



Federal Reserve Bank of Chicago

**Causality, Causality, Causality:
The View of Education Inputs and
Outputs from Economics**

Lisa Barrow and Cecilia Elena Rouse

WP 2005-15

**Causality, Causality, Causality:
The View of Education Inputs and Outputs from Economics**

Lisa Barrow
Federal Reserve Bank of Chicago

Cecilia Elena Rouse
Princeton University and NBER

November 1, 2005

Prepared for the Consortium for Policy Research in Education, State of Education Policy Research Meeting, February 14-15, 2005. We thank Brian Jacob, Jesse Rothstein and Diane Whitmore Schanzenbach for helpful conversations, Helen Ladd and the editors for insightful comments and Kyung-Hong Park for research assistance. Any errors in fact or interpretation are ours. The opinions in this paper do not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

I. Introduction

Frustrated with decades of research on education that seemingly amounts to little accumulated knowledge on how to improve student academic outcomes, policymakers and researchers are taking stock of what we do and do not know about the effectiveness of educational inputs.¹ As an example, in 2002 the U.S. Department of Education created the “What Works Clearinghouse” (WWC)—a data base meant to provide educators and policymakers with a “trusted source” of information on what “scientifically-based” education research has to say about what works and does not work in education. The fact that the federal government was willing to spend \$18.5 million (U.S. Department of Education 2002) to fund such an enterprise reflects the view of many that ultimately we know little about which inputs matter for student success in education.

Why do we seem to know so little? Many economists would argue it is because research has not emphasized isolating causal relationships between education inputs and student outcomes (Angrist 2004). Rather, education research has focused on other aspects of the issue, such as differences across settings, which usually has not been the major concern for researchers placing a priority on causality.² If one believes student outcomes are uniquely tied to the educational setting, then it is fruitless to try to draw general conclusions about the “average

¹ We emphasize that “inputs” can be interpreted either narrowly or broadly. District organization (e.g., primary, middle, high school vs. K-8 and high school) can be interpreted as an input as can the structure of teacher contracts. Similarly inputs may be defined as class size, text books, or computers in the class room. We take the broader interpretation but will only discuss the evidence regarding a few of the thousands of potential inputs to educational outcomes.

² Of course all researchers attempt to estimate the causal (or unbiased) effect of an input on educational outcomes. However, all research necessarily demands sacrifice and some researchers will sacrifice causality rather than not estimate differences across settings; others would make the opposite decision. We attempt to draw a distinction between these emphases.

effect” of an education input (which, in this view, does not apply to anyone in reality) (Cook 2001).³ However, the WWC as well as many others in the education research field have started to highlight research that emphasizes isolating the causal relationships between education inputs and student outcomes. Some refer to this emphasis as on “identification” (i.e., “identifying” the impact of a particular input as distinct from other factors) or “internal validity” as termed by Campbell and Stanley (1963).

In this paper we discuss methodologies for estimating the causal effect of resources on education outcomes; we also review what we believe to be the best evidence from economics on a few important inputs: spending, class size, teacher quality, the length of the school year, and technology. In general we conclude that while the number of papers using credible strategies is thin⁴, there is certainly evidence that what schools do matters. But, many unanswered questions remain.

II. The Theoretical and Empirical Ideal

A. Economists’ View of Education Resources⁵

Economists typically analyze a school’s performance and the effectiveness of its inputs using an “education production function.” The school produces education using inputs and a

³ See Cook (2001) for a thoughtful discussion of why education research has by-and-large rejected randomized experiments.

⁴ For example, in a recent review of curriculum-based interventions to improve middle school math achievement, the WWC staff found 77 studies. Of these 77 studies only 10 (studying 5 interventions) were found to have met the WWC standards for evidence which place an emphasis on causal inference (or internal validity).

⁵ Parts of this section are drawn from Rouse (2005).

production technology. One can then measure the effect of particular inputs on the output (education), usually for each student. Specifically, one can think of a production function as:

$$E_{ist} = f(NS_{it}, R_{ist}, X_{ist}, e_{ist}) \quad (1)$$

where E_{ist} represents the output for student i in school s in year t ; NS_{it} represents non-school inputs into student i 's educational attainment, such as her natural "ability," the extracurricular inputs provided by her parents (e.g., music lessons, extra tutoring in subjects), parental inputs (e.g., reading to their children, doing "educationally-rich" activities at home), and her educational history (that is, her achievement level in 4th grade is not only a function of her current school, but also of her schooling in kindergarten, 1st, 2nd, and 3rd grade)⁶; R_{ist} represents the resources under the control of school s in year t (e.g., class sizes, quality of teaching staff, and the curriculum); X_{ist} represents the school inputs that are not typically under the control of public schools (e.g., the quality of a student's "peers"), and e_{ist} is an error term that represents all of the other "stuff" that is not otherwise represented (e.g., measurement error).⁷

⁶ Why do we categorize a student's educational history as a "non-school" input? Because we are attempting to distinguish between *contemporaneous* inputs under the control of the current school and inputs over which the current school has no control. Obviously, a school cannot change what happened to a student in the past.

⁷ Other researchers in the social sciences often represent this educational production function using hierarchical linear models (HLM) in which they explicitly account for more of the organizational structure, i.e., schools are made up of classes which are made up of individual students (Bryk and Raudenbush 1992). We address issues of identifying causal relationships in the framework commonly used in economics rather than in HLM; however, the issues of identification discussed below are also relevant to identification in the HLM framework. While HLM models allow for more nuanced and structured modeling of the parameters and error term, whether the coefficient estimates are unbiased continues to rest on whether the covariates included in the regression fully account for all confounding factors that might affect student achievement and are correlated with the school input in question. In this regard, HLM is similar to OLS estimation discussed below.

The function, f , represents the “production function” or the educational practices that transform the inputs into what a student actually learns. The formulation of the education production function depicted in equation (1) also highlights some of the issues that complicate the design of methodologies for estimating the effectiveness of specific school resources—few of the measures that one would ideally include are observable. Take, for example, educational output, E_{ist} . We rely on our schools to help children learn academic subjects as well as to help them become full-functioning, happy adults by teaching democratic values, responsibility, cooperation, consideration, and other aspects of working well with others.⁸ As such typical outcomes (such as test scores or labor market wages) clearly reflect only part of what we expect from schools. Further, standardized tests do not fully reflect the academic achievement of students. They typically focus on only a few subjects, and in order to keep the testing affordable and not-too-intrusive, are relatively short and mostly rely on multiple choice questions (which are less costly to score). Thus, tests generally provide incomplete and noisy measures of educational output which make it much harder to detect the effectiveness of inputs.⁹

Because of non-school factors and other inputs beyond the school’s control, one cannot easily generate a causal estimate of the effect of school quality on outcomes.¹⁰ For example, the

⁸ That said, in this paper we often refer to the educational output as “student achievement” for ease of exposition.

⁹ More formally, a noisy measure of educational output (the dependent variable in a regression model) will increase the residual variance which will increase the size of the standard errors. As such, one will be less likely to reject the null hypothesis that an input has no effect on student outcomes.

¹⁰ In this regard, while economists refer to equation (1) as a production function, in many respects it is not. In order to truly recover the parameters of a production function one would need to hold *everything* else constant. Thus, if one were studying the effect of lowering class size on a student’s achievement, one would require that all other educational inputs, such as

test scores of more disadvantaged students (in School A) will likely be lower than the test scores of more advantaged students (in School B). If the quality of school resources in each school is correlated with the socioeconomic status of the students, it will be difficult to disentangle the role of school inputs from the influences of non-school factors. Suppose School B has more qualified teachers and more computers in the classroom than does School A. To study the effect of teacher quality and computers on school outputs, one must develop an analytical strategy that adequately controls for the non-school factors. In many cases we suspect the school serving more advantaged students will also have higher quality school inputs. Since (in this example) more computers and more qualified teachers are positively correlated with student family background, education production function estimates will overstate the effectiveness of school resources if one does not adequately control for family background.¹¹

Why is identification so important? From a policy perspective, if one implements a program based on estimates of school effectiveness that are overstated (understated), then the benefits to society will be smaller (larger) than anticipated by the research. If the misstatement (bias) is small, this is not a big problem, however in many cases the bias could be quite large

teaching styles, curriculum, extracurricular activities, and non-school factors remain constant. Because there are no data that allow the researcher to control for all such inputs, the literature typically asks a more general question: what is the effect of an exogenous decrease in, say, class size, on student achievement not requiring that all other inputs remain constant? In this example, teachers may change their teaching style in response to a smaller class size or parents may ease up on complementary educational activities such as tutoring (believing their child is receiving a higher quality education while in school). This more systemic response is what one might expect from an exogenous change in educational policy. See Todd and Wolpin (2003) for a more in-depth discussion of this conceptual issue.

¹¹ In some cases the bias may be negative. For example, special needs and English language learner classes tend to be much smaller than classes for regular or gifted students (Boozer and Rouse 2001). In this case, one may erroneously conclude that smaller class sizes lead to *worse* student outcomes.

leading to no societal benefits, or worse, adverse outcomes. (Or, in the case of understatement, a potentially beneficial program may not be adopted.)

