DATA SHARE AND USE AGREEMENT

between

ILLINOIS STATE BOARD OF EDUCATION

and

THE CENTER FOR RESEARCH ON EDUCATION OUTCOMES (CREDO)

This Data Share and Use Agreement (hereinafter "Agreement") is hereby made and entered into by and between the Illinois State Board of Education ("ISBE") and The Board of Trustees of The Leland Stanford Junior University for CREDO (the "Recipient") (each a "Party" and collectively "the Parties").

I. RECITALS

1. ISBE owns and maintains information, including individually identifiable information, on students, including prior Illinois students ("Student Data"); teachers, including teacher certification and service record data ("Teacher Data"); programs; schools or institutions; and districts (collectively "Confidential Data") necessary for required federal reporting and to audit and evaluate education programs and to perform studies for, or on behalf of, public elementary and secondary schools, all in a manner consistent with the Family Educational Rights and Privacy Act (20 U.S.C. § 1232g) ("FERPA"), the Illinois School Students Records Act (105 ILCS 10/1, *et seq.*) ("ISSRA"), and other applicable laws.

2. The term "individually identifiable information" means information that is identifiable to a particular individual, program, classroom, school, institution or district, including but not limited to the following: (a) a first and last name; (b) a home or other physical address, including street name and name of a city, town, or county; (c) an e-mail address; (d) a telephone number; (e) a social security, employer identification, or student identification number; (f) test scores; or (g) clinical information, including any questionnaires, notes, or other documentation.

3. ISBE is authorized by law to secure, compile, catalog, publish and preserve information and data relative to the public school system of Illinois, making such comparison as will assist the General Assembly in determining the priorities of educational programs to be of value to the public school system of Illinois and of other states (105 ILCS 5/2-3.31).

4. ISBE is the state agency in Illinois responsible for educational policies and guidelines for public schools, pre-school through grade 12, and Vocational Education in Illinois and is responsible for analyzing the present and future aims, needs and requirements of education in Illinois (105 ILCS 5/1A-4 (C)).

5. Recipient is a leading independent voice in the discussion of how to improve education in America, with an emphasis on rigorous program and policy analysis as the means of informing and improving education decision making

6. Recipient will engage in research and evaluation of the Confidential Data as ISBE's authorized representative. This research and evaluation, as more fully set forth in Exhibit A, the Specifications for Shared Data and Use ("Specifications"), attached hereto, will evaluate the impact of charter school attendance on student academic progress and emerging policy issues that affect both the traditional public school and charter school sectors, as defined in Section A. of the Specifications, ("Research Project").

7. Federal law allows the release of educational records or personally identifiable information of students without the consent of students or parents ("individually identifiable student information"), so long as the disclosure is to authorized representatives of state educational authorities for purposes of audit or evaluation of state supported education programs, or to organizations or individuals conducting studies for, or on behalf of, educational agencies or institutions for the purpose of improving instruction; so long as the studies are conducted in such a manner that they do not allow for the personal identification of students and parents by persons other than representatives of such organizations who have been determined by the state educational authority to have legitimate interests in the information and who are under the direct control of the state educational authority; and the information is destroyed when no longer needed for the purpose for which the study was conducted. 20 U.S.C. § 1232g(b)(1)(C)(i)(III) and (b)(1)(F) and 34 C.F.R. § 99.31 (a)(1)(i)(B), § 99.31(a)(3)(iv) and (a)(6)(i) and (a)(6)(ii).

8. The Parties wish to enter into this Agreement in order to:

   A. Establish Recipient as ISBE's authorized representative for purposes of Recipient providing ISBE with research, analysis, audit and/or evaluation of the State's educational system for the improvement of educational instruction;

   B. Establish the necessary data sharing arrangements between the Parties to provide data necessary to conduct research, analysis, and evaluation;

C. Facilitate the audit or evaluation of education programs for, or on behalf of, ISBE in a manner permitted by FERPA, ISSRA, and other applicable law ("Audit or Evaluation"); and

D. Facilitate the performance of studies for, or on behalf of, ISBE in a manner permitted by FERPA, ISSRA, and other applicable law ("Research").

II. DATA SHARING AUTHORIZATIONS

1. ISBE hereby designates and recognizes Recipient as its authorized representative for purposes of Research, Audit and Evaluation related to the impact of Illinois charter schools on student learning compared to matched students from traditional public schools in Illinois. In addition, Recipient will include Illinois in a multi-state study of charter school performance that will produce a pooled assessment of the impact of charter schools on student achievement and growth as well as state-by-state comparisons.

2. ISBE agrees to share with Recipient the ISBE data set forth in the Specifications solely for the limited purposes and extent as specified in Exhibit A. Recipient agrees to use the Confidential Data for the Research Project only as authorized pursuant to this Agreement. Any act by Recipient that involves a use beyond that set forth herein shall be deemed in its entirety to be a prohibited use of the Confidential Data.

3. Recipient will securely store and maintain the Confidential Data at Stanford University in room 114 of the Lou Henry Hoover Building and at Texas Schools Project in room 2.504 of the Waterview Science and Technology Center, in accordance with the terms and conditions set forth in Exhibit E ("Data Security Plan"), attached hereto.

4. All officials and employees authorized to request, receive and obtain information, including the Confidential Data, from Recipient under this Agreement are identified in Exhibit A, Attachment 1. Any further disclosure to officials and employees identified by Recipient that will have access to information as provided under this Agreement shall be documented through completion of a form meeting the requirements of Exhibit C ("Form for Amendment of Exhibit A, Attachment 1") and an executed copy(ies) of Exhibit E, Attachment 1 ("Security Pledge for the Use of Confidential Data"). No disclosure of information provided under this Agreement shall be permitted until a form meeting the requirements of Exhibit C and an executed Security Pledge for the Use of Confidential Data is received and approved by ISBE.

