1.142	-.072
	(.330)	(.366)	(.424)	(.498)	(.583)	(.681)	(.311)	(.356)	(.471)	(.554)	(.556)	(.652)
	[.407]	[.454]	[.495]	[.589]	[.703]	[.834]	[.370]	[.428]	[.552]	[.662]	[.661]	[.790]
Group 3	-3.790	-2.348	.492	.725	.097	.874	-4.750	-3.793	.581	.625	-.075	.423
	(.235)	(.240)	(.302)	(.327)	(.415)	(.447)	(.212)	(.224)	(.322)	(.349)	(.380)	(.411)
	[.326]	[.322]	[.387]	[.412]	[.557]	[.588]	[.281]	[.290]	[.417]	[.448]	[.503]	[.536]
SES Index	-.087	-.058	-.360	-.355	-.354	-.338	-.062	-.041	-.340	-.339	-.291	-.280
	(.008)	(.008)	(.010)	(.011)	(.014)	(.015)	(.007)	(.007)	(.0110	(.012)	(.013)	(.014)
	[.010]	[.010]	[.012]	[.013]	[.017]	[.018]	[.009]	[.009]	[.013]	[.014]	[.016]	[.016]
Enrollment		.053		.009		.028		.036		.002		.019
		(.003)		(.005)		(.006)		(.003)		(.005)		(.006)
		[.005]		[.006]		[.009]		[.005]		[.007]		[.009]
N	2,019		2,019		2,018		2,049		2,049		2,049	

Notes: f_{sc} = enrollment / [INT ((enrollment-1)/40) + 1]

Group 1 = 1 if ($f_{sc} \leq 16$)

Group 2 = 1 if ($16 < f_{sc} \leq 24$)

Group 3 = 1 if ($24 < f_{sc} \leq 32$)

Table IV: 2SLS Estimates for 1991, 5th grade

Regressor	Reading comprehension						Math					
	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV	(7) OLS	(8) IV	(9) IV	(10) IV	(11) IV	(12) IV
Mean (S.D.)			74.4 (8.1)							67.3 (9.9)		
Class-size	-.031	-0.158	-0.275	-0.305	-0.260	-.220	.076	-0.013	-0.230	-0.310	-0.261	-.242
	(.022)	(0.033)	(0.055)	(0.072)	(0.068)	(.066)	(.030)	(0.045)	(0.074)	(0.098)	(0.092)	(.090)
	[.026]	[0.040]	[0.066]	[0.087]	[0.081]	.078]	[.036]	[0.056]	[0.092]	[0.122]	[0.113]	[.108]
SES-Index	-.350	-0.372	-0.369	-0.372	-0.369	-.413	-.340	-0.355	-0.350	-0.355	-0.350	-.345
	(.011)	(0.012)	(0.012)	(0.012)	(0.012)	(.013)	(.015)	(0.016)	(0.016)	(0.016)	(0.016)	(.018)
	[.012]	[0.014]	[0.014]	[0.014]	[0.013]	[.015]	[.018]	[0.019]	[0.019]	[0.019]	[0.019]	[.020]
Enrollment			0.022		0.012	.030			0.041		0.062	.071
			(0.007)		(0.021)	(.020)			(0.009)		(0.028)	(.027)
			[0.009]		[0.026]	[.024]			[0.012]		[0.037]	[.034]
Enrollment^2					0.000	-.000					-0.000	-.000
					(0.000)	(.000)					(0.000)	(.000)
					[0.000]	[.000]					[0.000]	[.000]
Ln(enrollment)				1.648						3.344		
				(0.599)						(0.809)		
				[0.734]						[1.028]		
Region, Urban, and Religion Effects	no	no	no	no	no	yes	no	no	no	no	no	yes
N			2,019							2,018		

Notes: Unit of observation is the average score in the class, standard errors in parenthesis.

Table V: 2SLS Estimates for 1991, 4th grade

Regressor	Reading Comprehension						Math					
	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV	(7) OLS	(8) IV	(9) IV	(10) IV	(11) IV	(12) IV
Mean				72.5						68.9		
(S.D.)				(8.0)						(8.8)		
Class-size	-.053	-0.110	-0.133	-0.097	-0.074	-.053	.055	0.049	-0.050	-0.059	-0.033	-.022
	(.024)	(0.033)	(0.049)	(0.058)	(0.056)	(.054)	(.028)	(0.039)	(0.057)	(0.068)	(0.066)	(.064)
	[.028]	[0.040]	[0.059]	[0.071]	[0.067]	[.064]	[.033]	[0.048]	[0.071]	[0.085]	[0.080]	[.077]
SES-Index	-.339	-0.346	-0.345	-0.347	-0.346	-.389	-.289	-0.290	-0.284	-0.286	-0.284	-.294
	(.011)	(0.012)	(0.012)	(0.012)	(0.012)	(.013)	(.013)	(0.014)	(0.014)	(0.014)	(0.014)	(.016)
	[.013]	[0.014]	[0.014]	[0.014]	[0.014]	[.016]	[.016]	[0.017]	[0.017]	[0.017]	[0.017]	[.019]
Enrollment			0.005		-0.040	-.019			0.020		0.007	.021
			(0.006)		(0.019)	(.018)			(0.007)		(0.022)	(.021)
			[0.008]		[0.024]	[.023]			[0.010]		[0.029]	[.028]
Enrollment^2					0.000	.000					0.000	-.000
					(0.000)	(.000)					(0.000)	(.000)
					[0.000]	[.000]					[0.000]	[.000]
Ln(enrollment)				-0.152						1.290		
				(0.500)						(0.588)		
				[0.619]						[0.741]		
Region, Urban, and Religion Effects	no	no	no	no	no	yes	no	no	no	no	no	yes
N				2,049						2048		

Notes: Unit of observation is the average score in the class, standard errors in parenthesis.