B. *The Ideal Way to Measure the Impact of Schooling Inputs*

To identify the causal impact of school resources, ideally one would begin with a group of students and educate them during the year with the first educational input in question (or the status quo). At the end of the year one would assess the students or administer an appropriate test, the results of which would perfectly reflect what the students know.¹²

Next, one would take the same group of students and revert them back to their initial conditions at the beginning of the *first* year. That is, they would be the same age, have the same living conditions, etc. This second year, one would then educate the students with the second educational input in question. (For example, in the first year the input might be teachers with regular teaching credentials and in the second it might be teachers who have gone through an alternative certification program.) At the end of the year, one would again assess what each student knows.¹³ The difference between the students' outcomes using the first input and those using the second would isolate the value of the second input relative to the first.

Why is this the ideal design for measuring the value of an input? First, because the same

¹² Note that while we describe the ideal outcome as a test, in theory one could use any other outcome (such as adult wages or voting behavior).

¹³ In the ideal methodology one need not administer a test at the beginning of each school year because the students (and all of their characteristics) are identical in both years. If one were to administer a test at the beginning of the year, the difference between what the students know at the beginning of the year and the end of the year could (mostly) be attributed to the input used that year since the students are the same in each testing period. This would constitute the input's "value-added" in the each year.

students are educated using each educational input starting from the same initial conditions, one has guaranteed that all background characteristics of the students are the same, including their prior exposure to high and low quality schooling, their family situation, and their innate ability. In other words, one has effectively controlled for X_{ist} and NS_{it} . Second, because the assessments perfectly reflect what students know, there is no measurement error. The combination of these two features means that one can isolate the relative influence of the educational inputs. The key is that by observing how students fare under both regimes one has a “counterfactual” outcome against which to compare the outcome using the main input of interest. Namely, we observe how much the students learn using the first input as well as how much they would have learned had we instead used the second input. As we will discuss, it turns out that establishing a credible counterfactual outcome is among the most difficult tasks faced by the analyst.

Clearly, the ideal evaluation is impossible to implement. No one can turn back time to assess the students under the exact same conditions in each year. In addition, there has not been an assessment devised that *perfectly* reflects what students know. Rather, existing tests reflect only a part of what students know, and there are permanent confounding factors (such as different test-taking abilities) and random confounding factors (such as some students not feeling well on the day of the test or not getting enough sleep before the test). The analyst’s task is, nevertheless, to implement a methodology that comes as close to the ideal approach as possible.

One must also consider how generalizable the outcome of any evaluation would be (that is, whether the evaluation has “external validity” (Campbell and Stanley 1963). If we started with a “representative” group of students, then the ideal evaluation would uncover the average relative effectiveness of the input even if the effectiveness of the input varies over the

population. If it is the case that the first or second input is more effective for some students than others, the students in our study must be representative of the population of students in order to assess the effectiveness of each input on average. Consider the extreme example in which the first input is effective for teaching girls but not effective for teaching boys, and the second input is as effective for teaching boys as the first input is for teaching girls but not at all effective for teaching girls. Further, assume that 50 percent of the student population is female and 50 percent is male. In this extreme example, the two inputs are equally effective on average. However, if the students in the study were disproportionately female, we would incorrectly conclude that, on average, the first input is more effective than the second input. Thus, only by starting with a group of students who are representative of the population (in general, or the population of interest for the policy) can one guarantee that the exercise will uncover the true average treatment effect of the input. Of course, in this extreme but simple example with an ideal set up in which one knows all of the characteristics of the students, one could estimate the effectiveness of the inputs for different subgroups of the population and discover, in fact, that the first input was more effective for girls and the second input was more effective for boys.

This issue of “heterogenous treatment effects” is part of the reason many in education cast a jaundiced eye toward randomized experiments (and many other quantitative methodologies). However, the average effect (even if there are different effects for different subpopulations or in different settings) is important for setting policy, especially if it is difficult for policymakers to target a policy narrowly or effectively. As such, the first-order question is whether an intervention works in general or for very broad categories of schools or students. Among the important subsequent questions is whether it is more effective for some groups or

situations than for others.

In a related manner, one characteristic of many methodologies that emphasize causality is that the researcher does not delve into the intricacies of why the intervention may have mattered (or not mattered). For example, in the studies of class size reduction using the randomized Project STAR data from Tennessee, researchers have not identified *why* class size reduction mattered. Was it because there were fewer students in the class *per se* (i.e., a peer effect story) or because the teachers changed their teaching styles? It is important to note that this is not an inherent limitation of the randomization (or other methodologies which emphasize causality). Rather it follows from the researcher placing a greater emphasis on causality such that he or she will not attempt to address issues of which subcomponents of an intervention might have mattered, unless there was randomization along those dimensions as well. That is, unless the researcher randomly assigned teachers to teaching styles in addition to differing class sizes, he or she will worry that teachers who adopted certain styles may be different from those who did not in ways that are not observable. Using survey data (or other observational data) and applying ordinary least squares regression will not solve this problem (unless, of course, they contain *all* of the relevant background variables).

In the next two sections we review methodologies researchers have used in their quest to study the effectiveness of school inputs.

III. Methods Using Observational Data

Observational data are gathered from observing existing situations in schools. That is, they contain the existing input levels in schools (e.g., class sizes and teacher qualifications) as

well as information on students attending the schools. These are “observational” because there is no attempt on the part of the analyst to manipulate the situation generating the data. Most of the literature on the effect of educational inputs on student outcomes relies on observational data typically because they are most readily available. However, the fundamental problem with observational data is that individuals and schools choose their situations, such that one must control for all factors that led the individual or school to their choice that might also be correlated with the outcome of interest. Each of the approaches discussed, below, attempts to address this fundamental problem in a different way.

A. *Ordinary Least Squares Regression*

Traditionally, researchers have used ordinary least squares regression (OLS) to study the impact of school resources on outcomes (e.g., Coleman et al 1966). In the cross-sectional case, the analyst relies on a specification such as,

$$E_{ist} = \alpha + \beta R_{ist} + \lambda NS_{it} + \delta X_{ist} + e_{ist} \quad (2)$$

where the variables are the same as those in equation (1) and α , β , λ , and δ are parameters to be estimated. If all of the other factors (NS_{it} and X_{ist}) are observed in the data set such that one can include them in the regression (thereby holding them constant), then one can generate an unbiased (causal) estimate of β . However, we know of no data that contain all of the other factors for which one must control. Rather, most cross-sectional data contain only limited information on important non-school and school factors. For example, in school administrative data one rarely, if ever, has an accurate measure of family income. As a result, researchers control for whether the student was eligible for the National School Lunch Program, a proxy for

income that is very crude at best.

Today many researchers believe that cross-sectional OLS estimates are likely inaccurate (i.e., they are statistically biased). As a result, they turn to data that follow a student over time—longitudinal data. With longitudinal data one can control for observed and unobserved student characteristics, particularly those that do not change over time. In addition, because these data have information on students over multiple years, one comes closer to the comprehensive data required in equation (1). Because of accountability requirements in *The No Child Left Behind Act of 2001*, many states are beginning to collect such data on individual students statewide. This advance in data collection will be an invaluable resource for education researchers going forward.¹⁴

Since these data have not been readily available, one approach that researchers have used is known as a “value-added” specification (e.g., Summers and Wolfe 1977). These equations take the form:

$$E_{ist} - E_{ist-1} = \alpha' + \beta'R_{ist} + \lambda'NS_{it} + \delta'X_{ist} + e'_{ist} \quad (3)$$

where E_{ist-1} is the student’s outcome in the previous year. In this case, one estimates the effect of a (concurrent) resource R_{ist} on the change in a student’s outcome. If E_{ist-1} fully captures the effect of all previous schooling and non-schooling inputs on the student’s achievement, then one can generate an unbiased estimate of β' , the effectiveness of the school input in question. However, it seems unlikely that a noisy measure of a student’s performance in the prior year (as reflected

¹⁴ Texas, North Carolina, and Florida already have fairly rich databases and have provided researchers with access to them. That said, we know of no administrative data that contain all of the information that would be relevant for replicating the ideal research design.

in test scores) will fully control for all relevant factors.¹⁵

In general, the basic problem with using OLS regression to estimate the effect of school resources on student achievement—using either cross-sectional or longitudinal data—is that one is uncertain whether or not one has controlled for all important factors¹⁶ in the regression. As such, much of the latest and most compelling research on the impact of school resources on student achievement has moved away from simple OLS regression.

C. *Regression Discontinuity*

Imagine that an educational input is assigned to students based on the value of some measure. For example, suppose that a state imposes a maximum class size of 25 students per teacher. If the number of students exceeds 25 students, then the students are to have a teacher plus a teacher's aide. If there are more than 40 students then the school must create two classes (each with one teacher). Because of the cutoffs imposed by the law, students in schools with 39 students in, say, the 3rd grade will experience a much larger class size (39 students) than students in schools with 41 students in the 3rd grade who will be educated in class sizes of 20 and 21. The key is that the variation in whether a school has 39 students in the 3rd grade or 41 students likely occurs by chance. More specifically, it is unlikely there are other factors that determine whether there are 39 or 41 students in the 3rd grade that also affect student outcomes. As such, one can compare the outcomes of students in schools with 39 students to those of students in schools

¹⁵ See Todd and Wolpin (2003) for a more comprehensive discussion of what different empirical models using longitudinal data identify and under what assumptions.

¹⁶ Statistically “important factors” are those that influence the student’s performance on the outcome measure and that are correlated with the resource in question.

with 41 students and attribute any difference to the effect of class size.

