How good is *maimonides' rule?*

strengths and pitfalls of using institutional rules as instruments to assess class size effects

Paulo A. Meyer M. Nascimento

LORDÊLO, JAC., and DAZZANI, MV., orgs. *Avaliação educacional*: desatando e reatando nós [online]. Salvador: EDUFBA, 2009. 349 p. ISBN 978-85-232-0654-3. Available from SciELO Books <http://books.scielo.org>.

How good is *maimonides' rule?* strengths and pitfalls of using institutional rules as instruments to assess class size effects

Paulo A. Meyer M. Nascimento (IPEA)[1]

INTRODUCTION

Class size reductions are commonly perceived as a policy that can positively impact on students' outcomes.

The empirical evidence, though, is not as strong as the conventional belief. Traditionally, no consistent relationship has been identified between class sizes and student achievement. (VIGNOLES, et al., 2000; WEBBINK, 2005) However, a growing number of studies have been challenging the traditional findings by better tackling potential biases caused by omitted variables, such as choices made by parents and schools. Most of these have found some positive effects associated with smaller classes (e.g. ANGRIST; LAVY, 1999; AKERHIELM, 1995; BLATCHFORD, et al., 2002; CASE; DEATON, 1998; KRUEGER, 1999; RIVKIN; HANUSHEK; KAIN, 2005), although some still find either no relationship or larger classes being beneficial (e.g. DOBBELSTEEN; LEVIN; OOSTERBEEK, 2002; HOXBY, 2000). These analyses rely either on controlled experiments or on natural or institutional changes that generate an exogenous variation in class sizes comparable to randomised experiments. (WEBBINK, 2005)

1 The author thanks the British Council and the Bahia State Foundation for Research Support (FAPESB), for financing the studies that produced this and other papers. He is also thankful to Robert Evan Verhine and Anna Vignoles for their valuable comments which helped to improve the paper. All remaining errors are his own.

This chapter will discuss how effectively institutional rules (i.e. laws and regulations, such as Maimonides' Rule in Israel) allow assessing the relationship between class size and student achievement. In order to do that, it will focus on the strengths and limitations of one of the most influential papers among the above cited ones. The study under review is Angrist and Lavy (1999), often referred to by others. It uses an instrumental variable that has been adapted and applied elsewhere. (e.g. DOBBELSTEEN; LEVIN; OOSTERBEEK, 2002; GARY-BOBO; MAHJOUB, 2006; LEUVEN, OOSTERBEEK; RONNING, 2005; URQUIOLA, 2006).

Before exploring in detail the reviewed study, a brief discussion will situate the reader on the methodological issues concerning the evaluation of class size effects (section 2) and the ways to possibly deal with them (section 3). Then the reviewed paper will be described (section 4) and analysed in terms of its purposes (section 5), as well as in relation to the limitations of quantitative methods on policy recommendations – and how adding a qualitative element could provide further insights into the quantitative findings (section 6). Final remarks are subsequently made in section 7.

ISSUES CONCERNING THE EVALUATION OF CLASS SIZE EFFECTS

Education economists often dedicate efforts to research the effects of educational resources on student achievement, and class size is regarded as one possible proxy for resources. Economic evaluations of educational interventions are frequently conducted through education production functions. These see student performance (e.g. in standardised tests) as the outcome of a cumulative process of knowledge acquisition. (TODD; WOLPIN, 2003, cited in WEBBINK, 2005) This implies that today's performance is not a consequence only of current efforts of the child or family and school inputs, but rather a result of the full history of family (dis)advantage, school effects and student endowments.

THE ENDOGENEITY PROBLEM

The virtual impossibility of collecting data on all the possible relevant characteristics of students, their families and schools, reinforced by the failure of the economic theory in specifying all the factors affecting achievement, bring about a key methodological issue. (ANGRIST; KRUEGER, 2001; WEBBINK, 2005, citing KRUEGER, 1999) This issue is *the probable omission of relevant variables* and, consequently, a *bias on the estimated effects of the intervention,* "in case these unobserved factors are related both to the intervention and to the performance indicator". (WEBBINK, 2005, p. 538)

The *endogeneity of school quality* is a special case of the omitted variable bias. It arises because school quality is not independent of schooling choices and, therefore, instead of exogenously determined, it is actually the outcome of the interactions between the decisions made by the educational stakeholders (parents, schools and the government). As long as such stakeholders are able to make choices that affect each other's behaviour and decisions, "school quality will be positively correlated with the wealth and social advantage of children's families [and thus] some of the apparent gain from additional school quality will in fact be a 'return' to pupils' socio-economic background". (VIGNOLES, et al., 2000, p. 10)
In Figure 1 below, the key difficulty would lie in isolating the net effects of school resources (vector 5), assuming that the child's socio-economic characteristics affect his/her achievement both directly and indirectly (through influences on resource allocation and on the child's initial ability).