Table VI: Aditional Results for 5th Graders

	Reading Comprehension						Math					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Instrument	linear	linear	3D	3D	3D	linear	linear	linear	3D	3D	3D	linear
Classize	-.115	-.207	-.260	-.242	-.168	-.154	-.215	-.232	-.244	-.297	-.237	-.080
	(.081)	(.065)	(.061)	(.079)	(.075)	(.060)	(.107)	(.089)	(.083)	(.107)	(.104)	(.082)
	[.091]	[.076]	[.071]	[.091]	[.086]	[.073]	[.123]	[.107]	[.099]	[.129]	[.122]	[.104]
SES	-.316	-.646	-.368	-.368	-.643	-.223	-.371	-.522	-.351	-.352	-.522	-.103
	(.019)	(.030)	(.012)	(.012)	(.030)	(.062)	(.025)	(.041)	(.016)	(.016)	(.042)	(.084)
	[.021]	[.035]	[.014]	[.014]	[.035]	[.069]	[.028]	[.050]	[.019]	[.019]	[.051]	[.097]
SES^2		.005			.005			.004			.004	
		(.001)			(.001)			(.001)			(.001)	
		[.001]			[.001]			[.001]			[.001]	
Enrollment	.017	.041	-.021	.007	.032	.021	.067	.079	.042	.070	.081	.037
	(.022)	(.019)	(.007)	(.023)	(.021)	(.006)	(.029)	(.027)	(.009)	(.031)	(.029)	(.009)
	[.025]	[.024]	[.009]	[.028]	[.026]	[.008]	[.034]	[.034]	[.013]	[.019]	[.037]	[.012]
$Enrollment^2$	.000	-.000		.000	-.000		-.000	-.000		-.000	-.000	
	(.000)	(.000)		(.000)	(.000)		(.000)	(.000)		(.000)	(.000)	
	[.000]	[.000]		[.000]	[.000]		[.000]	[.000]		[.000]	[.000]	
Region	no	yes	no	no	yes	no	no	yes	no	no	yes	no
Urban	no	yes	no	no	yes	no	no	yes	no	no	yes	no
Religion	no	yes	no	no	yes	yes	no	yes	no	no	yes	yes
Town Effects	yes	no	no	no	no	no	yes	no	no	no	no	no
Interaction						-.007						-.010
Classize*SES						(.002)						(.003)
						[.003]						[.004]
	2019						2019					

Table VI(cont.): Aditional Results for 4th Graders

	Reading Comprehension						Math					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Instrument	linear	linear	3D	3D	3D	linear	linear	linear	3D	3D	3D	linear
Classize	-.066	-.022	-.180	-.127	-.076	-.089	-.065	-.003	-.074	-.061	-.043	.024
	(.062)	(.053)	(.053)	(.062)	(.060)	(.055)	(.071)	(.064)	(.062)	(.073)	(.071)	(.065)
	[.070]	[.063]	[.063]	[.074]	[.069]	[.066]	[.081]	[.076]	[.075]	[.088]	[.084]	[.080]
SES	-.285	-.604	-.347	-.348	-.606	-.342	-.288	-.422	-.285	-.285	-.423	-.185
	(.017)	(.032)	(.012)	(.012)	(.032)	(.065)	(.020)	(..038)	(.014)	(.014)	(.038)	(.077)
	[.019]	[.038]	[.014]	[.014]	[.038]	[.073]	[.022]	[.046]	[.017]	[.017]	[.046]	[.087]
SES^2		.005			.005			.003			.003	
		(.001)			(.001)			(.001)			(.001)	
		[.001]			[.001]			[.001]			[.001]	
Enrollment	.008	-.011	.009	-.028	.000	.007	.034	.025	.022	.013	.034	.019
	(.020)	(.018)	(.006)	(.020)	(.019)	(.006)	(.023)	(.021)	(.008)	(.023)	(.022)	(.007)
	[.023]	[.023]	[.007]	[.026]	[.024]	[.008]	[.027]	[.028]	[.010]	[.031]	[.029]	[.010]
$Enrollment^2$	.000	.000		.000	.000		-.000	-.000		.000	-.000	
	(.000)	(.000)		(.000)	(.000)		(.000)	(.000)		(.000)	(.000)	
	[.000]	[.000]		[.000]	[.000]		[.000]	[.000]		[.000]	[.000]	
Region	no	yes	no	no	yes	no	no	yes	no	no	yes	no
Urban	no	yes	no	no	yes	no	no	yes	no	no	yes	no
Religion	no	yes	no	no	yes	yes	no	yes	no	no	yes	yes
Town Effects	yes	no	no	no	no	no	yes	no	no	no	no	no
Interaction Classize*SES						-.001						-.004
						(.002)						(.003)
						[.003]						[.003]
N			2049						2049			

Table VII: Estimates for 3rd Graders

	Classize	Reading comprehension				Math			
Regressor	(1) RF	(2) OLS	(3) RF	(4) IV	(5) IV	(6) OLS	(7) RF	(8) IV	(9) IV
Mean (S.D.)			86.3 (6.1)				84.1 (6.8)		
Classize		-.020 (.024) [.027]		-.052 (.039) [.047]	-.040 (.047) [.055]	.023 (.028) [.032]		-.005 (.046) [.056]	-.068 (.055) [.065]
SES Index	-.044 (.007) [.009]	-.176 (.010) [.011]	-.175 (.010) [.011]	-.177 (.010) [.012]	-.177 (.010) [.012]	-.110 (.012) [.013]	-.112 (.011) [.013]	-.112 (.012) [.014]	-.110 (.012) [.013]
Enrollment	.019 (.003) [.005]	.0004 (.004) [.005]	.002 (.004) [.006]	.003 (.005) [.006]	-.006 (.016) [.021]	.006 (.005) [.006]	.008 (.005) [.007]	.008 (.006) [.008]	.058 (.019) [.025]
$Enrollment^2$					.000 (.000) [.000]				-.000 (.000) [.000]
f_{sc}	.691 (.020) [.025]		-.036 (.027) [.033]				-.003 (.032) [.038]		
N	2,111		2,111				2,111		

Notes: Unit of observation is the average score in the class, standard errors in parenthesis.
f_{sc} = enrollment / [INT ((enrollment-1)/40) + 1]