This basic methodology is known as a “regression discontinuity” design (Cook and Campbell 1979), and it has grown in popularity in research on education quality. More generally this design will work when the input in question (in this example, class size) is at least partly determined by a known discontinuous function¹⁷ of an observed characteristic (in this example, 3rd grade enrollment). Because of the discontinuous relationship between the input in question and the observed characteristic, the researcher can control directly for the observed characteristic while still identifying the effect of the input in question on student outcomes making the strategy much more compelling than typical OLS.¹⁸

An important disadvantage of regression discontinuity designs is that the range of values over which one gets identifying variation tends to be rather small. (For example, the most compelling comparison in the class size example is between schools with enrollments of 39 vs. 41; one can imagine that schools with 15 rather than 35 3rd-grade students also vary along other dimensions that may or may not be observable.) In addition, if the effect of class size on student

¹⁷ That is, class sizes and enrollment do not simply increase one-for-one forever, but there is a change in the relationship at some point. In this example, class sizes increase one-for-one until there are 40 students in the class and then class sizes abruptly (and discontinuously) decrease to 20 and 21 students.

¹⁸ In order for regression discontinuity design methods to provide credible estimates of the effect of educational inputs on student outcomes, the key individuals involved (e.g., parents, principals, teachers, students) must not have control over the exact value of the measure on which eligibility for the input will be based. Thus, these key individuals must not have control over the exact size of the 3rd grade class in our example. If the underlying measure can be manipulated, then one could manipulate the school enrollment to engineer the desired class size. If the desired class size is correlated with other unobserved determinants of student outcomes, such as commitment to education, then the estimate of the effectiveness of class size will be biased. In this example, the methodology seems more credible when applied to public schools—that do not have complete control over their enrollment—than to private schools.

achievement in the range for which one can generate unbiased estimates is different from that in other ranges, then the estimated parameters may not generalize.¹⁹ However, because regression discontinuity provides a credible way to estimate a parameter with internal validity, it provides an invaluable tool for education research. Further, while it may not appear to be very practical, regression discontinuity is a candidate analytical design whenever there are cutoffs for program participation. In Section V, below, we discuss papers that use this design to study the effectiveness of professional development, smaller class sizes, and summer school and grade retention.

D. *Natural Experiments (Instrumental Variables)*

“Natural experiments” provide another approach for analyzing observational data in a way that comes closer to the ideal experiment than OLS. In this approach, researchers attempt to locate determinants of schooling inputs that would not be expected to independently alter their educational outcomes.

Here’s the basic idea used in this methodology. Suppose we were interested in studying the effect of financial resources on student outcomes. And, we knew of a determinant of financial resources, say, a change in the state education financing formula, that would increase the amount of money allocated to one group of schools. Suppose further we were certain that this change in the financing formula did not have any direct effect on the students’ outcomes,

¹⁹ For example, when class sizes vary by only 1 or 2 students—which may be the viable range for policy changes—teachers may not change their teaching styles significantly. However, when class sizes change a lot (e.g., 39 students vs. 20 students) then many other educational practices may change making it difficult to isolate the effect of class size, per se. Or, the effect of class size reduction may matter for, say, classes with over 30 students but not for classes with 20-25 students.

except through the impact on the schools' revenues.²⁰ We would then estimate the effect of state aid on outcomes in two steps: In the first step we would estimate the effect of the state aid on school revenue. In the next step we would measure the effect of the change in state aid on students' outcomes. If we found that the outcomes of the students improved, then we could be sure that increased revenues were the cause of the outcome improvement since we were certain that the change in state aid had no *direct* effect on outcomes. The ratio of the outcome improvement caused by the change in state aid to the change in the educational input caused by the state aid is a straightforward estimate of the causal effect of financial resources on student achievement. This instrumental variables (IV) estimator uses the "exogenous" event (a change in the state financing formula) as the instrumental variable.²¹ This is, indeed, the approach taken in the recent paper by Jonathan Guryan (2003) to study the impact of "money" on student outcomes (see section V, below).

As another example of how this estimation strategy works, consider the recent paper by

²⁰ Thus, for example, one would need to be careful with state aid formulas that were designed to be redistributive. In this case, schools in poor areas would likely receive more state aid and yet their students would likely perform worse on tests than students in wealthier areas. As such, one would not want to simply use the *level* of state aid as the outside determinant of resources. However, some *changes* in the formula may have been driven by factors that are uncorrelated with the characteristics of the districts such that these *changes* would be valid outside determinants of state aid.

²¹ Some investigators refer to the fact that schools use varying levels of a particular input as a "natural experiment." (For example, we have heard researchers propose studying the effect of a whole school reform model by "exploiting the natural variation" arising from the fact that some schools have adopted the model and others have not.) This is not what most economists would refer to as a "natural experiment" particularly since the method of analysis that follows is simply OLS (in which one relates whether or not a school uses the whole school reform model in question to student outcomes). One is still left with the question as to why some schools adopted the model and others did not, and whether this same (unobserved) factor that led them to adopt it is correlated with student outcomes.

Angrist and Lavy (2002) that studies the effect of technology on student achievement. Angrist and Lavy note that schools in Israel that received a technology grant were more likely to use computer-assisted instruction (CAI) (again, see section V, below). The grant program is a suitable instrumental variable so long as one can assume that any difference between schools that received funding through the program and those that did not is only the use of computers in the schools and not other observable characteristics of the schools (e.g., schools with more motivated principals were more likely to apply for the grant program and have better performing schools). Angrist and Lavy study the correlation of participation in the program with other school environment characteristics (e.g., class size, hours of instruction, non-computer technology) and conclude that the program increased CAI instruction without changing other characteristics of the schools. That said, this highlights the main empirical challenge in IV strategies: the researcher must make the claim that the instrumental variable only affects the outcome through its effect on the educational input (the endogenous variable) in question. As such, the researcher must make assumptions about unobservable factors—which are inherently difficult to prove or disprove.²²

Another disadvantage of IV strategies is that like regression discontinuity designs, if there is heterogeneity in the effect of an input on student outcomes, the estimated effect may not generalize to other segments of the population. The reason is that IV identifies the effect of an input on student achievement among those students (or schools) that are induced to change their behavior because of the instrumental variable (Imbens and Angrist 1994). Thus, in the case of

²² In this regard, IV shares much in common with OLS. Both strategies rely on assuming that the error term is not correlated with either the input in question (in the case of OLS) or the instrumental variable (in the case of IV), conditional on the (other) observed covariates.

the technology grant program in Israel, IV identifies an effect of CAI only for those schools that increased their intensity of CAI because of the grant program (there are others that may have received money from the program that would have increased their CAI intensity even without the program). If student achievement is particularly responsive to increases in CAI intensity in these schools, then the IV estimate of the effectiveness of technology will overstate the average effect across all schools. While quite popular, especially among economists, it is extremely difficult to find credible instrumental variables such that this methodology is unlikely to become a mainstay of education research.

IV. Methods Using Experimental Data

Finally we come to what many describe as the “gold standard” in evaluation methodology: randomized designs. In this case, one group of students is randomly assigned to be educated with the input in question (the treatment group) and a second group of students is (randomly) assigned to be educated with the status quo or another input (the control group). One then tests the students at the end of the evaluation. The difference in student outcomes between those in the treatment group and those in the control group represents the effect of the input in question relative to an alternative. Why is this the ideal design for measuring the impact of school inputs on student outcomes? First, the random assignment of students to an educational input initially controls for all background characteristics of students, including their prior exposure to high and low quality schooling. That is, on average, both groups would have the same distribution of students along observable and unobservable dimensions such that one need

not control for X_{ist} and NS_{it} .²³ Further, because a random event determined to which group (treatment or control) a student was assigned, one need not be concerned that students who expected to benefit more from an input were those likely to be educated using it (a concern under OLS), and the error term is uncorrelated with the input in question. Randomization makes the identification of causality in experiments more transparent than many other methodologies; thus, experiments are quite compelling.

That said, experiments are not well-suited to answering all questions. First, the more aggregated the unit observation, the more “cumbersome” a randomized evaluation becomes. For example, to study the effect of district-wide open enrollment on student outcomes using an experimental design one must randomly assign *districts* to treatment and control groups. While there are more than 14,000 districts in the U.S. from which to choose, the logistics of getting a sufficient number of districts to cooperate, such that one would have a large enough sample with which to draw conclusions, would be quite daunting.

In addition, in an “ideal” experiment the random assignment process completely determines the status of individuals in the treatment and control groups. In many experimental settings involving human subjects, however, there is slippage between the random assignment status of experimental subjects and whether or not they actually receive the treatment.²⁴ For

²³ While researchers will often also attempt to administer a pre-test at the beginning of a randomized evaluation it is not necessary to do so to generate an unbiased estimate of the effect of the input in question on student outcomes, if the randomization is conducted correctly. The reason is that the characteristics (both observed and unobserved) of the students in treatment and control groups are the same, on average, at the beginning of the evaluation. Hence, controlling for the student’s pre-test will not change the estimate of the effectiveness of the input in question. Controlling for the pre-test can, at times, improve statistical precision (i.e., lower the standard errors).