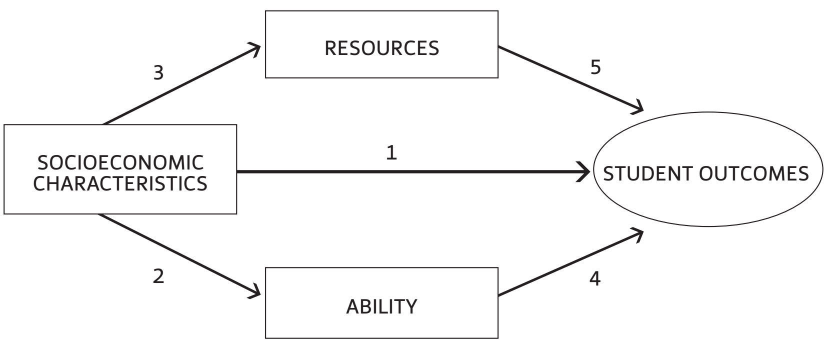

Figure 1: A theoretical model of factors affecting student outcomes
Source: Unnever, Kerckhoff and Robinson (2000)

In the case of assessing the causal effects of class size on student performance, the endogeneity problem reflects two facts. On the one hand, parents may choose where their offspring will study partly based on the size of the classes advertised by schools. On the other hand, the size and composition of the classes are determined by schools. (DOBBELSTEEN, LEVIN; OOSTERBEEK, 2002) Consequently, naïve Ordinary Least Squares (OLS) regressions, by not considering this problem, produce biased estimators that will be inconsistent even in infinite samples, because the covariance between the error term and the class size measure is not zero. (AVERETT; McLENNAN, 2004; LEVACIC, et al., 2005)

Therefore, “the key issue is whether the relationships observed between class size and achievement can be due to further ‘confounding’ factors which have not been accounted for”. (BLATCHFORD, et al., 2002, p. 171)

HETEROGENEITY

Reducing class sizes might impact differently on the achievement of diverse student groups. The benefits of smaller classes may be more associated with certain types of students and thus to specific class compositions.

Indeed, Averett and McLennan (2004, p. 52) summarise that, for studies effectively controlling for the endogeneity of class sizes, “[...] the weight of the evidence indicates to us that smaller classes do result in higher student achievement for some students and some classes”. Hence, it is important to analyse not only the average class size effect for the entire sample, but also its impact on different groups.

Quantile regressions, randomised control trials and matching methods could be cited as examples of techniques that could control for heterogeneity. Alternatively, estimating different regressions for pupils with different characteristics (e.g. gender, social class, ethnicity) could provide some information on the cases when class size reductions are more effective.

FUNCTIONAL FORM

Most of the studies examining class size effects assume a linear relationship between achievement and class size. (BLATCHFORD et al., 2002) This assumption implies the same impact of reductions regardless of initial class size.

Figlio (1999) rejects the linearity assumption for the effects of school inputs on student outcomes and proposes a *translog* functional form. Using quantile regressions, Eide and Showalter (1998, cited in VIGNOLES, et al., 2000) also contest this assumption. Concerning specifically the class size debate, Blatchford, et al. (2002) allow for different marginal effects by employing cubic regression spline.

The uncertainty about the nature of the relationship is attributable to the lack of an established theoretical model guiding empirical research. (FIGLIO, 1999) Notwithstanding, it is expected that quality research will make it possible to test the functional forms used with statistical rigour. (VIGNOLES, et al., 2000)

THE CLUSTERING COMPONENT IN THE DATA

Education is provided to students in different classes, schools and districts or local authorities. The educational experience is likely to be diverse for each pupil, but some common resources (e.g. teacher, facilities) and experiences are shared by those at the same school and, especially, by those in the same class. Therefore, when tested by standardised tests, those students sharing a higher number of common school factors are likely to present a narrower range of test scores in comparison to the entire score distribution across schools. In other words, there are *intraschool correlations in pupils' responses.* (STEELE; VIGNOLES; JENKINS, 2007)

Due to the hierarchical nature of educational data, models aiming to predict the class size effects on student performance should allow for between-pupil, between-classroom and between-school variations. (BLATCHFORD, et al., 2002)

INNOVATIVE METHODS TO IDENTIFY CAUSAL EFFECTS IN EDUCATION

To overcome the methodological issues identified in the previous section researchers employ various methods. The major objective is to generate or isolate exogenous variations in class sizes, in order to deal with the key issue – i.e., the endogeneity problem. These methods are summarised below.