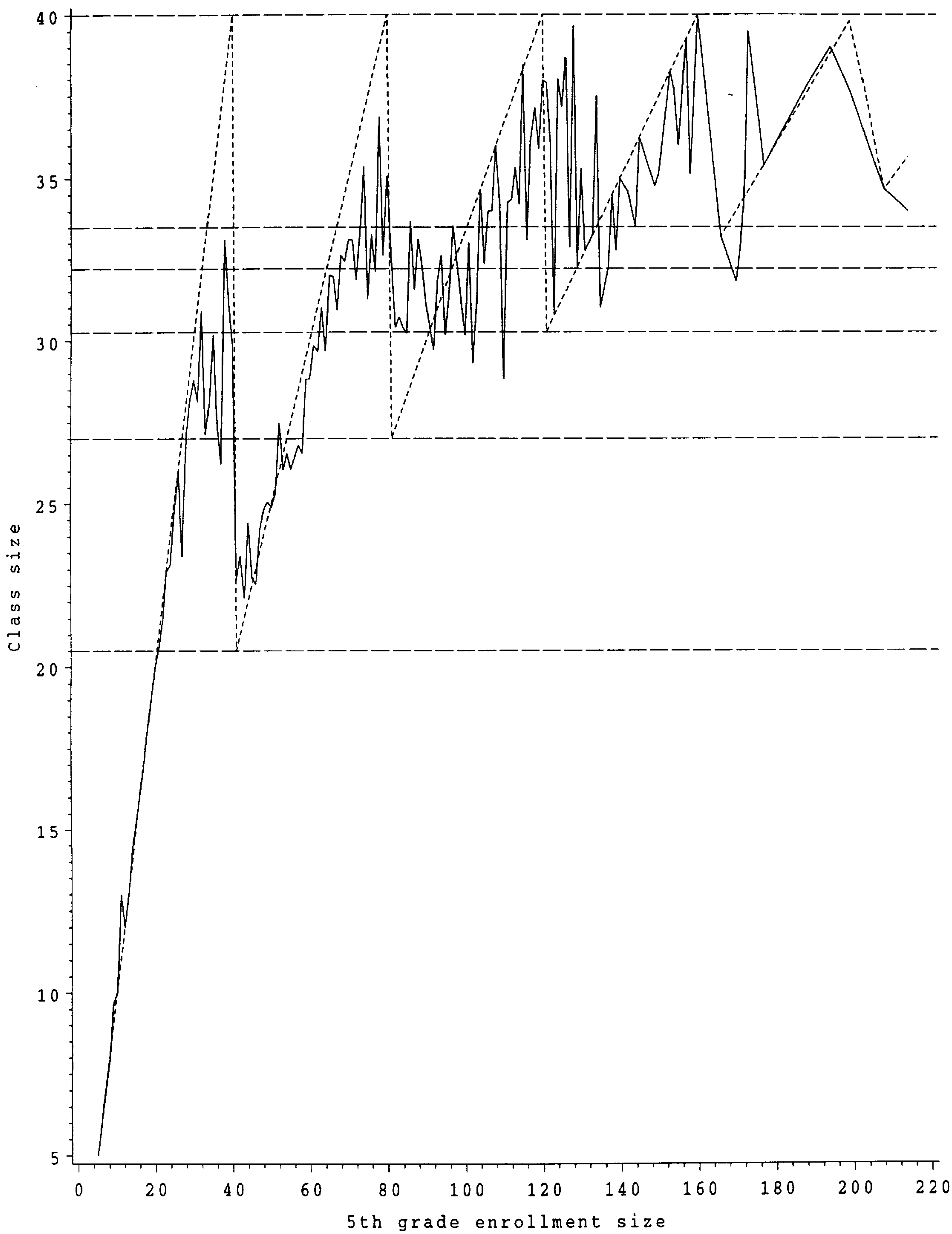

Figure 1. Fifth-grade class size by CBS fifth-grade enrollment count,
for Jewish children in regular public schools.
The figure plots actual size and the size predicted by enrollment.

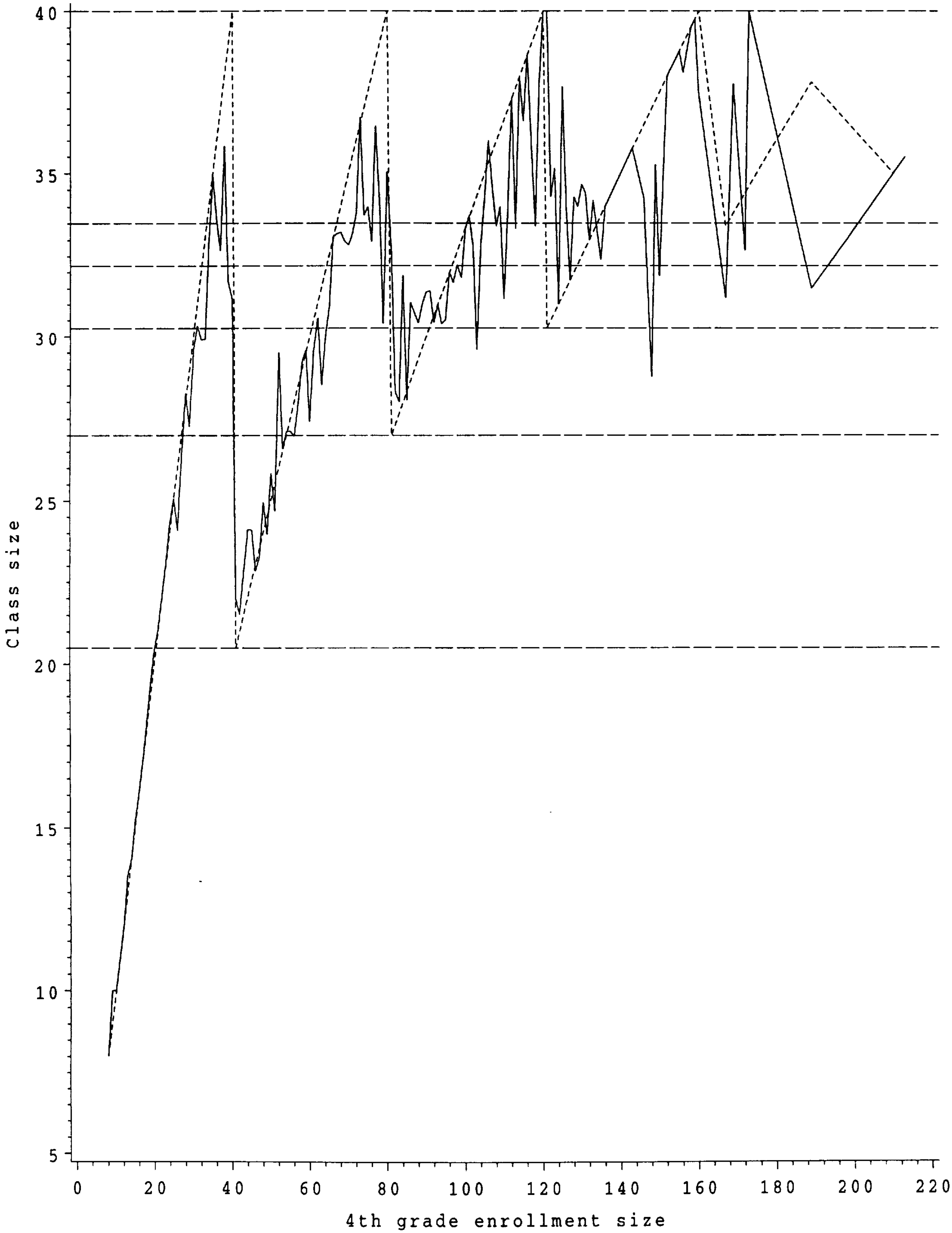

Figure 2. Fourth-grade class size by CBS fourth-grade enrollment count, for Jewish children in regular public schools. The figure plots actual size and the size predicted by enrollment.