5. All contractors, subcontractors, or agents authorized to request, receive or obtain information, including Confidential Data, from Recipient under this Agreement are identified in Exhibit A, Attachment 2. Any further disclosure to contractors, subcontractors, and agents identified by Recipient that will have access to information as provided under this Agreement shall be documented through completion of a form meeting the requirements of Exhibit D ("Form for Amendment of Exhibit A, Attachment 2") and an executed copy(ies) of Exhibit E, Attachment 1 ("Security Pledge for the Use of Confidential Data"). No disclosure of information provided under this Agreement shall be permitted until a form meeting the requirements of Exhibit D and an executed Security Pledge for the Use of Confidential Data is received and approved by ISBE.

6. Recipient agrees that it is ultimately responsible for ensuring that any third-party, including any employee, contractor, subcontractor, or agent of Recipient, operates in accordance with the terms and conditions of this Agreement.

7. Recipient is responsible for ensuring that any third party acquirer of ISBE Confidential Data, employed by, under contract to or working in collaboration with Recipient, operate within the terms of this Agreement; not retain any Confidential Data permanently, but is required to return any Confidential Data to ISBE upon request, or to Recipient within 10 days of Recipient's request, and/or upon completion of the work or termination, cancellation or expiration of this Agreement. ISBE may request that the Confidential Data be permanently destroyed and that written certification of destruction be sent to ISBE.

8. ISBE makes no representations or warranties, express or implied, with respect to the Confidential Data. ISBE shall not be liable to the Recipient for amounts representing the loss of profits, loss of business or indirect, consequential or punitive damages in connection with the provision or use of Confidential Data under this Agreement.

9. The individuals signing below on behalf of ISBE represent that, with respect to any agreement between any third-party and the Recipient, ISBE's signatures attached hereto are intended solely as an acknowledgement of the separate data agreements with the third-party and do not suggest or imply acceptance of the terms and conditions of any agreements between the Recipient and the third-party, nor constitute an endorsement or approval of any such agreements by the State of Illinois.

III. DATA ACCESS, USE AND SECURITY

1. **Restrictions on Recipient.** The data access, use, and security restrictions set forth in this Section shall apply to the receipt, use, disclosure, and maintenance of Confidential Data by Recipient. Recipient agrees to the following:

A. Confidential Data may only be used for the purpose or purposes authorized pursuant to this Agreement.

B. Recipient will comply with all applicable laws, materials, regulations and all other State and Federal requirements with respect to the protection of privacy, security and dissemination of the shared data.

C. Recipient will comply with the relevant requirements of FERPA (20 U.S.C. § 1232g) and ISSRA (105 ILCS 10/1 *et seq.*), regarding the confidentiality of Student Data, and specifically "education records" as defined in FERPA and "school student records" as defined in ISSRA. Any use of information contained in student education records to be released must be approved by ISBE. To protect the confidentiality of student education records, Recipient will limit access to student education records to those employees who reasonably need access to them in order to perform their responsibilities under this Agreement.

D. Recipient will follow ISBE's confidentiality requirements for all ISBE data, pursuant to the Data Processing Confidentiality Act (30 ILCS 585/0.01 *et seq.*). Information obtained from any individual shall comply with the following terms and conditions, which include, but are not limited to:

- Be confidential;
- Not be published or open to public inspection;
- Not be used directly in any court in any pending action or proceeding; and
- Not be admissible in evidence in any action or proceeding.

All records and other information maintained by ISBE regarding any person are confidential and shall be protected from unauthorized use and/or disclosure under this Agreement. Any dissemination or use of the Confidential Data for other than the primary purpose of this Agreement without the express written authority of ISBE is specifically prohibited. Confidential Data released under this Agreement are solely for the use of Recipient and are to be used only for the specific purposes as described in the Specifications.

E. In the event that any Confidential Data is required to be disclosed in response to a valid order of a court of competent jurisdiction or other governmental body of the United States or any political subdivisions thereof, Recipient shall first (a) notify ISBE of the order and provide a complete copy of such order to ISBE and (b) permit ISBE to seek an appropriate protective order. Recipient shall fully cooperate with ISBE if ISBE wishes to apply to such court for a protective order. Recipient shall only disclose the Confidential Data to the extent necessary and for the purposes of the court or other governmental body. Furthermore, Recipient must comply with the notice requirements of FERPA (34

C.F.R. § 99.31(a)(9)(ii) when and if it is required to disclose any Student Data in accordance with a lawfully issued subpoena or court order. 34 C.F.R. § 99.33(b)(2).

F. Nothing in this Agreement may be construed to allow Recipient to maintain, use, disclose, or share the Confidential Data in a manner not allowed by state or federal law or regulation, including but not limited to FERPA (20 U.S.C. § 1232g) and ISSRA (105 ILCS 10/1, *et seq.*).

G. Recipient will not share Confidential Data with anyone, except those employees and contractors, subcontractors and agents of Recipient as identified in Exhibit A, Attachments 1 and 2, as may be amended from time to time in accordance with the terms and conditions of this Agreement. Recipient will instruct all persons having access to Confidential Data on the use and confidentiality restrictions set forth in this Agreement and sanctions for unauthorized disclosure and shall require all employees, contractors, subcontractors, or agents of any kind to comply with all applicable provisions of FERPA and other state and federal laws with respect to the Confidential Data. Recipient shall provide executed copies of Exhibit E, Attachment 1 ("Security Pledge for the Use of Confidential Data") for each such person and upon ISBE's request, shall produce a written acknowledgement from all such persons verifying that the instruction required under this Section has occurred.