RANDOMISED CONTROLLED TRIALS

The most effective way to deal with the endogeneity problem is to run well-designed randomised controlled experiments (GLEWWE, 2002), because "random assignment of students to an experimental and a control group eliminates bias of factors not observed by the researcher". (KRUEGER, 1999 apud WEBBINK, 2005, p. 540) Nevertheless, experiments in education are costly and may face ethical and political barriers, be affected by attrition and be difficult to replicate in the wider population (LEVACIC, 2005), limiting the applicability of their findings to specific policy interventions. (TODD; WOLPIN, 2003) Furthermore, the absence of a "blind" trial may lead to "Hawthorne effects" (i.e., participants' awareness of being in an experiment and their motivation to make it "work" driving the results towards a particular direction, as pointed out by Hoxby, 2000).

The most famous class size experiment is the American Student/Teacher Achievement Ratio (STAR) project in the 1980s. (DOBBELSTEEN, LEVIN; OOSTERBEEK, 2002; KRUEGER, 1999; LEVACIC, 2005; WEBBINK, 2005) This experiment's results pointed to higher achievement in smaller classes (KRUEGER, 1999), although the validity of such findings has been contested based on the criticisms mentioned above.

MATCHING

Matching is a method "based on the assumption that conditioning on observables eliminates selective differences between programme participants and non-participants that are not correctly attributed to the programme being evaluated". (HECKMAN, ICHIMURA; TODD, 1998, p. 282) This method aims to isolate the effect of an intervention (e.g. class size reduction) by comparing two sets of observation units (the 'treatment' and the 'control' groups) that are assumed identical in all other relevant observable characteristics, except for receiving 'treatment'. This can be done by using the "so-called balancing scores *b(X),* i.e. functions of the relevant observed covariates *X* such that the conditional distribution of *X* given *b(X)* is independent of assignment into treatment". (CALIENDO; KOPEINIG, 2008, p. 32) A possible and widely used balancing score is the propensity score matching, which is "the probability of participating in a programme given observed characteristics *X*". (CALIENDO; KOPEINIG, 2008)

This technique is largely used in very diverse fields of study, as Caliendo and Kopeinig (2008) indicate. In the evaluation of class size effects, however, matching has not been used as much, perhaps because it demands extensive sets of high quality pre-treatment observables.

VALUE-ADDED MODELS

Value-added models include the pupils' previous ability in the equation, either as a control variable or as a component of the outcome variable (e.g. student achievement defined as test score obtained in time 1 *minus* test score obtained in time 0). It improves single equation educational production function models because it controls for most of the child's family, school and learning history up to the point when the baseline test was administrated.

However, value-added models do not sufficiently overcome the endogeneity problem because "even allowing for the child's initial attainment, it may still be the case that the effect of the resourcing that the child actually experiences is systematically related to their family background and prior attainment". (VIGNOLES, et al., 2000, p. 13)

INSTRUMENTAL VARIABLES

Researchers use instrumental variable (IV) methods to obtain consistent estimators when relevant variables are likely to be omitted. (WOOLDRIGE, 2002) The IV approach is a statistical treatment of the data aiming to isolate the exogenous part of the variation in class sizes (i.e. the part that is *not* determined by unobservable factors, such as education stakeholders' behaviour and pupils' ability, and is, therefore, random). The IV is "a variable correlated with the intervention but otherwise unrelated to the test score". (WEBBINK, 2005, p. 540) The net effect of class size, or vector 5 in Figure 1 above, can be estimated with Two Stage Least Squares (2SLS): first, regressing class size on the IV and other exogenous variables, and then regressing the test score on *the variation in class size due to the IV,* together with other controls.

The IV approach has been frequently used to account for endogeneity of school quality variables. (STEELE; VIGNOLES; JENKINS, 2007) Instruments have been mainly derived from *institutional rules* or *natural variation.* (NASCIMENTO, 2008; WEBBINK, 2005) The paper reviewed in the next sections is a widely cited example of an institutional rule taken as an instrument.

ANGRIST AND LAVY'S (1999) ASSESSMENT OF CLASS SIZE EFFECTS IN ISRAEL

THE DATA SET

Angrist and Lavy (1999) use data on test score from a national testing programme administrated in Israeli primary schools at the end of the 1990-1991 and 1991-1992 academic years. Mathematics and (Hebrew) reading skills were tested. Fourth and fifth graders were given the achievement tests in the first year and third graders in the second year. Data on class sizes came from administrative sources and were collected between March and June of the school year starting in September.

The unit of observation is the class – "although micro data on students are available for third graders in 1992, for comparability with the 1991 data, we aggregated the 1992 micro data up to the class level" (pp. 537), which does not seem to concern the authors as "the literature on class size often treats the class as the unit of analysis and not the pupil" (pp. 546).

Average Maths and Reading scores for each class were linked with data on school characteristics and class size from the administrative sources. Specifically, "[t]he linked class-level data sets include information on average test scores in each class, the spring class size, beginning-of-the-year enrolment in the school for each grade, a town identifier, and a school-level index of students' socioeconomic status that we call percent disadvantaged (PD). Also included are the variables identifying the ethnic character (Jewish/Arab) and religious affiliation (religious/secular) of schools" (pp. 537-538).