²⁴ This problem is exacerbated in multi-year studies.

example, in the Tennessee Project STAR experiment some students randomly assigned to small classes ended up in larger classes and vice versa. In a randomized experiment, if one simply compares the outcomes of students *originally assigned* to the treatment group to those originally assigned to the control group (regardless of which treatment the student actually received), one estimates the “intention-to-treat” effect (Rubin 1974; Efron and Feldman 1991). While of interest, the intent-to-treat effect may be unsatisfying for those educators and policymakers who desire an estimate of the effects of a particular input for students who are educated actually using the input—the effect of “treatment on the treated.” The estimated intent-to-treat effect does not establish whether the input in question—properly implemented—is better than an alternative. Note that the difference between the intent to treat and the treatment on the treated is “take up” (or implementation). Students randomly assigned to the treatment group may actually choose to use another input (that is not the input in question), and students randomly assigned to the control group may actually use the input in the question (i.e., get “treated”).

While the effect of treatment on the treated is important, we believe there are at least two reasons why we should also be interested in intention-to-treat. First, it is the only policy instrument available to policy makers. If the state of Tennessee decides to lower class sizes for all students, all policymakers in Tennessee can do is mandate lower class sizes. If all schools comply with the new law, then one might expect results that follow from the treatment on the treated parameter. In reality, some schools may not be able to meet the new lower class size requirements. If so, any anticipated gains in student achievement will be diluted. Take the extreme case in which no schools are able to reduce class sizes (say, for example, because of lack of building capacity). In this case, even if student achievement is much higher with smaller class sizes, there will be no achievement gains from the program because the program was not

implemented. When considering a policy change, policymakers must consider both halves of the issue: both how the treatment affects the treated and whether the program can and will be implemented as desired.²⁵ Because it combines both halves of the issue, the intention-to-treat estimates reflect the overall potential gains from an educational policy change. Second, as in many experimental settings, the randomization only occurred in the intention-to-treat, and, as such, this estimate is the only unambiguously unbiased estimate that one can obtain from an OLS regression, assuming the initial selection was truly random. That said, one can estimate the effect of treatment on the treated by using an IV strategy in which one uses the random assignment as an instrumental variable for whether or not the student received the treatment (the input in question). This is a case in which, if the initial selection was truly random, the instrumental variable will not be correlated with the error term in the outcome equation and therefore will be valid.

Although experimental approaches probably come closest to the ideal evaluation design, they do have some analytical shortcomings which are worth highlighting. For example, they tend to be rather “blunt” instruments. One implements an experimental design out of concerns for obtaining an unbiased estimate of the effect of the input on student achievement. However, one can only truly get such an unbiased estimate from the point of random assignment, and unless the experimental design is sufficiently complicated, one can really only answer one question (whether those assigned to the treatment group fared differently from those assigned to the control group). Cost concerns (and complicated implementation) often preclude comparing

²⁵ This discussion is analogous to the distinction between “efficacy” (i.e., how well a drug might work in theory) and “effectiveness” (i.e., how well a drug works in practice given the fact that patients may not follow the protocol exactly).

the effectiveness of more than two or three different inputs in one study. Note, however, that one can ask whether the effect differs for subgroups identified based on characteristics measured before random assignment takes place, provided sample sizes are large enough. Similarly, it is rare that researchers using experimental designs attempt to determine (at least causally) which dimensions of an intervention may have worked or not worked. Again, the reason is that unless random assignment also occurred along the “subdimensions” any such analysis will not necessarily yield causal estimates of the effectiveness of the subdimensions.

Keeping in mind these many empirical challenges, we now turn to a brief summary of what these methodologies suggest about the importance of schooling inputs on student outcomes.

V. Does School Quality Matter?

A. Coleman, Hanushek, and Card and Krueger

The Coleman Report (Coleman et al. 1966) is credited with launching an explosion of studies estimating the relationships between educational outcomes and school inputs. Many papers were written criticizing the methodology used in the Coleman Report including arguments that longitudinal studies or well-designed experiments were needed to make causal inferences (e.g., Sewell 1967). Further, even the Report’s authors note that their cross-sectional analysis does not provide a strong basis for causal interpretation. However, the Report was broadly interpreted to find that schools do not matter; instead, family background and peers explained most of the variation in education outcomes.

By the mid-1980s, Hanushek (1986) includes 147 studies in a survey of the literature relating educational outcomes to school inputs. Ten years later, Hanushek (1996) finds more than double the number of studies to survey. The reviews and conclusions of Hanushek’s

analyses reinforced the findings of the Coleman Report, and by the early 1990s, many people were firmly convinced that “money does not matter,” namely, that once family inputs into schooling were taken into account, school resources did not matter. As Hanushek (1997) writes, “Simple resource policies hold little hope for improving student outcomes.” He further concludes, “Three decades of intensive research leave a clear picture that school resource variations are not closely related to variations in student outcomes and, by implication, that aggressive spending programs are unlikely to be good investment programs unless coupled with other fundamental reforms.” (Hanushek 1996)

Although Hanushek’s meta-analyses have been extremely influential, researchers have criticized them along a number of dimensions. Hedges, Lain, and Greenwald (1994) note that many of the studies Hanushek surveys can be faulted for methodological reasons similar to those discussed above. For example, many of the surveyed studies are based on cross-sectional, observational data and do not have longitudinal data on student outcomes or natural experiment features. Further, Hanushek relies on simple “vote counting” in his analysis. Using more sophisticated meta-analytical techniques, Hedges, Lain, and Greenwald conclude that among the studies surveyed in Hanushek (1989), per pupil expenditures, teacher experience, and teacher-pupil ratios are positively related to student outcomes. They also find that the effect sizes for per pupil expenditures are large and educationally important.

Krueger (2003) makes a more basic point in criticizing Hanushek for weighting all estimates equally and thus giving more weight to studies that publish more estimates. Focusing on the class size results included in Hanushek (1996), Krueger uses alternative weighting strategies, including giving equal weight to each study rather than equal weight to each estimate, and finds support for a positive relationship between smaller class sizes and better student

outcomes.

Today, while researchers recognize the importance of family background and other non-school inputs in determining educational outcomes, many have come to question the findings of the Hanushek meta-analyses as well as the validity of many of the individual studies estimating education production functions. Card and Krueger (1992) perhaps marks the turning of the tide on the view that schools do not matter. Instead of focusing on direct education outcomes, Card and Krueger focus on how school quality affects the returns to schooling (i.e., the increase in earnings associated with an additional year of schooling). Assuming that school quality for men working in a given labor market varies exogenously by their state of birth and cohort, Card and Krueger find that men who were educated in states and years with higher quality schools—schools with lower pupil-teacher ratios, longer school years, and higher relative teacher pay—earn more for an additional year of education than men educated in states and years with lower quality schools. These results are consistent with earlier work finding positive relationships between school quality and earnings (e.g., Johnson and Stafford 1973; Rizzuto and Wachtel 1980) and work that attributes much of the closing of the black-white wage gap to improvements in school quality for African American students (e.g., Smith and Welch 1989). While some researchers challenged the assumptions used in Card and Krueger (1992), others began to consider that school resources may affect students' earnings after leaving school without having measurable effects on academic achievement while in school (Burtless 1996).

C. *Recent Studies*

Since Card and Krueger (1992), there have been many new papers examining the effects of school inputs on student achievement, several of which use estimation strategies aimed at

identifying the causal relationships between school inputs and student outcomes. In this section we review a few of the best studies in economics assessing school spending, class size, teacher quality, time in school, and technology.

1. Spending

Because some students may be more expensive to educate than others and schools and districts differ in the types of students they serve, simply looking at the relationship between average student test scores and per pupil spending may indicate that greater school spending is associated with lower student achievement. As a result, researchers rely on alternative strategies for identifying the causal relationship between spending and student outcomes. Barrow and Rouse (2004) and Guryan (2003) use changes in state school financing aid formulas as instrumental variables to isolate plausibly exogenous changes in school spending. Barrow and Rouse (2004) examine the general question of whether spending on schools is valued by the “market” by looking at the effects of increased school spending on local property values. Indeed, the authors find that school spending is valued, on average, since they estimate that property values increased by the expected amount in school districts that received an extra \$1 per pupil in state school financing. If potential residents did not value the additional spending because school districts were viewed as spending excessively or wastefully, additional state aid should not have resulted in such large increases in property values.

Guryan (2003) looks more specifically at the relationship between school spending and student achievement in Massachusetts. He finds that additional state aid resulting from a change in the financing formulas led to a significant increase in math, reading, and science test scores for both 4th and 8th grade students. Specifically, he estimates that a \$1000 increase in per-pupil

spending leads to a one-third to one-half of a standard deviation increase in average test scores. In sum, Barrow and Rouse (2004) and Guryan (2003) both suggest that money matters when it comes to public schools. Below we look at studies that examine more specifically whether different inputs matter.

2. Class size

Although the effect of class size on student achievement has most often been studied using observational data, Boozer and Rouse (2001) provide a clear demonstration of how estimates of class size effects can be misleading due to the relationship between class size and student ability as well as how school-level measures of pupil-teacher ratios can mask significant within-school variation in actual class size. Thus, one should be suspect of estimates that do not make use of more sophisticated estimation techniques to uncover the causal relationship between class size and student achievement. Fortunately, class size is one of the education topics that has been studied using a variety of estimation techniques, including regression discontinuity, instrumental variables, and randomized evaluation.