H. Recipient will not disclose any individually identifiable information or Confidential Data under this Agreement in a manner which could identify an individual student, person, program, school, institution, or district except as authorized by ISBE and applicable law. Disclosure includes, without limitation, disclosure of information, research, or analysis in a manner that permits the personal identification of parents and students, as such terms are defined in the FERPA regulations (34 C.F.R Part 99), or individual identification of a person, program, school, institution, or district; and includes, de-identified or aggregate data in cell sizes of less than ten (10) for each category or subcategory of data, and de-identified or aggregate data in cell sizes of more than ten (10) for each category or subcategory that, when disaggregated could lead to indirect disclosure through the disclosure, through the cumulative effects of disclosures, or when combined with other data element(s) in the public domain.

I. Recipient may not re-disclose Student Data to any other person or entity unless permitted by law and approved in advance under an amendment to this Agreement and, if required by law, the written consent of the parent or student if such student may consent to disclosure under FERPA. Re-disclosure of Student Data includes, without limitation, disclosure of information, research, or analysis in a manner that permits the personal identification of parents and students, as such terms are defined in the FERPA regulations (34 C.F.R. Part 99); and includes, de-identified or aggregate data in cell sizes of less than ten (10) for each category or subcategory of data, and de-identified or aggregate data in cell sizes of more than ten (10) for each category or subcategory that, when disaggregated could lead to indirect disclosure through the disclosure, through the cumulative effects of disclosures, or when combined with other data element(s) in the public domain.

J. Recipient certifies that it has the capacity to restrict access to the Confidential Data and maintain the security of electronic information, as more fully set forth in Exhibit E ("Data Security Plan"). Recipient shall develop, implement, maintain and use appropriate administrative, technical and physical security measures to preserve the confidentiality, integrity and availability of all electronically maintained or transmitted Confidential Data received from, or on behalf of, ISBE. Recipient acknowledges that the use of unsecured telecommunications, including the Internet, to transmit individually identifiable or deducible information derived from the Confidential Data specified in Exhibit A is strictly prohibited. Recipient agrees that all data transmissions will be encrypted and provided through a secure FTP site. These measures will be extended by contract to all employees, contractors, subcontractors, or agents that will receive Confidential Data provided by this Agreement and used by Recipient.

K. Recipient will not provide any of the Confidential Data obtained pursuant to this Agreement to any party ineligible to receive data protected by FERPA or prohibited from receiving data from any entity by virtue of a finding under subsections 99.67(c), (d) or (e) of Title 34 of the Code of Federal Regulations. 34 C.F.R. § 99.67 (c), (d) and (e).

L. Recipient agrees to fully report to ISBE within one day of discovery any infraction of the confidentiality provisions and any use or disclosure of Confidential Data not authorized by this Agreement or in writing by ISBE. Recipient's report shall identify: (i) the nature of the unauthorized use or disclosure; (ii) the Confidential Data used or disclosed; (iii) who made the unauthorized use and/or received the unauthorized disclosure; (iv) what Recipient has done or shall do to mitigate any deleterious effect of the unauthorized use or disclosure; and (v) what corrective action Recipient has taken or shall take to prevent future similar unauthorized use and/or disclosure. Recipient shall provide such other information, including a written report, as reasonably requested by ISBE.

M. Recipient agrees that Confidential Data shall not be archived or sent to a records center and shall not be retained for any period longer than the earlier of: (i) the expiration of the Term of this Agreement or (ii) the completion of the Research Project (the "Completion Date ").

N. Recipient agrees to secure any and all data received pursuant to this Agreement and agrees to establish, secure and retain records of access and use of all Confidential Data received pursuant to this Agreement. Recipient agrees to allow ISBE on-site inspection and access to all relevant data files and servers to verify data security and usage, as well as audit access, throughout the Term of this Agreement and for a period of three (3) years following the Completion Date, whichever is longer. The three (3) year period shall be extended for the duration of any audit in progress during the Term. No fees shall be assessed for such access, audit, or review, and Recipient agrees to cooperate with ISBE's efforts to verify data security and usage.

O. Recipient must create and maintain a record of any disclosure of Confidential Data made to any other person or entity pursuant to this Agreement. The record of disclosure must record the name of any person or organization receiving the Confidential Data and their legitimate interest in the Confidential Data for Audit, Evaluation, or Research. The

record must also describe the Confidential Data included within the disclosure by class, school, district, or other appropriate grouping. Upon ISBE's request, Recipient must provide a copy of the record of further disclosures to ISBE.

P. Any breach of the security of any Confidential Data provided to any person or entity under this Agreement shall be subject to the terms and provisions of the Personal Information Protection Act (815 ILCS 530/1, *et seq.*).

Q. Recipient represents and agrees that any and all approvals for the research to be conducted using the Confidential Data, where required by law, from the Recipient or the Recipient's Institutional Review Board ("IRB") have been obtained. ISBE may request a copy of any review completed by Recipient or the Recipient's IRB related to the Confidential Data; and Recipient shall provide ISBE with a copy of the requested review within ten (10) working days of ISBE's written request.

R. Recipient may not assign its obligations under this Agreement, or any part of its interest in this Agreement, without the prior written consent of ISBE. Any assignment made without said consent shall be null and void.