The PD index, "a function of pupils' fathers' education and continent of birth, and family size" (pp. 538, footnote 8), was identified as a key control variable "because it is correlated with both enrolment size and test scores" (pp. 538). Arab schools were not included in the study, not only because they were not given Reading tests in 1991, but because no PD index was computed or published for them until 1994. Independent religious schools were also excluded, because their curriculum differed significantly from the others. Therefore, the study was restricted to the Jewish public school system. That means over 2,000 classes per grade and about 62,000 pupils in total were considered.

The average public school class in the data set had about 30 pupils, and in average 78 students were enrolled per grade in each school. Ten percent of classes had more than 37 pupils, and ten percent had less than 22.

The authors note that the average score distributions were similar for fourth and fifth graders, "but mean scores are markedly higher, and the standard deviations of scores lower for third graders" (pp. 538). They attribute these differences to a probable "[...] systematic test preparation effort on the part of teachers and school officials in 1992, in light of the political fallout resulting from what were felt to be disappointing test results in 1991". (ANGRIST; LAVY, 1999)

"MAIMONIDES' RULE" AS AN INSTRUMENT

Angrist and Lavy (1999) explain that the Israeli public school system determines the division of enrolment cohorts into classes based on a maximum class size rule derived by an interpretation of the Babylonian Talmud made by the great twelfth century Rabbinic scholar, *Maimonides*. This consuetudinary rule states a maximum class size of *40*.

The rule can be used as an identification method as it "induces a discontinuity in the relationship between enrolment and class size at enrolment multiples of 40" (pp. 540). It means that class size varies "abruptly and predictably" (HOXBY, 2000, p. 1252) every time enrolment exceeds a multiple of 40, when a new class needs to be formed. Put another way, at the points of discontinuity the school average class size drops from 40 (when grade enrolment is 40, 80, 120 etc) to 20.5 (when the 41st, 81st, 121st etc student is enrolled), reflecting an exogenous variation created by the institutional rule.

Therefore, Maimonides' rule may be a good instrument for the discontinuity sample (group of schools with enrolments in a range close to the points of discontinuity).

Angrist and Lavy (1999) first propose a mixed-effect model that accounts for school characteristics and class size, plus error terms for each level (class, school and pupil). Student background (PD index) is aggregated at the school level and goes into the vector of school characteristics. For the *i*th student in class *c* and school *s*, the authors write the education production function as follows:

$$y_{isc} = X'_s\beta + n_{sc}\alpha + \mu_c + \eta_s + \epsilon_{isc} \qquad (1)$$

They interpret the above equation as describing the average potential outcomes of students under alternative assignments of class sizes, controlling for any effects of school characteristics and accounting for random effects related to the child, the class and the school.

Given that they are only able to do comparable analysis between grades at class-level, because the micro-level data are unavailable for fourth and fifth graders, and considering the literature tendency to treat the class as the unit of analysis, the authors group the first equation in the class-level estimating equation of the form below (over bars denote averages):

(2)
$$\bar{y}_{sc} = X'_s\beta + n_{sc}\alpha + \eta_s + [\bar{\mu}_c + \bar{\epsilon}_{sc}]$$

These initial equations, however, do not account for the endogeneity problem. To do that, a 2SLS is performed. The notation for the first-stage relationship is expressed as (f_{sc} being the class-size function derived from the Maimonides' rule):

(3)
$$n_{sc} = X'_s\pi_0 + f_{sc}\pi_1 + \xi_{sc}$$

Therefore, the institutional rule of no more than 40 is the instrument predicting the actual class size experienced by a particular student.

After equation 3 is estimated, the second stage is done using equation 2, which gives, for the discontinuity sample, the effect on the test scores of the variation in class size due to the IV. The authors present a variety of both OLS and 2SLS estimates for each grade.

MAIN FINDINGS AND CONCLUSIONS

Angrist and Lavy (1999) find some positive effects on class size reductions, particularly for fifth graders, for whom such effects were large for both Reading and Maths scores. Smaller effects were found for Reading scores of fourth graders, whose Maths scores were not significantly affected by the size of the classes, "though pooled estimates for fourth and fifth graders are significant and precise on both tests" (pp. 569). No relationship was found for third graders. The authors conjecture that these results might indicate a cumulative effect of class size reductions (i.e. years experiencing small classes may be required before any benefits become noticeable).

Published when a bill imposing a new maximum class size of 30 was being discussed in the Israeli Parliament, Angrist and Lavy's (1999) paper asserts that such a change "could have an impact equivalent to moving two deciles in the 1991 distribution of class averages" (pp. 569).