Angrist and Lavy (1999) use a regression-discontinuity estimator to look at the effect of class size on student test scores in Israel. Public schools in Israel have a maximum class size of 40 pupils which generates a non-linear, non-monotonic relationship between grade enrollment and class size. As discussed above, this will generate large differences in class size between grades with enrollment of 39 students and grades with enrollment of 41 students. For fourth and fifth grade students, Angrist and Lavy (1999) find that reductions in class size increase test scores by statistically significant and educationally important amounts. They do not find similar effects for third grade students.

Two other papers have used regression discontinuity and/or instrumental variables as well. Hoxby (2000) uses class size minimums and maximums in Connecticut to look at the effect on student test scores of changes in class size driven by movements in enrollment populations that push schools over and under the class size thresholds. She finds mixed results on the relationship between class size and student performance. Boozer and Rouse (2001) use state class size maximums as an instrumental variable for student-level class size in the NELS88 and find that smaller classes improve student achievement.

Perhaps the best known and most convincing evidence on the impact of class size comes from the Tennessee Student/Teacher Achievement Ratio experiment (Project STAR) in which Tennessee kindergarten students were randomly assigned to small classes (13 to 17 students per teacher), regularly-sized classes (22 to 25 students per teacher), or regularly-sized classes with a teacher's aid (22 to 25 students per teacher). The experiment continued through the third grade and then students were returned to regularly sized classes. Finn and Achilles (1990) and Krueger (1999) find that students in the smaller classes outperformed students in the larger classes on standardized tests. Additionally in a longer-term follow-up of Project STAR, Krueger and Whitmore (2001) find that students who were randomly assigned to smaller classes were significantly more likely to take a college entrance exam and that this effect was greater for African American students.

At this point, many education researchers and policymakers have been convinced that smaller class sizes can improve student outcomes on average. However, many unanswered questions remain. For example, we need to know more about the cost of class size reduction relative to other interventions and whether it is cost effective. In addition, California's experience with class size reduction in the 1990s highlighted that implementation—especially on

a large scale—can go awry (Bohrnstedt and Stecher 1999). Importantly, even the evidence from Project STAR suggests that the impact of class size reduction differs across schools and subpopulations of students (for example, Krueger (1999) found the largest effects for African American and low-income students). Clearly we need to know more about the conditions under which reducing class sizes will be most fruitful.

3. Teacher quality

The preponderance of evidence suggests that teachers matter for student outcomes. Hanushek, Rivkin, and Kain (2005) use Texas data on elementary students linked to teachers at the school-grade level in order to estimate the effect of teachers on student learning while Aaronson, Barrow, and Sander (2003) use Chicago Public Schools data on high school students linked to teachers at the classroom level to examine teacher quality. Both studies find large variation in teacher quality as measured by the effect of teachers on student test score gains. Hanushek, Rivkin, and Kain (2005) estimate that a one standard deviation increase in teacher quality at the grade level will increase student test scores by roughly 10 percent of a standard deviation while Aaronson, Barrow, and Sander (2003) find that a one standard deviation improvement in 9th grade math teacher quality for one semester is associated with a gain equal to 10 to 20 percent of the average math test score gain experienced in a typical school year.

When it comes to determining what makes a good teacher, the research is much less clear. Research by Clotfelter, Ladd, and Vigdor (2004) in North Carolina illustrates the great tendency for the most qualified teachers to teach in schools with the most advantaged students as well as for parents of more advantaged children to get their children into classes with more qualified teachers. This sorting of teachers and students makes it difficult to disentangle the

causal effects of various measures of teacher quality. In addition, the characteristics of teachers available in the large administrative data sets are typically limited to those that determine compensation, such as whether or not a teacher has a master's degree and how many years she has been teaching in the school district.

Researchers have found some evidence that teacher quality improves sharply after one or two years of experience (e.g., Clotfelter, Ladd, and Vigdor 2004; Hanushek, Rivkin, and Kain 2005). However, new teachers exit teaching at fairly high rates, and Aaronson, Barrow, and Sander (2003) find that teachers in the lowest quality decile in one year are 26 percent less likely to be teaching in the next year than teachers in the highest quality decile suggesting that some of the experience results may be driven by selection if only the higher quality teachers stay beyond one or two years. Aaronson, Barrow, and Sander (2003) also find some evidence that undergraduate major may be related to teacher quality, while Clotfelter, Ladd, and Vigdor (2004) find evidence that teachers who score best on licensing tests are indeed higher quality teachers.

Using a regression discontinuity design, Jacob and Lefgren (2004a) take advantage of the nonlinear relationship between school-level student achievement in Chicago Public Schools and the assignment of schools to probationary status in order to examine the relationship between professional development and student achievement. In 1996, elementary schools in which fewer than 15 percent of students met national norms on a standardized test of reading were placed on probation and given resources (up to \$90,000 in the first year) to purchase staff development services. Schools with more than 15 percent of students meeting national norms were not placed on probation and not given the additional resources. Jacob and Lefgren thus assume that whether a school has just fewer than 15 percent of students that met the reading norm or slightly more than 15 percent of students that met the norm is by chance.

The authors find that schools on probation primarily spent the additional resources on professional and staff development purchased from a wide variety of external sources including universities, nonprofit organizations, and independent consultants. The authors find that teachers report a 25 percent increase in the frequency of attending professional development programs, and others (e.g., Smylie et al. 2001) have reported a more substantial increase in the quality of the professional development teachers received. Unfortunately, however, the authors find no evidence that the increase in the quantity and quality of professional development induced by schools' probationary status translated into improved student achievement.

In sum, the best evidence suggests that teachers matter; however, we still have much to learn about how to identify quality teachers when making hiring decisions or how to increase teacher productivity with training or professional development.

4. Time in school

The length of the school year in the U.S. is a frequent target of criticism in discussions of why students in the U.S. score badly on standardized tests relative to other developed countries. Several studies document erosion of students' skills over the summer vacation (e.g., Cooper et al. 2000), and there is some evidence that summer school can improve student achievement (e.g., Jacob and Lefgren 2004b).

Jacob and Lefgren (2004b) utilize regression discontinuity to look at the effect of summer school and grade retention on student achievement in Chicago. In 1996, Chicago Public Schools instituted a policy of requiring 3rd and 6th grade students to attend summer school if they did not meet minimum test score thresholds. Students were then retained in grade if they did not achieve the minimum test score following summer school. The authors are able to use the discontinuity

of the treatment rule in order to assess the benefits of the summer school and grade retention policy on student achievement. Namely, students scoring just below the minimum “passing” test score and students scoring just above the minimum passing test score are assumed to be quite similar except that those scoring just below the threshold are assigned to summer school. Jacob and Lefgren (2004b) find that the net effect of summer school and grade retention was to increase student achievement among 3rd grade students. However, the authors find no similar achievement gains for 6th grade students.²⁶

Pischke (2003) specifically looks at the effect of school-year length by taking advantage of a natural experiment occurring in West Germany in the late 1960s. Adoption of a common fall-start to the school year led students in most states to experience two short school years, equivalent to roughly two-thirds of the standard length school year. In contrast, students in West-Berlin and Hamburg attended one long school year. Pischke (2003) finds that the shorter school year increased grade repetition among elementary school students, but that the shorter school year had no effect on the number of students attending the highest secondary school track or subsequent earnings as adults. Thus, there is little evidence of long-lasting negative effects of a shorter school year.

The Pischke (2003) results point to an important difficulty in estimating educational policy effects with observational data even in the presence of a natural experiment. Although we may believe that the natural experiment is valid in that it generated exogenous variation in the length of a school year and that it should only affect student outcomes through the experiment’s

²⁶ The authors also use regression discontinuity to look at the effect of grade retention alone using the post summer school test scores. Once again, they find that grade retention is beneficial to 3rd grade student achievement, but has no effect on 6th grade student achievement.

effect on the length of a school year, it is quite likely that teachers changed their behavior to compensate for the temporarily shortened school year. Since the behavioral response to a short-term change in the school year may be different from the responses generated by a permanent change in the length of the school year, the results may lack external validity.

5. Technology

Research on the success of computer-aided instruction (CAI) has yielded mixed evidence at best. Some research using observational data has shown computers can offer highly individualized instruction, allow students to learn at their own paces, enhance assessment, and increase student motivation (e.g., Sandholz et al. 1997; Means and Olson 1995; Lepper 1985). In contrast, other research reports that computers are frequently poorly embraced by teachers, can disrupt classrooms, and fail to increase student achievement in any measurable way (e.g., Cuban 2001; Becker 2000; Angrist and Lavy 2002; Rouse and Krueger 2004).

A common critique of the literature is that both student outcomes and what constitutes “computer use” are poorly defined (Cuban, 2001). For example, while Angrist and Lavy (2002) are able to use an instrumental variables estimator to look at the effect of CAI on student test scores, the intensity of computer use is defined by the teacher’s response to a rather vague question about how often they used “computer software or instructional computer programmes” (Angrist and Lavy 2002). The authors find no evidence that greater use of CAI improved student test scores in math or Hebrew.

Borman and Rachuba (2001) and Rouse and Krueger (2004) have the advantages of being able to evaluate the effect of much more specific computer use—the use of a particular instructional software—and to implement random assignment of students to treatment and

control groups. Both studies evaluate the popular *Fast ForWord* (FFW) computerized reading instruction program using random assignment within schools in large urban school districts. The studies' findings are remarkably similar: both rule out large impacts of computerized instruction with estimated effects that are not statistically different from zero.