S. Recipient recognizes and agrees that the Confidential Data it obtains under this Agreement is the property of ISBE and shall be disposed of or returned to ISBE within ten (10) days, upon ISBE's request, if Confidential Data has not been pooled with other states' data and 60 days if Confidential Data has already been pooled with other states' data. All Confidential Data received pursuant to this Agreement shall be disposed of upon termination, cancellation, expiration, or other conclusion of this Agreement. Disposal means the return of the Confidential Data to ISBE or destruction of the Confidential Data as directed by ISBE, including purging of all copies from the Recipient's computer systems. If Recipient destroys the Confidential Data, it shall provide ISBE with a written certificate identifying the data and confirming the method and date of destruction. Recipient agrees to require all employees, contractors, subcontractors, or agents of any kind to comply with this provision.

2. Recipient must ensure that any third-party recipient of the Confidential Data working under or in collaboration with Recipient agrees by contractual terms to the provisions of this Agreement for the sharing, disclosure, re-disclosure, use, maintenance, security and destruction of the Confidential Data.

3. The terms and provisions of this Section III shall apply to the use of Confidential Data received by Recipient for so long as Recipient retains the data and shall survive the expiration or earlier termination of this Agreement.

IV. TERM AND TERMINATION

1. **Term**. This Agreement shall become effective on the date of signature of the last signatory to the Agreement and, subject to any earlier termination as provided herein, shall remain in full force and effect through and including March 31, 2014 (the "Term"). At the sole option of ISBE and subject to the mutual agreement of the Parties, this Agreement may be renewed for five annual terms.

2. **Termination**. This Agreement may be terminated by either Party upon thirty (30) days written notice to the other Party.

3. **Termination for Breach**. Notwithstanding any other provisions to the contrary, this Agreement is subject to immediate cancellation by ISBE for failure of Recipient or its authorized employee, contractor, subcontractor, or agent to adhere to any provision set forth in this Agreement.

4. **Termination upon Unauthorized Data Disclosure**. Notwithstanding any other provisions to the contrary, ISBE may immediately terminate its participation in this Agreement if any Confidential Data disclosed by ISBE to Recipient is used in any manner which violates the terms and provisions of this Agreement, ISSRA and/or FERPA.

5. **Survival**. Notwithstanding any other provision in this Agreement, the terms of this Agreement regarding the use, confidentiality, and secure maintenance of data shall survive the termination of the Agreement and continue in full force and effect.

V. GENERAL PROVISIONS

1. **Amendment**. This Agreement may be amended only by a written instrument signed by the Parties. Notwithstanding the foregoing, any amendment to modify or add to the scope of the research or data to be shared must be signed by Recipient and the Illinois State Superintendent of Education and ISBE General Counsel and shall be in the form as set forth in Exhibit B (the "Form for Amendment of Research Scope or Shared Data").

2. **Reservation of Data Release**. ISBE reserves the right to refuse any data request involving individually identifiable information data or school/program/institution/district level data. However, nothing herein shall prohibit individuals or entities from releasing data pertaining to themselves or their own school, program, institution or district.

3. **Comment and Approval Period**. ISBE shall be provided for its review, any and all research and other reports produced using its data. The Recipient will provide ISBE with one electronic and at least one paper copy of a final draft and all final versions of all approved

reports to be released, along with other documents associated with any of the Specifications, as set forth in Exhibit A. ISBE expressly reserves the right to review, comment, and approve any use of the data shared or collected pursuant to this Agreement before its public release. The period of such review will be 60 days from ISBE's receipt of the material to be publicly released for all Illinois specific research and reports and 30 days from ISBE's receipt of the material to be publicly released for all other research and reports. For all Illinois specific research and reports for which the review period shall be 60 days, after the review period has lapsed, if ISBE has not approved the use in writing, the Recipient may not release the material publicly with ISBE's data included, but shall remove all ISBE data from the proposed publication and associated research analysis and provide a copy of the proposed publication, sans ISBE data, to ISBE prior to final publication. For all other research and reports for which the review period shall be 30 days, after the review period has lapsed, if ISBE has not approved the use in writing, the Recipient may release the material publicly with ISBE's data included, but shall remove all reference to ISBE data and Illinois from the proposed publication and associated research analysis and provide a copy of the proposed publication, sans reference to ISBE data and Illinois, to ISBE prior to final publication. ISBE reserves the right to demand, and, if requested, Recipient shall include in any material to be publicly released that includes the use of ISBE data, a rejoinder to be provided in writing by ISBE.

4. **Public Announcements**. All media releases and public announcements by either Party relating to this Agreement, the Research Project, or the Specifications shall be coordinated with and approved in writing by ISBE.

5. **Authorized Representatives**. The following persons are authorized to approve an amendment to this Agreement on behalf of the Parties (each is an "Authorized Representative;" collectively, the "Approval Representatives"):

   For ISBE: The State Superintendent of Education, or designee, with form approved by the General Counsel, or designee(s);

   For Recipient: Sally O'Neil, Manager, Industrial Contracts, Stanford University.

6. **Notices.** All notices or other correspondence required to be given pursuant to this Agreement shall be sent by mail or delivered to the Parties' Approval Representatives at the following addresses:

For ISBE:

General Counsel
Illinois State Board of Education
100 North First Street
Springfield, Illinois 62777
Fax: (217) 524-3911

For Recipient:

Margaret E. Raymond
Director of CREDO
434 Galvez Mall
Stanford University
Stanford, CA 94305-6010
Email: macke@stanford.edu
Fax: (650) 723-1687

In case of an emergency or when immediate assistance is needed:

The person to contact on behalf of ISBE is:

Brandon Williams
Telephone (217) 782-4824
bwilliam@isbe.net

The person to contact on behalf of Recipient is:

Margaret E. Raymond
Telephone: (650) 725-3431
Email: macke@stanford.edu

7. **Entirety.** This Agreement, together with the Exhibits attached hereto, constitutes the entire Agreement among the Parties with respect to the subject matter hereof and supersedes any other negotiations, agreements, or communications, whether written or oral, that have been made by any Party.