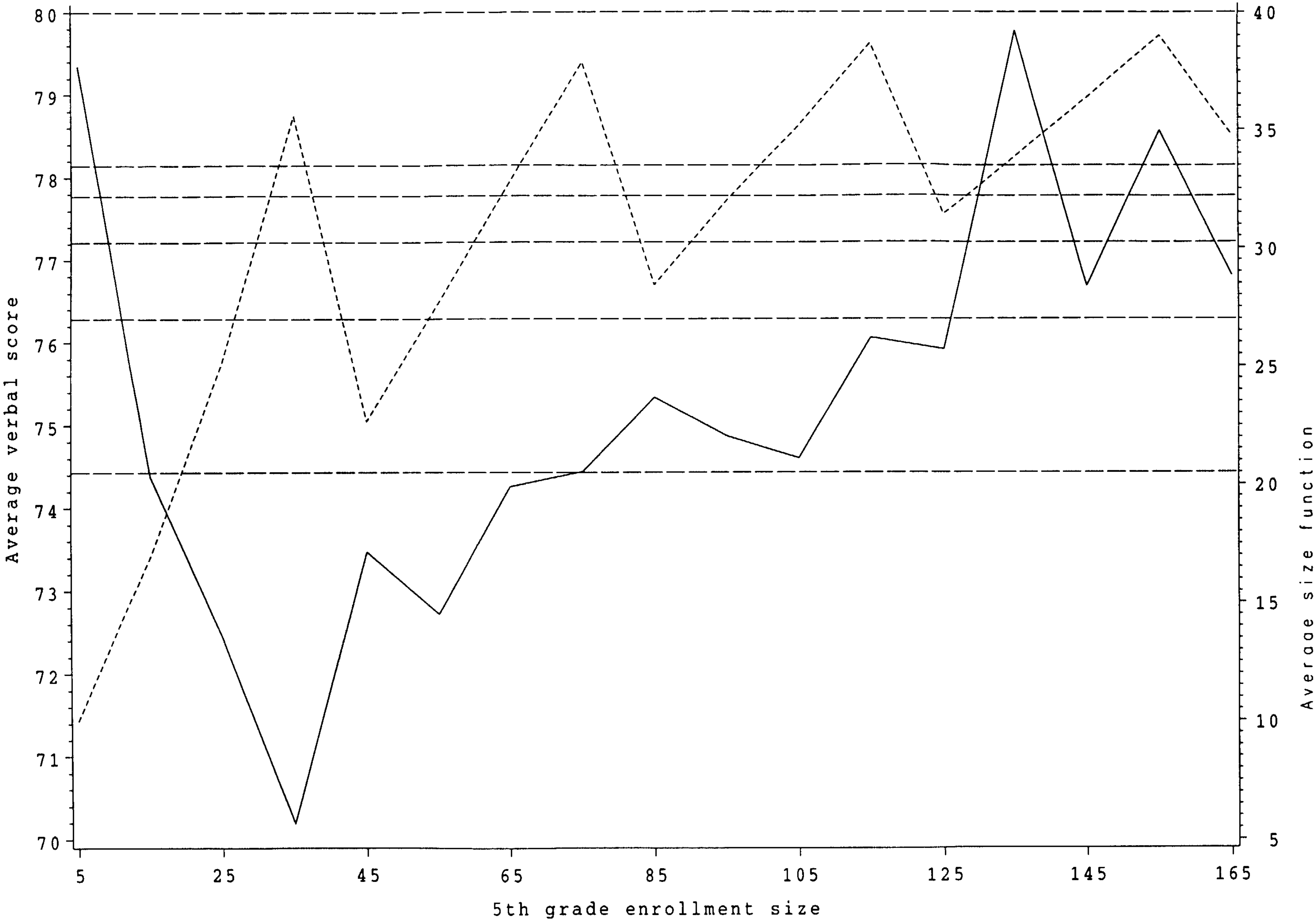

Figure 3a. Average 5TH GRADE VERBAL SCORES by enrollment size, for Jewish children in regular public schools. The figure shows average scores by groups of 10 and the average size function from enrollment data.