It is also considered that the results presented are unlikely to be relevant for other countries, where schools may experience other class size averages and range, in addition to cultural, political and economic differences. This is a consequence of the fact that "instrumental variables provide an estimate for a specific group—namely, people whose behavior can be manipulated by the instrument". (ANGRIST; KRUEGER, 2001, p. 77)

Finally, the study highlights that without tackling the endogeneity problem, the findings would guide to misleading conclusions, as "[t]he raw positive correlation between achievement and class size is clearly an artefact of the association between smaller classes and the proportion of pupils from disadvantaged backgrounds" (pp. 569), suggesting that those who look solely at traditional literature may be prematurely assuming no causal effects of school resources on learning.

A CRITIQUE OF ANGRIST AND LAVY (1999)

THE DISTINCTIVENESS OF THE STUDY

Angrist and Lavy (1999) became quickly a classical reference for discussing the IV approach as a method to isolate the exogenous portion of class size determination. This may be partly due to the clarity and simplicity of the adopted identification strategy, and mostly because, for specific class size ranges, the instrument shows correlation with the explanatory variable of interest and is generally accepted as uncorrelated with achievement.

The study is, therefore, illustrative of how institutional rules can establish a discontinuous relationship between enrolment and class size and, consequently, be considered analogous to running a field experiment. (DOBBELSTEEN, LEVIN; OOSTERBEEK, 2002)

Moreover, it is certainly relevant, in terms of policy recommendations, for the Jewish educational public system in Israel, where Maimonides' rule operates.

THE LIMITATIONS OF THE STUDY

In spite of its contributions to ongoing class size debate, the study by Angrist and Lavy (1999) suffers from a number of limitations.

First, as the authors themselves point out, the presented findings are unlikely to be relevant for policy decisions in any setting other than the Jewish public school system in Israel. This is so because different educational systems are subject to different incentives and constraints raised by cultural, social, economic, political and institutional factors. Even within Israel, the school systems excluded from the analysis (Arab schools and Jewish independent private schools) are unlikely to present the same responsiveness to class size rules.

Second, as admitted by the authors, the adopted identification strategy is valid only for a restricted sub-sample of Jewish public schools (the discontinuity sample) where grade enrolment sizes are around the class discontinuity points (40, 80, 120 etc). As Hoxby (2000, p. 1253) puts it:

> [I]f one uses a rule that binds only in some schools, one learns about the effects of class size only for those schools. For instance, in Angrist and Lavy's [1999] data, the maximum class size rule does not bind in districts that serve well-off households. It is useful to estimate the effect of class size only for less-well-off students, but one must be careful to interpret the results appropriately. If better-off districts actually have maximum class size rules of their own that they follow, then using a statewide rule that does not bind everywhere is throwing away useful variation.

Indeed, the analysis of Angrist and Lavy (1999) is relevant only for the discontinuity sample, which includes solely schools with grade enrolments in the set of intervals {[36,45], [76,85], [116,125]}. This corresponds to less than 25% of all classes. Concerning the groups most affected, though, Angrist and Lavy (1999, p. 567) remark that:

> [T]he analysis of affected groups indicates that the estimates presented here are affected disproportionately by smaller schools and by schools with fewer disadvantaged pupils than

> average, although the variation in impact along this second dimension is much more modest than the first.

Apart from the groups that are most affected by the Maimonides' rule, two different class or school sub-samples may present distinct class size ranges within which variations are worth checking in order to identify the upper size limit above which reductions are more efficacious. (HOXBY, 2000) For example, for one sub-sample the useful variation in class size may be situated between 16 and 25 students, while for another it could be between 20 and 29. Then it would be unclear whether "the greater efficacy was due to decreasing returns to reductions in class size or greater efficacy in the sort of schools that typically choose higher maximum class sizes". (HOXBY, 2000, p. 1254, footnote 10) This problem would demand an examination of the observation units with lower and higher maximum class sizes – which Angrist and Lavy (1999) are unable to do for the classes and schools unaffected by the Maimonides' rule.

Hoxby (2000) draws attention to another shortcoming of Angrist and Lavy (1999). She demonstrates that "predicted class size is *not* a valid instrument *except* when the rule triggers a change in the number of classes" (HOXBY, 2000, p. 1252), because it varies with actual enrolment, which is, in turn, correlated with school characteristics and the residual term. This implies a requirement to discard the "suspected" observations (those not affected by the rule) that lie between the discontinuity points, although "Angrist and Lavy [1999] [...] are able to do only some of the desirable discarding because their cross-sectional data contain too few occurrences of enrolment in the right ranges". (HOXBY, 2000, p. 1253)

OVERALL STRENGTHS AND WEAKNESSES

Given the aspects highlighted in subsections 5.1. and 5.2. above, it could be argued that Angrist and Lavy (1999) represents a key paper in the class size debate because it pioneered the application of an institutional rule that does generate exogenous variations in class sizes. Although the study is a worldwide reference for research on the effects of class size reduction, the findings are specifically relevant for a particular educational setting – a shortcoming that has been widely observed.