8. **Severability.** If any provision of this Agreement shall be held invalid, illegal, or unenforceable, such provision shall be deemed deleted from this Agreement and replaced by a valid and enforceable provision which so far as possible achieves the Parties' intent in agreeing to the original provision. The remaining provisions of this Agreement shall continue in full force and effect.

9. **Governing Law.** This Agreement shall be governed by and construed in accordance with the laws of the State of Illinois. Any claim against the State or a State agency arising out of this Agreement must be filed exclusively with the Illinois Court of Claims (705 ILCS 505/1) when said claim is within the jurisdiction of the Court of Claims.

10. **Records**. Books and records, including information stored in databases or other computer systems, shall be maintained by Recipient for a period of three (3) years from the later of the Term of this Agreement or the Completion Date and by any subcontractor for a period of three (3) years from the later of the date of the final payment under the subcontract or completion of the subcontract. The three (3) year period shall be extended for the duration of any audit in progress during the Term. Books and records required to be maintained under this section shall be available for review or audit by representatives of ISBE, the Auditor General, and other governmental entities with monitoring authority upon reasonable notice and during normal business hours. Recipient and its employees, contractors, subcontractors and agents shall cooperate fully with any such audit. Recipient and its employees, contractors, subcontractors and agents shall not impose a charge for audit or examination of Recipient's or Recipient's contractor's, or subcontractor's books and records.

11. **Hold Harmless**. To the fullest extent allowed by Illinois law, the Recipient agrees to defend, indemnify and hold harmless ISBE, its officers, staff, employees, and agents against any and all claims, suits, damages and causes of action arising out of or in any way related to the activities to be carried out pursuant to the obligations of this Agreement, including but not limited to, the use or disclosure by Recipient, its employees, contractors or agents, of any information received from or through ISBE pursuant to the terms of this Agreement.

12. **Injunctive Relief**. Recipient agrees that an impending or existing violation of any provision of this Agreement would cause ISBE irreparable injury for which it would have no adequate remedy at law and that ISBE shall be entitled to seek immediate injunctive relief prohibiting such violation, in addition to any other rights and remedies available to it.

13. **Authority to Execute.** Each Party represents and warrants to the other Party that this Agreement has been duly authorized, executed and delivered by and on behalf of each such Party and constitutes the legal, valid and binding agreement of said Party.

14. **Counterparts.** This Agreement may be executed in several counterparts, each of which shall be an original and all of which shall constitute one and the same instrument. Facsimiles of signatures shall constitute acceptable, binding signatures for purposes of this Agreement.

15. **Recitals and Exhibits.** The recitals in Section I and the following Exhibits are hereby incorporated by reference and expressly made a part of this Agreement.

EXHIBIT A – SPECIFICATIONS FOR SHARED DATA AND USE (INCLUDING ATTACHMENTS 1, 2, 3 AND 4)

EXHIBIT B – FORM FOR AMENDMENT OF RESEARCH SCOPE OR SHARED DATA

EXHIBIT C – FORM FOR AMENDMENT OF EXHIBIT A, ATTACHMENT 1

EXHIBIT D – FORM FOR AMENDMENT OF EXHIBIT A, ATTACHMENT 2

EXHIBIT E – DATA SECURITY PLAN (INCLUDING ATTACHMENT 1)

IN WITNESS WHEREOF, the Parties have executed this Agreement on the dates set forth below.

For the **Illinois State Board of Education:**

[redacted] 4/29/13
Signature Date
Division Administrator,

Illinois State Board of Education

Peter Godard
Print Name

[redacted] 4/29/13
Signature Date
Chief Performance Officer
Illinois State Board of Education

Peter Godard
Print Name

[redacted] 5/2/13
Signature Date
General Counsel
Illinois State Board of Education

Nicki Bazer
Print Name

[redacted] 5/8/13
Signature Date
State Superintendent of Education
Illinois State Board of Education

Christopher A. Koch
Print Name

The Board of Trustees of The Leland Stanford Junior University

[redacted] April 17, 2013
Signature Date
Sally O'Neil
Manager, Industrial Contracts
Stanford University

Peter Godard
Illinois State Board of Education
Data Request Review Board

Dear Mr. Godard,

I write on behalf of CREDO at Stanford University to request a Data Share and Use Agreement with the Illinois State Board of Education. This letter and the accompanying documents provide the required documentation that the Data Request Review Board needs to ascertain if the proposed project meets your guidelines for approval.

I will be serving as the contact person for this request, although the formal requestor will be John Raisian, Ph.D., the Director of the Hoover Institution of Stanford University. The Principal Investigator for the proposed project is Eric A. Hanushek, the Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University. Our contact details are:

Margaret Raymond, Director
CREDO
Stanford University
434 Galvez Mall
Stanford CA 94305
(650) 725-3431

Eric A. Hanushek
Hoover Institution
Stanford University
434 Galvez Mall
Stanford CA 94305
(650) 736-0942

The name of the proposed Project is: The Effectiveness of Charter Schools in Illinois and the National Charter School Study II (in which Illinois will be a one of 28 participating states.) The study will compare the academic progress of students enrolled in Illinois charter schools with the outcomes those students would have received if they had attended the traditional public schools instead. The study will assess performance for the sector as a whole, and then provide more detailed analyses of charter school impacts by geography, student backgrounds, participation in special support programs and years of persisting enrollment.