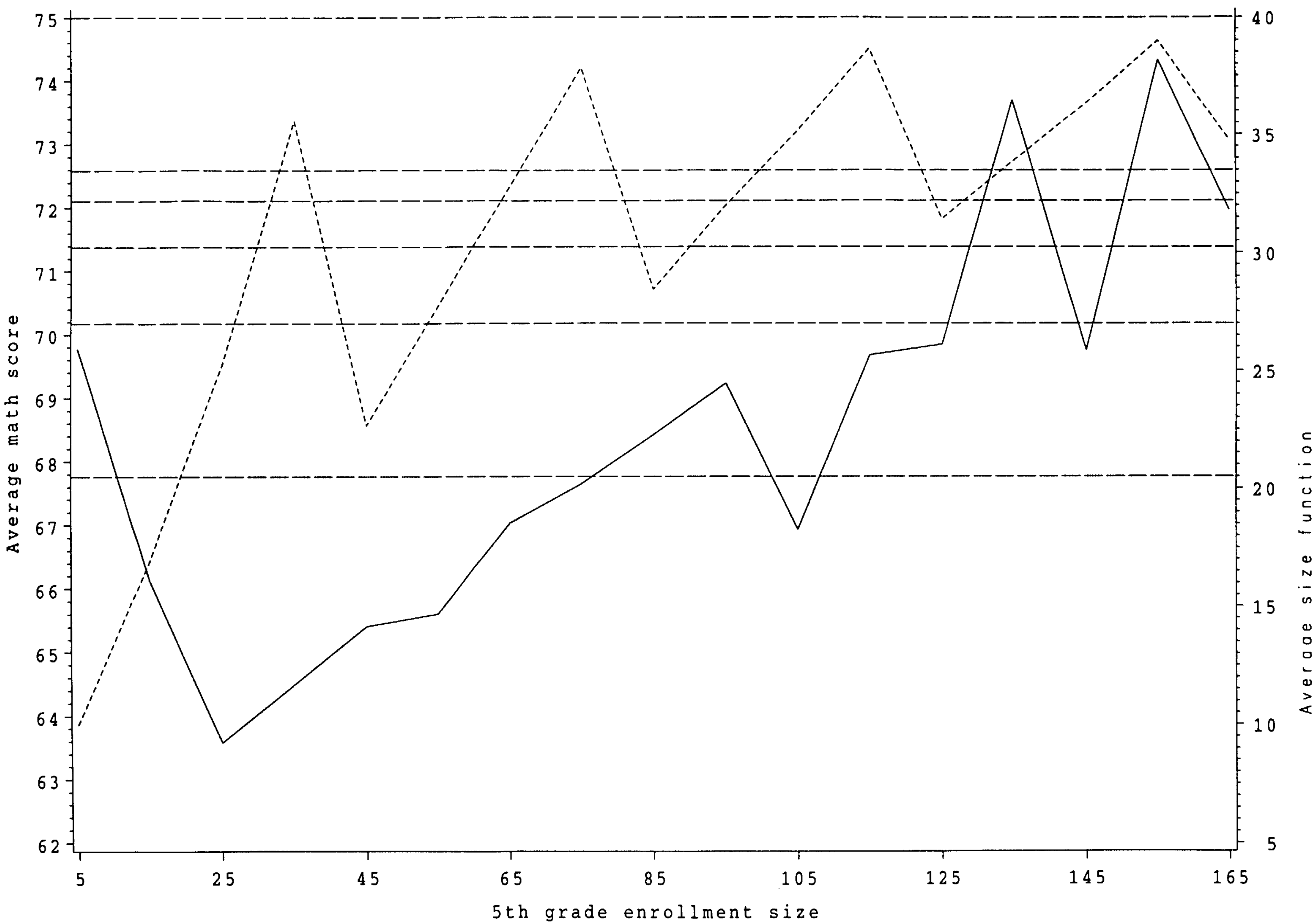

Figure 3b. Average 5TH GRADE MATH scores by enrollment size, for Jewish children in regular public schools. The figure shows average scores by groups of 10 and the average size function from enrollment data.

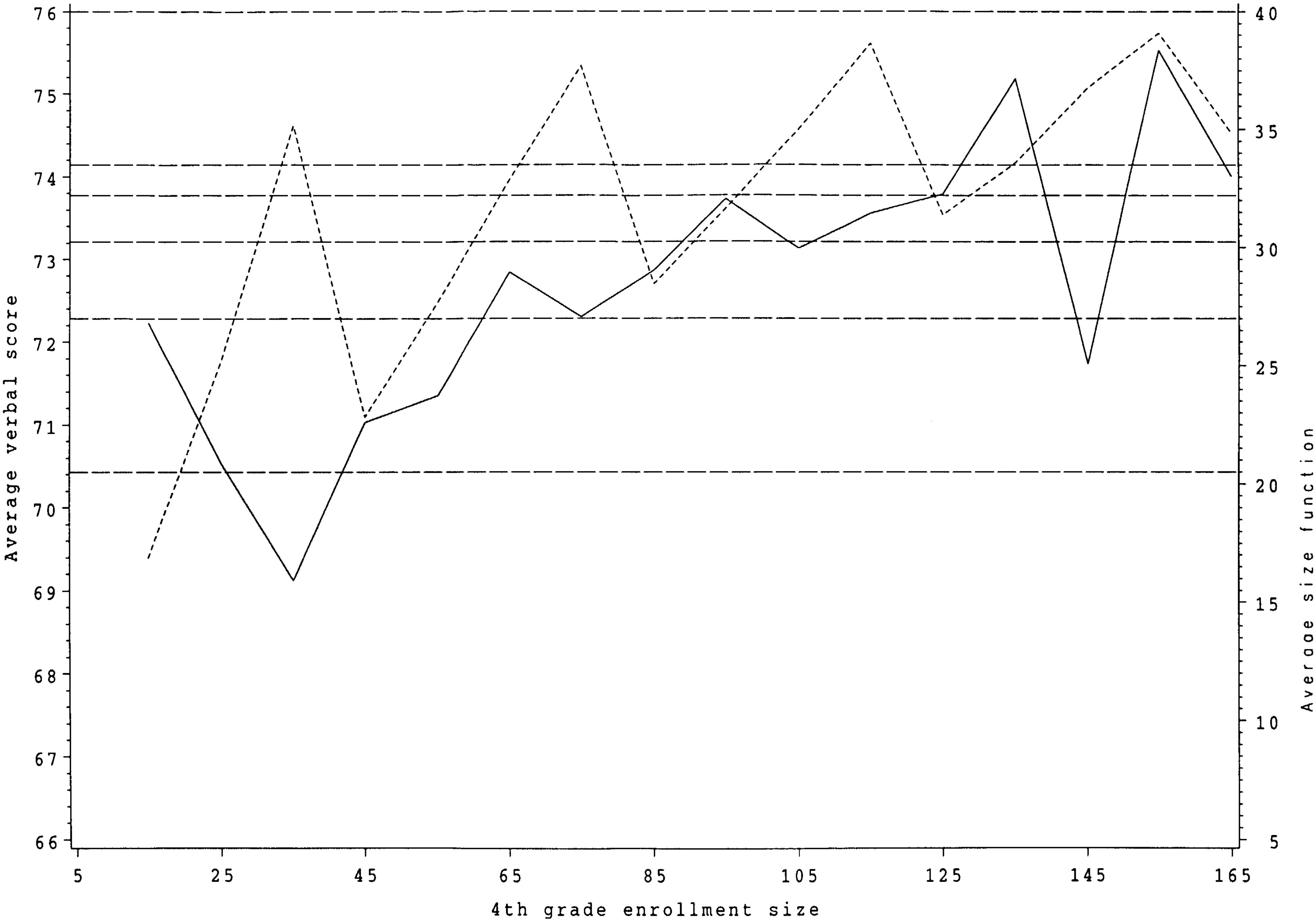

Figure 3c. Average 4TH GRADE VERBAL scores by enrollment size, for Jewish children in regular public schools. The figure shows average scores by groups of 10 and the average size function from enrollment data.