While these studies suggest that CAI does not significantly improve student educational outcomes, one might find that different computerized reading programs were successful or that the use of CAI in other subjects significantly raised students' learning in those subjects. Further, one might find that FFW was effective when used in other settings. Both randomized evaluations of FFW were conducted in schools, but it may be that schools are not the best environment in which to implement the program (FFW is also often used by psychologists and reading specialists in private practice). While the schools and teachers in the studies did their best to engage students and keep them on task, the many disruptions that occur during the semester may have compromised students' ability to benefit from the program; the same students may have benefitted from the program in a different setting. Currently, however, there is very little evidence that CAI is effective in schools.

VI. Conclusion

Educators and policy makers are increasingly intent on using scientifically-based evidence when making decisions about education policy. Thus, education research today must necessarily be focused on identifying the causal relationships between education inputs and student outcomes. The good news is that the body of credible research on causal relationships is growing, and we have started to gather evidence that some school inputs matter while others do not.

As this body of knowledge grows, we can also get inside the “black box” of the inputs that work. Once we understand that an input improves student outcomes, on average, we can look at the next set of questions: Do all students benefit from a particular input? Who benefits most from a particular input? Which aspects of multidimensional programs are most beneficial? (A challenge will be to develop studies that also generate causal estimates of this next generation of questions!) As we develop a knowledge base regarding what works in education, we will also need a better understanding about how to implement appropriate policies using that knowledge.

In addition, policymakers need information with which to assess the tradeoffs between different inputs to make sensible decisions. For example, Jacob and Lefgren (2004b) find a small, but statistically significant, positive effect of summer school and grade retention on student reading skills at a cost of about \$750 per student.²⁷ This cost per student may be compared to other interventions, such as class size reduction, that have larger effects (more than three times as large) on student reading skills but also cost more than \$2000 per student (Krueger 1999). As a result, implementing the summer school and grade retention may be more cost-effective for some school districts than reducing class sizes. This conclusion could not be based on estimates of the effectiveness of grade retention and summer school alone. Clearly, more information on such tradeoffs in educational practice is critical.

Policymakers must also understand that it is much more difficult to credibly evaluate the effectiveness of school policies *after the fact*. Rather, if research and evaluation are part of a

²⁷ Authors’ calculations based on the following assumptions. The current annual cost per pupil in the Chicago Public Schools is about \$9000. If the current school year is about 180 days, then the cost per pupil per day is \$50. The summer school program for 3rd graders was for 6 weeks for one-half day, or for 15 days. Thus, the cost per pupil for the summer school was about \$750.

new policy from the beginning, then researchers can collect the necessary data (which are often difficult—if not impossible—to collect after the policy has been implemented). Further, if a policy change is only to be implemented in a small number of locations, researchers can help policymakers design the selection of locations in a way that meets both political and research needs. Indeed, some of our best opportunities for learning more about the impact of education resources on student outcomes will come from just such partnerships between policymakers and researchers.

Finally we note that good policy is not based on the results of a single study, but rather from a pattern of results extending over time and across a number of settings. Let's take the evidence on small class sizes, as an example. The evidence from the Tennessee class size reduction experiment is important because it has been analyzed by multiple researchers, and the basic results have been found to be robust to alternative ways of analyzing the data. That said, without other credible evidence that smaller class sizes make a difference for students, one would not want to draw such conclusions. Another recent example of the caution with which one must approach a single study comes from the evidence on the *Fast ForWord* computerized language program. Results from Miller et al. (1999) suggest the program has a large and statistically significant effect on student outcomes. However, as discussed earlier, this finding was not found to be robust in alternative settings. Indeed, the purpose of the federally-funded WWC is to provide policymakers with summaries (or meta-analyses) of the best research on any particular topic. This effort reflects the fact that it is only by piecing together results from a variety of high-quality studies that we can begin to develop a picture of what does, and does not, work in education.

References

- Aaronson, D., L. Barrow, and W. Sander. 2003. "Teachers and Student Achievement in the Chicago Public High Schools." Unpublished manuscript. Federal Reserve Bank of Chicago, Chicago, Illinois. Available online at <http://www.chicagofed.org/publications/workingpapers/papers/wp2002-28.pdf> (accessed January 20, 2005).
- Angrist, J. D. 2004. "American Education Research Changes Tack." *Oxford Review of Economic Policy*, 20: 198- 212.
- Angrist, J. D. and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114: 533-575.
- Angrist, J. D. and V. Lavy. 2002. "New Evidence on Classroom Computers and Pupil Learning." *The Economic Journal* 112: 735-765.
- Barrow, L. and C. E. Rouse. 2004. "Using Market Valuation to Assess the Importance and Efficiency of Public School Spending." *Journal of Public Economics* 88: 1747-1769.
- Becker, H. J. 2000. "Who's Wired and Who's Not." *The Future of Children* 10: 44-75.
- Boozer, M. A. and C. E. Rouse. 2001. "Intraschool Variation in Class Size: Patterns and Implications." *Journal of Urban Economics* 50: 163-189.
- Bohrnstedt, G. W. and B. M. Stecher, eds. 1999. *Class Size Reduction in California: Early Evaluation Findings, 1996-1999*, <http://www.classsize.org/techreport/index.htm>. (accessed January 20, 2005).
- Borman, G. D. and L. T. Rachuba. 2001. *Evaluation of the Scientific Learning Corporation's Fast ForWord Computer-Based Training Program in the Baltimore City Public Schools*. A Report Prepared for the Abell Foundation.
- Bryk, A. S. and S. W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage Publications.
- Burtless, G. 1996. Introduction and summary to *Does money matter? The Effect of School Resources on Student Achievement and Adult Success*. Edited by G. Burtless. Washington, D.C.: Brookings Institution Press.
- Campbell, D. T. and J. C. Stanley. 1963. "Experimental and Quasi-Experimental Designs for Research on Teaching." In *Handbook of Research on Teaching*. Edited by N. L. Gage. Chicago: Rand McNally.
- Card, D. and A. B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the

- Characteristics of Public Schools in the United States.” *The Journal of Political Economy* 100: 1-40.
- Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. 2004. “Teacher Sorting, Teacher Shopping, and the Assessment of Teacher Effectiveness.” Unpublished manuscript, Duke University, Durham, North Carolina.
- Coleman, J. S. and E. Q. Campbell with C. F. Hobson, J. McPartland, A. M. Mood, F. D. Weinfield, and R. L. York. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U. S. Office of Education.
- Cook, T. D. 2001. “A Critical Appraisal of the Case Against Using Experiments to Assess School (or Community) Effects.” *Education Next Unabridged Articles*. No. 3 (Fall), <http://www.educationnext.org/unabridged/20013/cook.html> (accessed January 17, 2005).
- Cook, T. D. and D. T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.
- Cooper, H., K. Charlton, J. C. Valentine, and L. Muhlenbruck. 2000. *Making the Most of Summer School: A Meta-Analytic and Narrative Review*. Malden, Massachusetts: Society for Research in Child Development Monograph.
- Cuban, L. 2001. *Oversold and Underused: Computers in the Classroom*. Cambridge, Massachusetts: Harvard University Press.
- Efron, B., and D. Feldman. 1991. “Compliance as an Explanatory Variable in Clinical Trials.” *Journal of the American Statistical Association* 86: 9-17.
- Finn, J. D., and C. M. Achilles. 1990. “Answers and Questions About Class Size: A Statewide Experiment.” *American Educational Research Journal* 27: 557-577.
- Guryan, J. 2003. “Does Money Matter? Estimates from Education Finance Reform in Massachusetts.” Unpublished manuscript, University of Chicago, Chicago, Illinois.
- Hanushek, E. A. 1986. “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature* 24: 1141-1177.
- Hanushek, E. A. 1989. “The Impact of Differential Expenditures on School Performance.” *Educational Researcher* 18: 45-65 .
- Hanushek, E. A. 1996. “Measuring Investment in Education.” *Journal of Economic Perspectives* 10: 9-30.
- Hanushek, E. A. 1997. “Assessing the effects of school resources on student performance: An update.” *Educational Evaluation and Policy Analysis* 19: 141-164.

- Hanushek, E. A., S. G. Rivkin, and J. F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73: 417-458.
- Hedges, L. V., R. Laine, and R. Greenwald. 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Education Researcher* 23: 5-14.
- Hoxby, C. M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115: 1239-1285.
- Imbens, G. and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467-475.
- Jacob, B. A. and L. Lefgren. 2004a. "The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago." *Journal of Human Resources* 39: 50-79.
- Jacob, B. A. and L. Lefgren. 2004b. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics* 86: 226-244.
- Johnson, G. E. and Stafford, F. P. 1973. "Social Returns to Quantity and Quality of Schooling." *Journal of Human Resources* 8: 139-155.
- Krueger, A. B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114: 497-531.
- Krueger, A. B. 2003. "Economic Considerations and Class Size." *The Economic Journal* 113: F34-F63.
- Krueger, A. B. and D. M. Whitmore. 2001. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *The Economic Journal* 111: 1-28.
- Lepper, M. R. 1985. "Microcomputers in Education, Motivational and Social Issues." *American Psychologist* 40: 1-18.
- Means, B., and Olson, K. 1997. *Technology and Education Reform*. Office of Educational Research and Improvement, Contract No. RP91-172010. Washington, DC: U.S. Department of Education.
- Miller, S. L., M. M. Merzenich, P. Tallal, K. DeVivo, K. La-Rossa, N. Linn, A. Pycha, B. E. Peterson, and W. M. Jenkins. 1999. "Fast ForWord Training in Children with Low Reading Performance." *Nederlandse vereniging voor lopopedie en foniatrye: 1999 jaarcongres auditieve vaardigheden en spraak-taal*.