The project is both timely and pertinent to the mission of ISBE, as it will for the first time provide ISBE with detailed analysis of the performance of Illinois charter schools and, through the National Charter School Study, will deliver extensive comparative analysis from the other participating states so ISBE can benchmark the performance across the state with other states. Since charter schools comprise an important part of the overall school improvement strategy in Illinois, the analysis will be useful. In addition, with the advent of the Illinois State charter School Board, the analysis will serve as a foundation for future performance assessments.

Please do not hesitate to be in touch if I can answer any questions or provide you more information.

Best,

[redacted]

Margaret Raymond, Ph.D.
Director

credo
CENTER FOR RESEARCH ON EDUCATION OUTCOMES

Research Proposal
The National Charter School Study and Emerging Education Policies

a. Topic of the Research Project

The Center for Research on Education Outcomes (CREDO), a nonpartisan policy and program evaluation group at Stanford University, is currently conducting research to learn more about the effectiveness of charter schools and other public schools. One of our aims is to evaluate the impact of charter school attendance on student academic progress. This area of research becomes increasingly important as charter schools enroll greater shares of public school students and as school reform policies begin to blur the boundaries of charter and traditional school models. A second aim is to study a number of emerging policy issues that affect both the traditional public school and charter school sectors.

b. Purpose of the Research Project

CREDO has developed an ambitious research agenda of topics that are likely to be of considerable interest to policy makers for the next five years. While the bulk of CREDO's most visible work has focused on analysis of charter school effectiveness at the state and national levels, our prior work has also positioned CREDO to undertake analysis of other changes to public education. CREDO seeks to include the experience of Illinois in the upcoming range of studies.

CREDO is best known for its 2009 report on charter school impacts on student learning, *Multiple Choice: Charter School Performance in 16 States.* Since the release of the study in 2009, CREDO has conducted a number of other analyses of charter school effectiveness in Pennsylvania, New York City and Indiana. With additional years of student-level demographics and performance data for the original 16 states and the addition of another 12 states to our research collaborative, we are positioned to provide new and timely analyses that have immediate relevance in the policy landscape.

TEL: (650) 725-3431 FAX: (650) 723-1687
CREDO, Stanford University, Stanford CA, 94305-6010
credo.stanford.edu

c. Primary research questions to be answered by the Project

For each state and for the joint set of datasets, CREDO will conduct statistical analyses about the impacts of charter school enrollment – or participation in other education initiatives – on the learning growth of students. The analytic approach will differ for each of the questions the proposal seeks to address. The analysis is intended to create a national effect and then individual state effects on every question. Accordingly, each is discussed separately below.

1) How do charter school students compare to students who remain at traditional public schools? Do charter schools attract a representative subset of students or is the enrollment more limited?

There are two ways to examine this question: one is a simple descriptive profile of the charter and TPS student populations. This approach can be aggregated in a variety of ways: all charter students in a state compared with all TPS students, all charter school students compared with the populations in their feeder schools, all charter students against a TPS population weighted to mirror the composition of the charter schools, etc. Each aggregation would obviously employ all the available student records.

The second approach is to examine whether the students who leave traditional public schools to attend charters are systematically different from the TPS peers they leave behind. This question goes to the root of the "cream-skimming" allegations and arguments that charters are not serving their "fair share" of special needs students as well as to the counter argument charter supporters sometimes advance that they get the worst students. With access to student-level data, this question can be addressed in a rigorous way.

The analysis will test whether statistically significant differences exist in the academic achievement or recent growth (where supported by sufficient data) for students who leave their former TPS schools. The analytic instrument will be t-tests of means (with uneven samples) at varying levels of aggregation. This approach builds on the prior analysis of Zimmer et al (2009) that described the different starting scores of charter students at various levels of disaggregation. (Reference 1)

2) How do charter schools affect student outcomes?

Because the data is paired, and compared using the same state achievement tests for the same grade for the same year(s), any bias that exists in the alignment of achievement tests affects each member of each pair equally. This means that test score results can provide a relative basis of comparison within the pair, which is the focus of the analysis proposed here.

We propose to use longitudinal growth in standardized test scores as the measure of student learning. By differencing the standardized scores in successive years, an incremental measure of progress can be obtained. Because students are followed longitudinally, their background characteristics remain largely constant, thereby producing an unclouded measure of learning gains. These measures of standardized test score growth, referred to a z-score growth, form the outcome of interest in many of the questions about charter school impacts that the proposed study will seek to answer:

Charter School Impacts on Student Learning: The first concerns the rate of learning progress in charter school students compared to their virtual twins. The magnitudes of z-score growth for each member of a charter – TPS pair will be compared, $Z_{tC} - Z_{t-1C}$ vs. $Z_{tT} - Z_{t-1T}$ using econometric modeling. The model will provide statistical controls for personal and schooling factors as well as eligibility for program participation based on poverty, language or special education needs. The control variables will illuminate whether overall performance of the samples varies systematically by anything other than school type.

The VCR methodology permits this study to advance the insights of charter school impacts for students whose entire education (to date) has been in charter schools. As noted in the recent RAND study, most charter school studies focus on students who transfer from a TPS to a charter school (known as "switchers") to ensure that they share a common background. (Reference 1) Until now, there has been no reliable means of selecting control students for the students who start out in charter schools. However, the VCR method can select on all known characteristics, including third grade scores, for all students that match the factors of a student with earlier enrollment in a charter school. Based on our own research in other communities, we would expect the third grade average baseline scores of

"charter only" students to be higher than for switchers. We will compare as well differences in their growth performance.