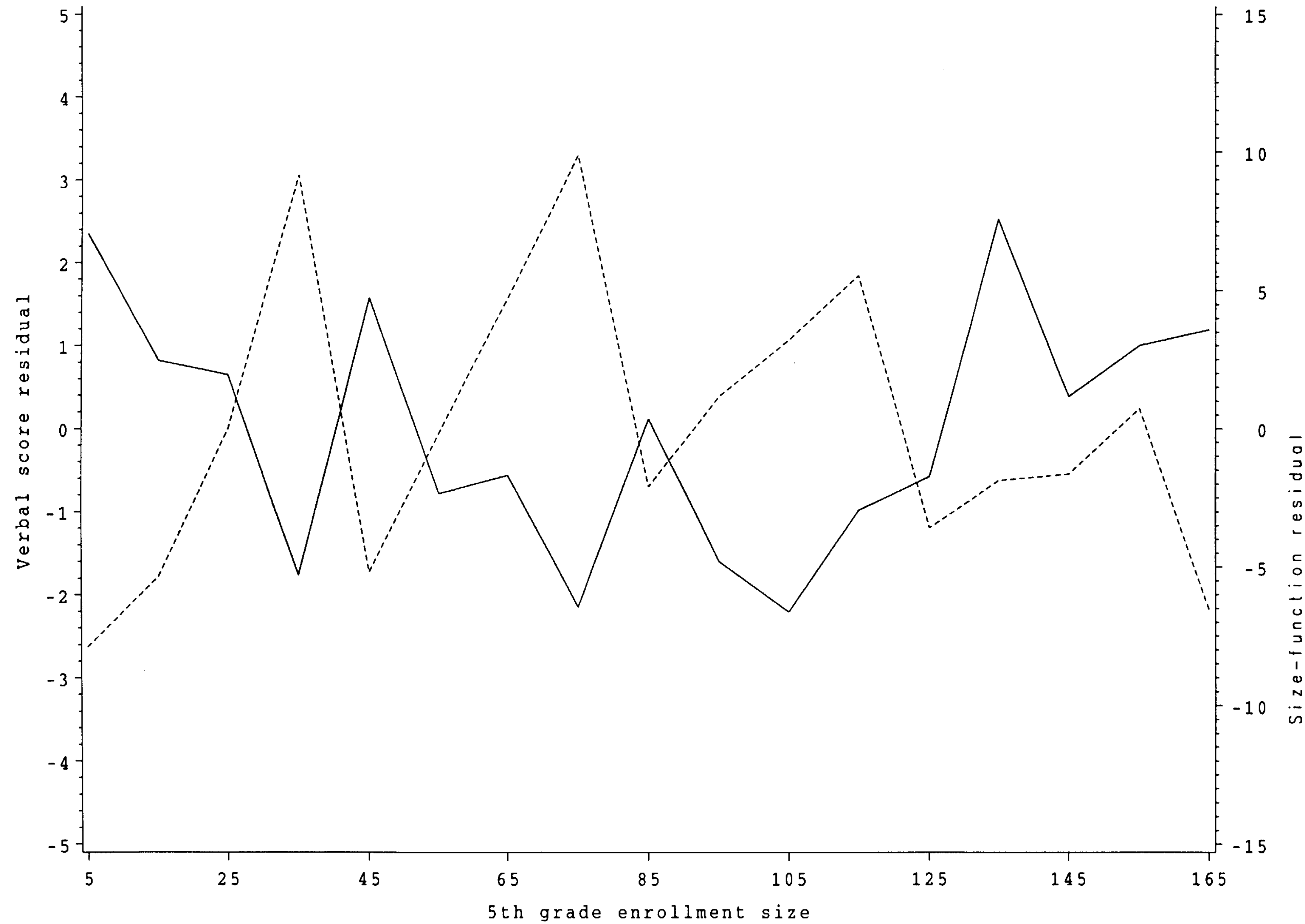

Figure 4a. Average 5TH GRADE VERBAL SCORES by enrollment size -- detrended, for Jewish children in regular public schools. The figure shows residuals from regression of averages on SES and enrollment.