- Pischke, J. 2003. "The Impact of Length of School Year on Student Performance and Earnings: Evidence from the German Short School Years." National Bureau of Economic Research Working paper 9964.
- Rizzuto, R. and Wachtel, P. 1980. "Further Evidence on the Returns to School Quality." *Journal of Human Resources* 15: 240-254.
- Rouse, C. E. "Accounting for Schools: Econometric Issues in Measuring School Quality." In *Measurement and Research Issues in a New Accountability Era*. Edited by Carol Anne Dwyer. New Jersey: Lawrence Erlbaum Associates, 2005.
- Rouse, C. E. and A. B. Krueger with L. Markman. 2004. "Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically-based' Reading Program." *Economics of Education Review* 23: 323-338.
- Rubin, D. 1974. "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies." *Journal of Educational Psychology* 66: 688-701.
- Sandholtz, J. H., C. Ringstaff, and D. C. Dwyer. 1997. *Teaching with Technology: Creating Student-Centered Classrooms*. New York: Teachers College Press.
- Sewell, W. H. 1967. Review of *Equality of Educational Opportunity* by J. S. Coleman and E. Q. Campbell with C. F. Hobson, J. McPartland, A. M. Mood, F. D. Weinfield, and R. L. York. *American Sociological Review* 32: 475-479.
- Smith, J. P. and F. R. Welch .1989. "Black Economic Progress After Myrdal." *Journal of Economic Literature* 27: 519-564.
- Smylie, Mark A., E. Allensworth, R. C. Greenberg, R. Harris, and S. Luppescu. 2001. *Teacher Professional Development in Chicago: Supporting Effective Practice*. Chicago: Consortium on Chicago School Research. Also available online at <http://www.consortium-chicago.org/publications/pdfs/p0d01.pdf>.
- Summers, A. A. and B. L. Wolfe. 1977. "Do Schools Make a Difference?" *American Economic Review* 67: 639-652.
- Todd, P. E. and K. I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal* 113: F3-F33.
- U.S. Department of Education. 2002. U.S. Department of Education Awards Contract for 'What Works Clearinghouse.' <http://www.ed.gov/news/pressreleases/2002/08/08072002a.html> (accessed January 20, 2005).

Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

Outsourcing Business Services and the Role of Central Administrative Offices <i>Yukako Ono</i>	WP-02-01
Strategic Responses to Regulatory Threat in the Credit Card Market* <i>Victor Stango</i>	WP-02-02
The Optimal Mix of Taxes on Money, Consumption and Income <i>Fiorella De Fiore and Pedro Teles</i>	WP-02-03
Expectation Traps and Monetary Policy <i>Stefania Albanesi, V. V. Chari and Lawrence J. Christiano</i>	WP-02-04
Monetary Policy in a Financial Crisis <i>Lawrence J. Christiano, Christopher Gust and Jorge Roldos</i>	WP-02-05
Regulatory Incentives and Consolidation: The Case of Commercial Bank Mergers and the Community Reinvestment Act <i>Raphael Bostic, Hamid Mehran, Anna Paulson and Marc Saidenberg</i>	WP-02-06
Technological Progress and the Geographic Expansion of the Banking Industry <i>Allen N. Berger and Robert DeYoung</i>	WP-02-07
Choosing the Right Parents: Changes in the Intergenerational Transmission of Inequality — Between 1980 and the Early 1990s <i>David I. Levine and Bhashkar Mazumder</i>	WP-02-08
The Immediacy Implications of Exchange Organization <i>James T. Moser</i>	WP-02-09
Maternal Employment and Overweight Children <i>Patricia M. Anderson, Kristin F. Butcher and Phillip B. Levine</i>	WP-02-10
The Costs and Benefits of Moral Suasion: Evidence from the Rescue of Long-Term Capital Management <i>Craig Furfine</i>	WP-02-11
On the Cyclical Behavior of Employment, Unemployment and Labor Force Participation <i>Marcelo Veracierto</i>	WP-02-12
Do Safeguard Tariffs and Antidumping Duties Open or Close Technology Gaps? <i>Meredith A. Crowley</i>	WP-02-13
Technology Shocks Matter <i>Jonas D. M. Fisher</i>	WP-02-14
Money as a Mechanism in a Bewley Economy <i>Edward J. Green and Ruilin Zhou</i>	WP-02-15

Working Paper Series *(continued)*

Optimal Fiscal and Monetary Policy: Equivalence Results <i>Isabel Correia, Juan Pablo Nicolini and Pedro Teles</i>	WP-02-16
Real Exchange Rate Fluctuations and the Dynamics of Retail Trade Industries on the U.S.-Canada Border <i>Jeffrey R. Campbell and Beverly Lapham</i>	WP-02-17
Bank Procyclicality, Credit Crunches, and Asymmetric Monetary Policy Effects: A Unifying Model <i>Robert R. Bliss and George G. Kaufman</i>	WP-02-18
Location of Headquarter Growth During the 90s <i>Thomas H. Klier</i>	WP-02-19
The Value of Banking Relationships During a Financial Crisis: Evidence from Failures of Japanese Banks <i>Elijah Brewer III, Hesna Genay, William Curt Hunter and George G. Kaufman</i>	WP-02-20
On the Distribution and Dynamics of Health Costs <i>Eric French and John Bailey Jones</i>	WP-02-21
The Effects of Progressive Taxation on Labor Supply when Hours and Wages are Jointly Determined <i>Daniel Aaronson and Eric French</i>	WP-02-22
Inter-industry Contagion and the Competitive Effects of Financial Distress Announcements: Evidence from Commercial Banks and Life Insurance Companies <i>Elijah Brewer III and William E. Jackson III</i>	WP-02-23
State-Contingent Bank Regulation With Unobserved Action and Unobserved Characteristics <i>David A. Marshall and Edward Simpson Prescott</i>	WP-02-24
Local Market Consolidation and Bank Productive Efficiency <i>Douglas D. Evanoff and Evren Örs</i>	WP-02-25
Life-Cycle Dynamics in Industrial Sectors. The Role of Banking Market Structure <i>Nicola Cetorelli</i>	WP-02-26
Private School Location and Neighborhood Characteristics <i>Lisa Barrow</i>	WP-02-27
Teachers and Student Achievement in the Chicago Public High Schools <i>Daniel Aaronson, Lisa Barrow and William Sander</i>	WP-02-28
The Crime of 1873: Back to the Scene <i>François R. Velde</i>	WP-02-29
Trade Structure, Industrial Structure, and International Business Cycles <i>Marianne Baxter and Michael A. Kouparitsas</i>	WP-02-30
Estimating the Returns to Community College Schooling for Displaced Workers <i>Louis Jacobson, Robert LaLonde and Daniel G. Sullivan</i>	WP-02-31

Working Paper Series *(continued)*

A Proposal for Efficiently Resolving Out-of-the-Money Swap Positions at Large Insolvent Banks <i>George G. Kaufman</i>	WP-03-01
Depositor Liquidity and Loss-Sharing in Bank Failure Resolutions <i>George G. Kaufman</i>	WP-03-02
Subordinated Debt and Prompt Corrective Regulatory Action <i>Douglas D. Evanoff and Larry D. Wall</i>	WP-03-03
When is Inter-Transaction Time Informative? <i>Craig Furfine</i>	WP-03-04
Tenure Choice with Location Selection: The Case of Hispanic Neighborhoods in Chicago <i>Maude Toussaint-Comeau and Sherrie L.W. Rhine</i>	WP-03-05
Distinguishing Limited Commitment from Moral Hazard in Models of Growth with Inequality* <i>Anna L. Paulson and Robert Townsend</i>	WP-03-06
Resolving Large Complex Financial Organizations <i>Robert R. Bliss</i>	WP-03-07
The Case of the Missing Productivity Growth: Or, Does information technology explain why productivity accelerated in the United States but not the United Kingdom? <i>Susanto Basu, John G. Fernald, Nicholas Oulton and Sylaja Srinivasan</i>	WP-03-08
Inside-Outside Money Competition <i>Ramon Marimon, Juan Pablo Nicolini and Pedro Teles</i>	WP-03-09
The Importance of Check-Cashing Businesses to the Unbanked: Racial/Ethnic Differences <i>William H. Greene, Sherrie L.W. Rhine and Maude Toussaint-Comeau</i>	WP-03-10
A Firm's First Year <i>Jaap H. Abbring and Jeffrey R. Campbell</i>	WP-03-11
Market Size Matters <i>Jeffrey R. Campbell and Hugo A. Hopenhayn</i>	WP-03-12
The Cost of Business Cycles under Endogenous Growth <i>Gadi Barlevy</i>	WP-03-13
The Past, Present, and Probable Future for Community Banks <i>Robert DeYoung, William C. Hunter and Gregory F. Udell</i>	WP-03-14
Measuring Productivity Growth in Asia: Do Market Imperfections Matter? <i>John Fernald and Brent Neiman</i>	WP-03-15
Revised Estimates of Intergenerational Income Mobility in the United States <i>Bhashkar Mazumder</i>	WP-03-16