As noted in prior studies even with an extensive array of controls for observable attributes, the possibility remains that students will differ in other unobserved ways. If student enrollment in charter schools is endogenously influenced – that is, that the variation in their performance is not random – then estimates of the influence of charter schools on their learning gains could be biased. To address the concern, the models will be estimated two ways: with and without student fixed effects. While the CREDO research team questions the validity of the selection argument (see above), it will still be advisable to examine the results with the fixed effect correction – if the results mirror the models without them, then the finding will be useful for furthering other research efforts.

The results of the econometric models will be refined by looking at interaction effects. In these models, the learning gains in charter schools will be studied to see if they differ by student-level characteristics, such as race, gender, grade, baseline starting scores or program eligibility. The aim of these models is to see if, beyond the average charter effect obtained in the model described above, there are subsets of students for which charter school attendance produces significantly stronger or weaker results. The insights from this stage of research could be beneficial to funders, authorizers and the Illinois Office of Charter Schools.

Impact of Charter School Characteristics on Performance: To extend the insights about charter school effectiveness, additional analyses will be performed. The additional analyses will focus on operational details of the schools that might affect their performance relative to their TPS counterparts. The factors that will be studied are: age of the school, the student persistence rate, membership in a CMO/EMO organization and conformance to feeder school grade spans.

Competitive Impacts of Charter Schools: The student level dataset will be used to conduct a third analysis. The purpose of the third study is to see if attributes of charter schools and their student learning have an impact on the extent or the pace of learning gains in traditional public schools or other charter schools. Market factors such as share, market

power and market segmentation will be calculated for each charter school. There are several possible ways to define the competitive space, and we expect to select a number for exploration. For example, the most likely group to respond to charter school performance would be the schools that lose students to them, so the feeder pool of traditional public schools would be one sphere to study. Another would be other charter schools with similar grade spans. In addition, market impacts will be examined over time to see if effects are lagged or perhaps change with additional saturation of the market.

3) **Other Hypotheses:** As noted above, we are interested in examining emerging education policies that affect the entire public school sector – both traditional and charter schools. We want to work with our partner states to develop the hypotheses that would be of most interest and value to you that also intersect with CREDO's capacity and research interests. For example, we have done some initial analyses in two areas, and we are interested in expanding the scope of these analyses to answer the following questions:
   - How do online schools affect student outcomes? Are differences in performance among online schools related to whether they are charter or TPS? Are performance differences related to the differing mix of online and in-class offerings?
   - Do turnarounds of failing schools improve student performance at those schools?

The hypotheses above are just two possibilities. We look forward to hearing about the questions that Illinois would like answered.

As in Hypothesis #2, we intend to use paired observations of students who do and do not participate in the new program to determine relative effectiveness. In the case of whole-school programs, we will first employ a matching strategy to identify demographically similar schools that are not participating in the program. Students from these similar schools will comprise the pool from which controls will be drawn using the control selection method described above.

Longitudinal growth in standardized test scores will be the primary measure of student learning in these analyses. Timely high school graduation is an

additional outcome of interest. In both cases, students will be the unit of analysis, and the analysis will include statistical controls for factors such as demographics, grade repeater status, and eligibility for meal programs, special education services and language services.

d. Explanation of how this Research Project is educationally significant

There are three separate strands of research that our work draws upon:

1) The correlates of academic achievement and growth,
2) The assessment of charter schools as an educational improvement policy, and
3) The impacts of charter schools on district practices and outcomes.

The first area, the correlates of academic achievement and growth, has a rich research history stretching back as far as Brown v. Board of Education. (Reference 2) Social justice and state and national economic prosperity are hindered when the education sector is differentially able to help students grow academically. (References 3, 4) Since part of the motivation for permitting charter schools and other education initiatives rests on the hope that different education models will create a narrowing of the achievement gap, this study design provides a rigorous control for differences in racial and social backgrounds to see if each initiative produces equivalent results for all student types. This is consistent with Goal 1 of ISBE: Every student will demonstrate academic achievement and be prepared for success after high school. Our work is statistically as rigorous as the best of the existing literature, such as Hanushek and Rivkin, Carolyn Hoxby, or the Center for Reinventing Public Education in that it examines a longitudinal sample of students. While not a Random Assignment Study (which is not appropriate because it would force placement of students that may differ from parental choice), it utilizes sample selection techniques that come as close as possible to RA designs using a quasi-experimental design. But because our data includes multiple states and many more schools than earlier student-level analysis, our design will both update and expand the current state of knowledge about student outcomes.

The second area, the assessment of charter schools as an educational improvement policy, seeks to examine the consequences of various approaches to school reform. In such studies, the student-level data is used to estimate program effects which are then compared to alternate policy results. For example, holding all other influences constant, one might wonder if California's charter school policy is as effective as say, Arizona or Texas, or whether the kind of impact charters have in large urban districts compares favorably with early childhood education policies in those same districts. In the area of charter schools, there is very little program effect information available, though it would bear strong resemblance to studies in the area of vouchers, literacy, or alternative certification. (References 5, 6, 7) The CREDO study will be able to provide comparative effects by state, by independent charter vs. network charter as well as by specific characteristics of the schools in our sample (teacher profile, age of charter, urbanicity, etc.).