The overall strengths and weaknesses of Angrist and Lavy (1999) are better assessed by looking at the extent to which they effectively deal with the methodological issues summarised in section 2.

The endogeneity problem

The paper does address the endogeneity problem. It uses the Maimonides' rule on maximum class sizes to isolate the net effect of class size variations on student achievement in Jewish public schools in Israel.

However, as discussed previously, the predicted class size derived from the Maimonides' rule is a valid instrument only for the cases where the institutional rule generates discontinuities in the average class sizes – this is called LATE – Local Average Treatment Effect; for elaboration on this, see Imbens and Angrist (1994), Angrist and Krueger (2001).

Therefore, for most of the total sample of classes, the endogeneity problem persists.

Heterogeneity

The paper controls for differences in the impact of class size reduction across groups by characterising affected groups on the basis of some observables and comparing their means. Heterogeneity is thus tackled for the groups of classes and schools for which the authors have data to run comparisons (namely, school and class composition with proxy provided by the PD index, and school size based on number of enrolments).

Additionally, it is worth keeping in mind that the Maimonides' rule seems to be effective only at schools with a high PD index. This fact *per se* restricts, regardless of the authors' intents, the paper's estimates to disadvantaged schools – a point highlighted by Hoxby (2000).

The functional form

As discussed before, there is no theoretical model guiding the class size debate on the appropriateness of alternative functional forms to describe hypothesised relationships with student achievement. Nonetheless, methodologically strong studies are expected to test different educational production functions. The reviewed article does not fail in that respect.

The clustering component in the data

The authors acknowledge clustering in a conventional econometric manner; they allow for class and school fixed effects and adapt the standard errors to capture clustering components. Nonetheless, they do not use the most efficient method to handle hierarchical data – i.e. multilevel models as described in statistical textbooks such as Bryk and Raudenbush (2002), Snijders and Bosker (1999) and Goldstein (1995).

In spite of these considerations, it is worth adding that the instrumental variable method used by Angrist and Lavy (1999) and by most of the recent research on education production functions (including Hoxby, 2000) is not exempted from questioning on its effectiveness to tackle endogeneity and the other methodological problems discussed here (see Steele, Vignoles and Jenkins (2007) for a detailed discussion).

BEYOND THE BIG NUMBERS: FURTHER INSIGHTS THAT COULD HAVE BEEN ADDED BY A QUALITATIVE STUDY

Some of the findings presented by Angrist and Lavy (1999) do not provide elements for vigorous conclusions or policy recommendations. This is exemplified in the following quote from the reviewed paper:

> The average score distributions for fourth and fifth grade classes are similar, but mean scores are markedly higher, and the standard deviations of scores lower for third graders. We believe the difference across grades is generated by a systematic test preparation effort on the part of teachers and school officials in 1992, in light of the political fallout

> resulting from what were felt to be disappointing test results in 1991. (ANGRIST; LAVY, 1999, p. 538)

The finding under consideration here is the significant difference in average score distributions between fourth and fifth graders, tested in the first year of the programme (1991), and third graders, tested in the programme's second year (1992). Based solely on quantitative data, Angrist and Lavy (1999) were only able to make an informed guess in attempting to explain the observed difference. They conjecture that such differences are possibly a result of the teachers' and school officials' response to the negative repercussion of overall bad results for fourth and fifth graders in 1991. The authors hypothesise that in 1992, as a reaction for the results obtained in the year before, teachers and school officials may have put extra effort to prepare third graders to sit for the national exams that year.

This is the authors' hypothesis. However, they cannot ensure that this was the case, because they only have quantitative data available. To uncover the actual behaviour of teachers and school officials after the release of the 1991 test results, and whether this behaviour was any different and likely to impact on pupils' achievement, a complementary qualitative research element would have been necessary. This qualitative element could assume various designs, such as semi-structured interviews, focus groups, observation or some other form of the ethnographic approach, and should be conducted with teachers and school administrators at the schools, particularly those that shown a bigger difference in average scores between the tests administrated in 1991 and 1992.

The complementarily roles of quantitative and qualitative methods in educational research have been identified and defended by many educational analysts. (EISNER, 1979; LOCKHEED; HARRIS, 2005; SALOMON, 1991) The interactions between the two methods enrich the research design, providing a wider range of conclusions and a deeper understanding of policy implications:

> Statistical models are good at estimating overall effects and, when extended into a multilevel framework, enable the

> variability of effects to be estimated. If, however, we aim for an understanding of just why this variability arises then data generated from qualitative approaches […] are valuable. (PLEWIS; MASON, 2005, p. 185)

Salomon (1991) discusses the complementary roles of quantitative and qualitative methods with examples of studies that seem to address the same question but are, in reality, exploring completely different aspects of it and reaching, therefore, considerably different conclusions – which are not necessarily mutually excluding. Greene et al. (1989), cited in Lockheed and Harris (2005), enumerates three ways in which quantitative and qualitative approaches complement each other:

- Qualitative studies are used first to construct theory and guide the development of instruments later used in quantitative studies.
- Qualitative methods are used at the same time as quantitative ones for triangulation.
- Qualitative methods are used after quantitative ones to provide deeper understandings of the findings brought by quantitative data.