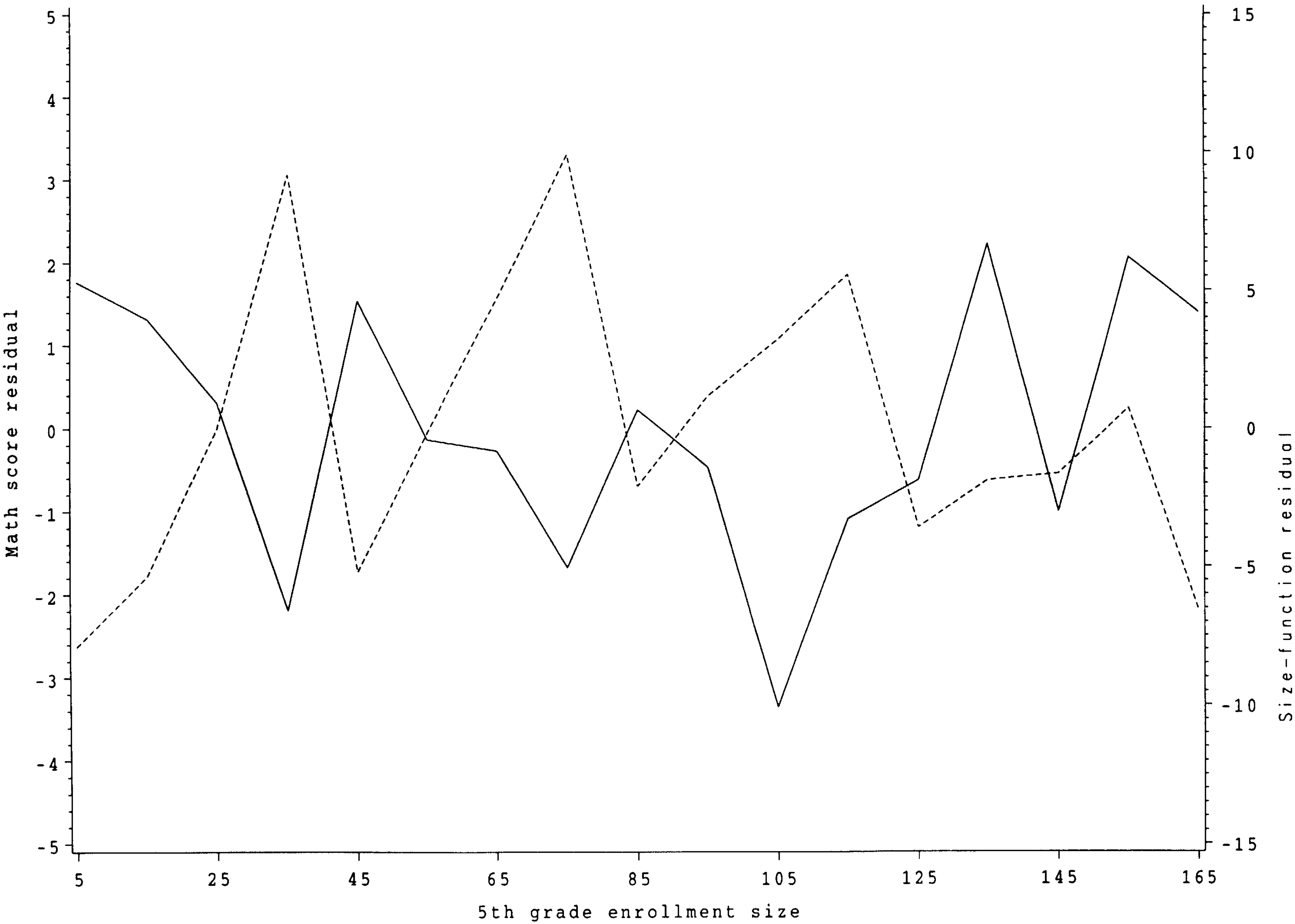

Figure 4b. Average 5TH GRADE MATH scores by enrollment size -- detrended, for Jewish children in regular public schools. The figure shows residuals from regression of averages on SES and enrollment.

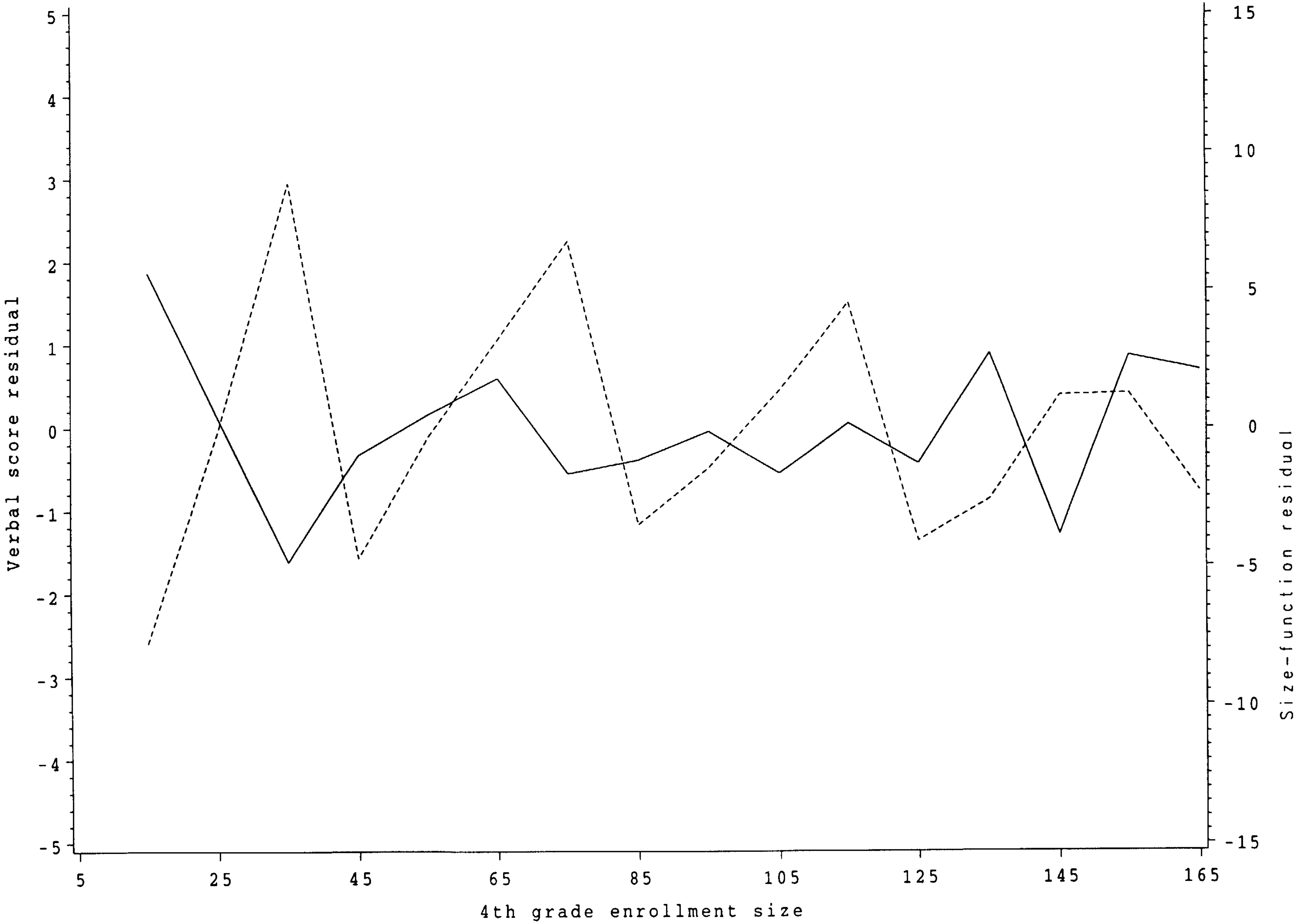

Figure 4c. Average 4TH GRADE VERBAL scores by enrollment size -- detrended, for Jewish children in regular public schools. The figure shows residuals from regression of averages on SES and enrollment.