Working Paper Series *(continued)*

Product Market Evidence on the Employment Effects of the Minimum Wage <i>Daniel Aaronson and Eric French</i>	WP-03-17
Estimating Models of On-the-Job Search using Record Statistics <i>Gadi Barlevy</i>	WP-03-18
Banking Market Conditions and Deposit Interest Rates <i>Richard J. Rosen</i>	WP-03-19
Creating a National State Rainy Day Fund: A Modest Proposal to Improve Future State Fiscal Performance <i>Richard Mattoon</i>	WP-03-20
Managerial Incentive and Financial Contagion <i>Sujit Chakravorti, Anna Llyina and Subir Lall</i>	WP-03-21
Women and the Phillips Curve: Do Women's and Men's Labor Market Outcomes Differentially Affect Real Wage Growth and Inflation? <i>Katharine Anderson, Lisa Barrow and Kristin F. Butcher</i>	WP-03-22
Evaluating the Calvo Model of Sticky Prices <i>Martin Eichenbaum and Jonas D.M. Fisher</i>	WP-03-23
The Growing Importance of Family and Community: An Analysis of Changes in the Sibling Correlation in Earnings <i>Bhashkar Mazumder and David I. Levine</i>	WP-03-24
Should We Teach Old Dogs New Tricks? The Impact of Community College Retraining on Older Displaced Workers <i>Louis Jacobson, Robert J. LaLonde and Daniel Sullivan</i>	WP-03-25
Trade Deflection and Trade Depression <i>Chad P. Brown and Meredith A. Crowley</i>	WP-03-26
China and Emerging Asia: Comrades or Competitors? <i>Alan G. Ahearne, John G. Fernald, Prakash Loungani and John W. Schindler</i>	WP-03-27
International Business Cycles Under Fixed and Flexible Exchange Rate Regimes <i>Michael A. Kouparitsas</i>	WP-03-28
Firing Costs and Business Cycle Fluctuations <i>Marcelo Veracierto</i>	WP-03-29
Spatial Organization of Firms <i>Yukako Ono</i>	WP-03-30
Government Equity and Money: John Law's System in 1720 France <i>François R. Velde</i>	WP-03-31
Deregulation and the Relationship Between Bank CEO Compensation and Risk-Taking <i>Elijah Brewer III, William Curt Hunter and William E. Jackson III</i>	WP-03-32

Working Paper Series *(continued)*

Compatibility and Pricing with Indirect Network Effects: Evidence from ATMs <i>Christopher R. Knittel and Victor Stango</i>	WP-03-33
Self-Employment as an Alternative to Unemployment <i>Ellen R. Rissman</i>	WP-03-34
Where the Headquarters are – Evidence from Large Public Companies 1990-2000 <i>Tyler Diacon and Thomas H. Klier</i>	WP-03-35
Standing Facilities and Interbank Borrowing: Evidence from the Federal Reserve’s New Discount Window <i>Craig Furfine</i>	WP-04-01
Netting, Financial Contracts, and Banks: The Economic Implications <i>William J. Bergman, Robert R. Bliss, Christian A. Johnson and George G. Kaufman</i>	WP-04-02
Real Effects of Bank Competition <i>Nicola Cetorelli</i>	WP-04-03
Finance as a Barrier To Entry: Bank Competition and Industry Structure in Local U.S. Markets? <i>Nicola Cetorelli and Philip E. Strahan</i>	WP-04-04
The Dynamics of Work and Debt <i>Jeffrey R. Campbell and Zvi Hercowitz</i>	WP-04-05
Fiscal Policy in the Aftermath of 9/11 <i>Jonas Fisher and Martin Eichenbaum</i>	WP-04-06
Merger Momentum and Investor Sentiment: The Stock Market Reaction To Merger Announcements <i>Richard J. Rosen</i>	WP-04-07
Earnings Inequality and the Business Cycle <i>Gadi Barlevy and Daniel Tsiddon</i>	WP-04-08
Platform Competition in Two-Sided Markets: The Case of Payment Networks <i>Sujit Chakravorti and Roberto Roson</i>	WP-04-09
Nominal Debt as a Burden on Monetary Policy <i>Javier Díaz-Giménez, Giorgia Giovannetti, Ramon Marimon, and Pedro Teles</i>	WP-04-10
On the Timing of Innovation in Stochastic Schumpeterian Growth Models <i>Gadi Barlevy</i>	WP-04-11
Policy Externalities: How US Antidumping Affects Japanese Exports to the EU <i>Chad P. Bown and Meredith A. Crowley</i>	WP-04-12
Sibling Similarities, Differences and Economic Inequality <i>Bhashkar Mazumder</i>	WP-04-13
Determinants of Business Cycle Comovement: A Robust Analysis <i>Marianne Baxter and Michael A. Kouparitsas</i>	WP-04-14

Working Paper Series *(continued)*

The Occupational Assimilation of Hispanics in the U.S.: Evidence from Panel Data <i>Maude Toussaint-Comeau</i>	WP-04-15
Reading, Writing, and Raisinets ¹ : Are School Finances Contributing to Children's Obesity? <i>Patricia M. Anderson and Kristin F. Butcher</i>	WP-04-16
Learning by Observing: Information Spillovers in the Execution and Valuation of Commercial Bank M&As <i>Gayle DeLong and Robert DeYoung</i>	WP-04-17
Prospects for Immigrant-Native Wealth Assimilation: Evidence from Financial Market Participation <i>Una Okonkwo Osili and Anna Paulson</i>	WP-04-18
Individuals and Institutions: Evidence from International Migrants in the U.S. <i>Una Okonkwo Osili and Anna Paulson</i>	WP-04-19
Are Technology Improvements Contractionary? <i>Susanto Basu, John Fernald and Miles Kimball</i>	WP-04-20
The Minimum Wage, Restaurant Prices and Labor Market Structure <i>Daniel Aaronson, Eric French and James MacDonald</i>	WP-04-21
Betcha can't acquire just one: merger programs and compensation <i>Richard J. Rosen</i>	WP-04-22
Not Working: Demographic Changes, Policy Changes, and the Distribution of Weeks (Not) Worked <i>Lisa Barrow and Kristin F. Butcher</i>	WP-04-23
The Role of Collateralized Household Debt in Macroeconomic Stabilization <i>Jeffrey R. Campbell and Zvi Hercowitz</i>	WP-04-24
Advertising and Pricing at Multiple-Output Firms: Evidence from U.S. Thrift Institutions <i>Robert DeYoung and Evren Örs</i>	WP-04-25
Monetary Policy with State Contingent Interest Rates <i>Bernardino Adão, Isabel Correia and Pedro Teles</i>	WP-04-26
Comparing location decisions of domestic and foreign auto supplier plants <i>Thomas Klier, Paul Ma and Daniel P. McMillen</i>	WP-04-27
China's export growth and US trade policy <i>Chad P. Bown and Meredith A. Crowley</i>	WP-04-28
Where do manufacturing firms locate their Headquarters? <i>J. Vernon Henderson and Yukako Ono</i>	WP-04-29
Monetary Policy with Single Instrument Feedback Rules <i>Bernardino Adão, Isabel Correia and Pedro Teles</i>	WP-04-30

Working Paper Series *(continued)*

Firm-Specific Capital, Nominal Rigidities and the Business Cycle <i>David Altig, Lawrence J. Christiano, Martin Eichenbaum and Jesper Linde</i>	WP-05-01
Do Returns to Schooling Differ by Race and Ethnicity? <i>Lisa Barrow and Cecilia Elena Rouse</i>	WP-05-02
Derivatives and Systemic Risk: Netting, Collateral, and Closeout <i>Robert R. Bliss and George G. Kaufman</i>	WP-05-03
Risk Overhang and Loan Portfolio Decisions <i>Robert DeYoung, Anne Gron and Andrew Winton</i>	WP-05-04
Characterizations in a random record model with a non-identically distributed initial record <i>Gadi Barlevy and H. N. Nagaraja</i>	WP-05-05
Price discovery in a market under stress: the U.S. Treasury market in fall 1998 <i>Craig H. Furfine and Eli M. Remolona</i>	WP-05-06
Politics and Efficiency of Separating Capital and Ordinary Government Budgets <i>Marco Bassetto with Thomas J. Sargent</i>	WP-05-07
Rigid Prices: Evidence from U.S. Scanner Data <i>Jeffrey R. Campbell and Benjamin Eden</i>	WP-05-08
Entrepreneurship, Frictions, and Wealth <i>Marco Cagetti and Mariacristina De Nardi</i>	WP-05-09
Wealth inequality: data and models <i>Marco Cagetti and Mariacristina De Nardi</i>	WP-05-10
What Determines Bilateral Trade Flows? <i>Marianne Baxter and Michael A. Kouparitsas</i>	WP-05-11
Intergenerational Economic Mobility in the U.S., 1940 to 2000 <i>Daniel Aaronson and Bhashkar Mazumder</i>	WP-05-12
Differential Mortality, Uncertain Medical Expenses, and the Saving of Elderly Singles <i>Mariacristina De Nardi, Eric French, and John Bailey Jones</i>	WP-05-13
Fixed Term Employment Contracts in an Equilibrium Search Model <i>Fernando Alvarez and Marcelo Veracierto</i>	WP-05-14
Causality, Causality, Causality: The View of Education Inputs and Outputs from Economics <i>Lisa Barrow and Cecilia Elena Rouse</i>	WP-05-15