In the example gin this section, an application of the third way is suggested: a qualitative field research follows the achievement tests to find out how the intervention impacts on the behaviour of the agents involved.

Notwithstanding, any of the three ways listed above could be used to shed light on gray areas of the educational production function approach. For example, quantitative data might reveal some resources being more effective than others, but they do not reveal *how* the use of certain resources, rather than others, is translated into better student achievement. It is likely that a triangulation, for example, using both quantitative and qualitative methods could provide some useful insights on that. Similarly, it is possible that the theoretical model that economists so frequently proclaim as lacking in the education production function approach will eventually emerge from qualitative studies aiming to map the inter-relationships that define the direction and strength of the vectors identified in Figure 1 shown in section 2 of this essay.

Certainly, from the findings presented by Angrist and Lavy (1999), yet other relevant information could be obtained by policy makers from qualitative studies. In particular, one could desire to know why some groups are more sensitive to class size reductions than others. What makes disadvantaged students benefit more from smaller classes than their better-off peers? And what changes in the class environment are perceived by teachers, school officials, parents and students that might explain why, for those attending smaller classes, the test scores were significantly different from their counterparts with similar background who happened to be in schools presenting enrolment sizes just under the discontinuity point?

Answers for these questions would not come from the data analysed by Angrist and Lavy (1999). If such answers can be found at all, then it is more likely that these would emerge from qualitative data, focusing, for example, on changes in the way teachers teach and pupils behave in smaller classes. Therefore, Israeli policy makers could probably have benefited much more from the 1991-1992 national testing programme if they had followed the quantitative data collection up with a qualitative research on the changes in school environment and on the perceptions and behaviour changes of teachers, school administrators, parents and pupils.

FINAL REMARKS

This chapter began by discussing the methodological issues concerning the evaluation of class size effects, and then summarised how the relevant literature deals with them. This sets the ground to analyse the strengths and weaknesses of Angrist and Lavy (1999). The paper by these two authors is well-known for taking advantage of an institutional rule on maximum class size existing in the Jewish public educational system in Israel to construct an effective instrumental variable that causes discontinuity in specific enrolment sizes and thus induces an exogenous variation in class sizes for sub-samples situated just after the point of that discontinuity. Following the critique on the reviewed study, some examples were given of how a qualitative element could have provided further insights into the quantitative findings.

It was shown that, despite some shortcomings that do not diminish their distinctiveness to the related literature, Angrist and Lavy (1999) provide useful policy insights for the class size debate in Israel, although the policy implications could have been far clearer if a qualitative approach attempted to uncover some points that were out of the inference capacity of quantitative methods.

Overall, the use of institutional rules as an instrumental variable to assess class size effects should be used with caution. When using an institutional change such as the Maimonides rule, one can look, at most, at its effects on a particular sample extracted from a specific context that was directly affected by the law statement. Researchers eager to apply similar instruments must be very careful when drawing conclusions, as their findings are likely to face the same constrains pointed out here to Maimonides´ rule. They should also be aware that there is more underlying the relationship between class sizes and student achievement than numbers alone can uncover – and thus a research design combining quantitative and qualitative methods would potentially provide broaden and deeper insights on this topic.

REFERENCES

AKERHIELM, K. Does class size matter? *Economics of Education Review,* v. 14, n. 3, p. 229-241, Sep. 1995.

ANGRIST, J. D.; LAVY, V. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, v. 114, n. 2, p. 533-575, Mar. 1999.

______; KRUEGER, A. B. Instrumental variable and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives*, v. 15, n. 4, p. 69-85, 2001.

AVERETT, S. L.; McLENNAN, M. C. Exploring the effect of class size on student achievement: what have we learned over the past two decades?

In: JOHNES, G.; JOHNES, J. (Ed.). *International handbook on the economics of education*. Cheltenham, UK; Northhampton, MA: Edward Elgar, c2004.

BLATCHFORD, P. et al. A study of class size effects in english school reception year classes. *British Educational Research Journal*, v. 28, n. 2, p. 169-185, 2002.

BRYK, A. S.; RAUDENBUSH, S. W. *Hierarchical linear models:* applications and data analysis methods. 2nd ed. Newbury Park: Sage Publications, 2002.

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, v. 22, n. 1, p. 31-72, 2008.

CASE, A.; DEATON, A. *School quality and educational outcomes in South Africa*. Princeton, New Jersey: Center for Research on Child Wellbeing, 1998. (Working Paper, #98-08). <http://crcw.princeton.edu/workingpapers/WP98-08-Case.pdf.>

DOBBELSTEEN, S.; LEVIN, J.; OOSTERBEEK, H. The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics*, v. 64, n. 1, p. 17-38, Feb. 2002.