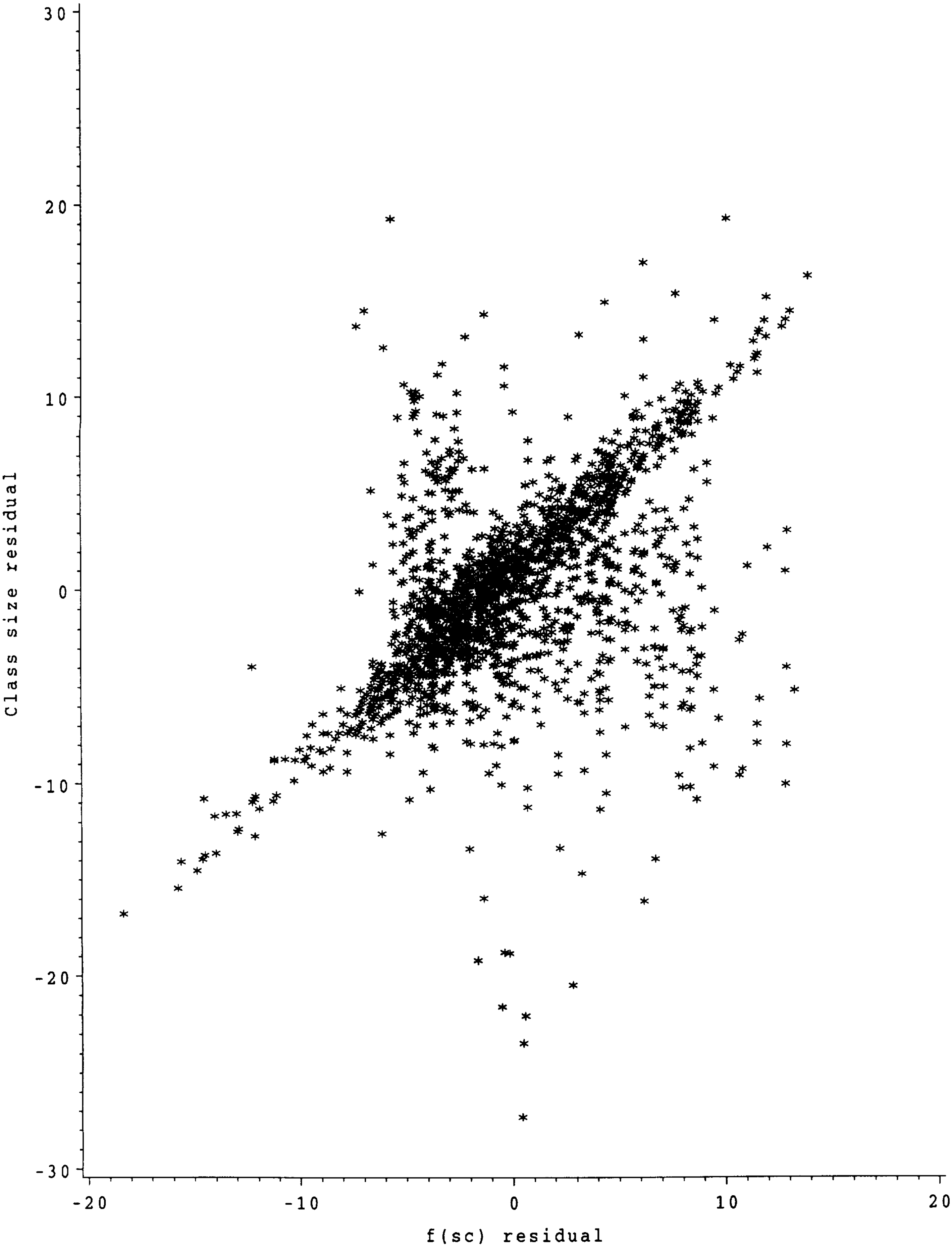

Figure 5a. Fifth-grade first stage,
controlling for SES and enrollment size.

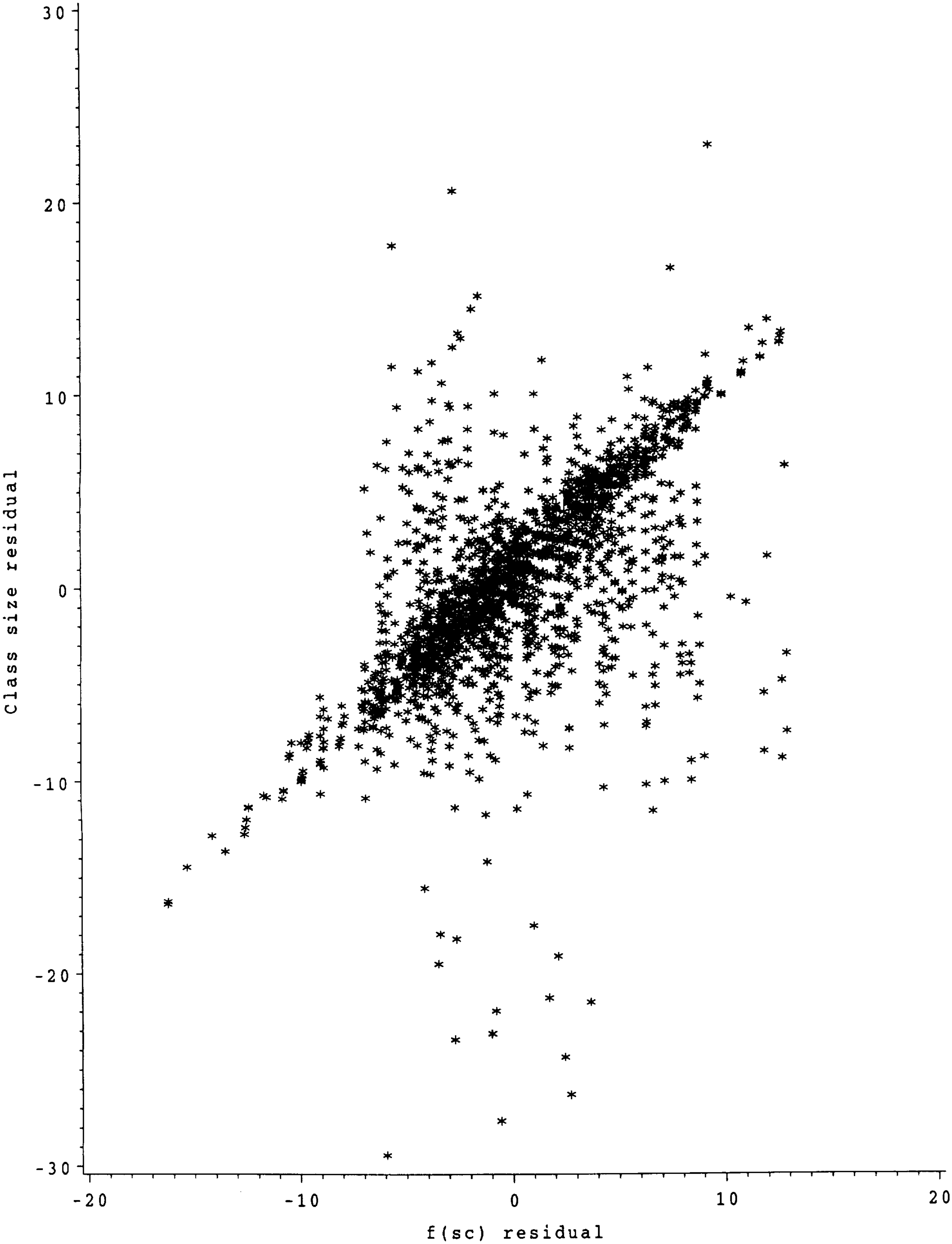

Figure 5b. Fourth-grade first stage,
controlling for SES and enrollment size.