EISNER, E. W. The use of qualitative forms of evaluation for improving educational practice. *Educational Evaluation and Policy Analysis*, v. 1, n. 6, p. 11-19, 1979.

FIGLIO, D. Functional form and the estimated effects of school resources. *Economics of Education Review*, v. 18, n. 1, p. 241-252, Apr. 1999.

GARY-BOBO, R. J.; MAHJOUB, M. B. Estimation of class-size effects, using 'Maimonides' Rule': the case of french junior high schools. London: Centre for Economic Policy Research, 2006. (CEPR Discussion Papers, 5754).

GLEWWE, P. Schools and skills in developing countries: education policies and socioeconomic outcomes. *Journal of Economic Literature*, v. 40, n. 2, p. 436-482, Jun. 2002

GOLDSTEIN, H. *Multilevel statistical models*. 2nd ed. London : E. Arnold ; New York : Oxford University Press, 1995.

HECKMAN, J. J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator. *Review of Economic Studies*, v. 65, n. 2, p. 261-294, Apr. 1998.

HOXBY, C. The effects of class size on student achievement: new evidence from population variation. *The Quarterly Journal of Economics*, v. 115, n. 4, p. 1239-1285, Nov. 2000.

IMBENS, G. W.; ANGRIST, J. D. Identification and estimation of local average treatment effects. *Econometrica,* v. 62, n. 2, p. 467-475, Marc. 1994.

KRUEGER, A. B. Experimental estimates of education production functions. *The Quarterly Journal of Economics,* v. 114, n. 2, p. 497-532, May 1999.

LEUVEN, E.; OOSTERBEEK, H.; RONNING, M. *Maimonides goes to Norway*: class size in primary and lower secondary schools and student achievement. In: SECOND NETWORK WORKSHOP OF THE RTN 'ECONOMICS OF EDUCATION

AND EDUCATION POLICY IN EUROPE'. Uppsala, Sweden: CEPR/IFAU/ Uppsala Universitet, 2005.

LEVACIC, R. *The resourcing puzzle:* the difficulties of establishing causal links between resourcing and student outcomes. London: Institute of Education University, 2005.

______. et al. *Estimating the relationship between school resources and pupil attainment at key stage 3*. London: Institute of Education University, 2005.

LOCKHEED, M.; HARRIS, A. M. Beneath education production functions: the case of primary education in Jamaica. *Peabody Journal of Education,* v. 80, n. 1, p. 6-28, 2005.

NASCIMENTO, P. A. M. M. School resources and student achievement: Worldwide findings and methodological issues. *Educate~ Special Issue*, p. 19-30, Mar. 2008.

PLEWIS, I.; MASON, P. What works and why: combining quantitative and qualitative approaches in large-scale evaluations. *International Journal of Social Research Methodology*, v. 8, n. 3, p. 185-194, Jul, 2005.

RIVKIN, S.; HANUSHEK, E. A.; KAIN, J. F. Teachers, schools, and academic achievement. *Econometrica*, v. 73, n. 2, p. 417-458, Mar. 2005.

SALOMON, G. Transcending the qualitative-quantitative debate: the analytic and systemic approaches to educational research. *Educational Researcher*, v. 20, n. 6, p. 10-18, 1991.

SNIJDERS, T. A. B.; BOSKER, R. J. *Multilevel analysis:* an introduction to basic and advanced multilevel modeling. London ; Thousand Oaks, Calif. : Sage Publications, 1999.

STEELE, F.;VIGNOLES, A.; JENKINS, A. The effect of school resources on pupil attainment: A multilevel simultaneous equation modelling approach. *Journal of Royal Statistical Society - Series A*, v. 170, n. 3, p. 801-824, 2007.

TODD, P. E.; WOLPIN, K. I. On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, v. 113, n. 485, p. F3-F33, 2003.

UNNEVER, J.; KERCKHOFF, A.; ROBINSON, T. District variations in the educational resources and student outcomes. *Economics of Education Review*, v. 19, n. 3, p. 245-259, 2000.

URQUIOLA, M. Identifying class size effects in developing countries: evidence from rural Bolivia. *Review of Economics and Statistics*, v. 88, n. 1, p. 171-177, May 2006.

VIGNOLES, A. et al. *The relationship between resource allocation and pupil attainment: a review* (Research report 228). London: Centre for the Economics of Economics Education, 2000.

WEBBINK, D. Causal effects in education. *Journal of Economic Surveys,* v. 19, n. 4, p. 535-560, 2005.

WOOLDRIDGE, J. M. *Introductory econometrics:* a modern approach. 2nd ed. Cincinnati, OH : South-Western College, 2002.