The third area of work that informs our design is the study of markets in transition. Specifically, this study will help illuminate the stimulation or competitive effects of charter schools on the districts in their immediate area. The theory is that through the introduction of competitive alternatives to Traditional Public Schools, schools and districts will be forced to respond in order to maintain their student enrollment. Response could take a number of different forms, but the aim of advocates of the "charters as competition" perspective expect that improved academic results will be a prime outcome of district efforts. There is a significant amount of economics literature regarding the introduction of competition into formerly monopoly markets (Reference 8), but in education, most of the prior work has looked at vouchers or private school impacts. (References 9, 10) The current study design will go part of the way to support this line of research: Our methodology will compare academic growth of students in charters and equivalent students in schools those students would attend if not for the charter. That relative measure of performance can then be used as an independent variable in a future analysis that examines two questions:

1. Do districts that face charter competition improve at faster rates than districts that do not face this competition?

2. Do districts respond differently to different levels of performance in their charter competitors?

e. Rationale for the Research Project

CREDO is committed to providing analysis that is relevant to policy makers, and this project is consistent with our previous work. The aim of this project is to identify the factors that can be causally associated with improvements in charter school outcomes in a systematic and significant way. This information can in turn be used by policy makers and local or state charter school authorizers to improve charter school quality

f. Methodology of the Research Project

For any study, the research design must set up the conditions so as to provide the fairest and most impartial hypotheses testing. Design is a matter of balancing data limitations, sampling options, bias control and effort/efficiency considerations. CREDO proposes a quasi-experimental design employing an innovative control sample selection technique. A description of our proposed approach and support for it are presented in the Sampling Design section below.

g. Data requested

CREDO is requesting student-level assessment and enrollment files from the 2004-05 through the 2011-12 school years for students in grades K-12 attending all public schools in Illinois. The list of known ISBE data tables and elements is included in Attachment 3 of the Data Share and Use Agreement. Generally, however, the request is for each student's full record, including:

- ☐ Student identification number from the state or district, or scrambled identifier that can be linked across years

- ☐ State achievement test scaled and raw scores (Reading/ELA and Math for all students plus end of course exam scores for high school students) for all available years starting with the 2004-2005 school year.

- ☐ Performance levels, or achievement or proficiency categories, usually given in the following categories (for each subject): below basic, basic, proficient, advanced and additional categories such as far below basic or above advanced, if applicable

- ☐ District, school and campus identification numbers for each school the student attended on the testing date each year

- ☐ Demographic/Student Information:
  - Race/Ethnicity
  - Gender
  - Lunch Status
  - Special Education Status
  - English Proficiency
  - Grade Level
  - Zip code of residency
  - Full student enrollment files for each year we are given, including date entered, date exited, days of possible attendance, days attended and school attended (ID & name of school).

- ☐ Additionally, for high school students:
  - Graduation flag
  - Course completion records

In addition to the student-level data described above, the following information is also requested:

- ☐ List of charter schools for each year by district, school and campus ID.

- ☐ List of juvenile detention centers by district, school and campus ID.

- ☐ Grade level means and standard deviations for the state reading and math tests for each year (aka technical report).

- ☐ Conditional standard error of measurement tables for each grade, year and subject test, i.e., standard error for each individual scaled or raw score.

- ☐ Cut scores for proficiency bands.

- ☐ Percentages of students in each school (by district and school ID): free/reduced-priced lunch, English Language Learner programs, and Special Education programs.

- ☐ Unique school identifier (or unique district, school and campus ID combination) that is linkable to federally published school data for Illinois.

h. Explanation why data available in the public domain is not sufficient for the Research Project

To be able to provide statistically sound answers to the research questions, the selection of appropriate controls is paramount. This requires making the selection at the individual student level and following students longitudinally over time to measure academic growth.

i. Sampling Design

Random assignment studies are of limited value in the study of charter schools and many other performance-driven education initiatives.

Generally, random assignment of subjects into study and control groups provides the most pristine design in terms of minimizing potential sources of bias. In other words, random assignment maximizes internal validity of the study. But in order for the random assignment to be credible, the subjects must be uniform in the factors that may influence their likelihood of selection. And with narrow exceptions, studies of education initiatives cannot meet the requirement of uniformity.

In the case of charter schools, the fact that charter school parents reveal their decisions about school placement with charter school enrollment raises the possibility that they are somehow differently motivated from other

parents. If differences in motivation exist, then the question follows whether the motivation to actively choose charter schools is somehow associated with other non-observed characteristics that might affect the performance of their children. (The typical argument posits that these parents "care about" their children's education more than other parents, and so are likely to engage in other pro-education behavior to their child's benefit.)

Hoxby's waiting list random assignment study in NYC illustrates the limited utility of the approach in the charter environment. Using the lotteries registration lists for New York City charter schools that were oversubscribed, she followed the academic outcomes of the lottery losers as well as winners to create her study and control groups. Profile analysis of the two samples showed little differences between the two groups, fulfilling the desired purpose of randomizing. (Reference 11) The problem with the study, however, is that it did not demonstrate that NYC charters with waiting lists are representative of the population of NYC charters. In fact, analysis in Ohio and Florida shows that they are more mature and report higher achievement levels than average. These considerations raise serious questions about the external validity of the approach.

Given the limited value of random assignment in this field of study, CREDO has adopted a rigorous quasi-experimental design, which provides a strong degree of control for potential sources of bias and better matches the conditions of the charter school student population.

CREDO Uses a Unique Sampling Method

Once a decision to pursue a quasi-experimental design is made, the choice of appropriate control cases becomes paramount.

With charter school research, the problem with appropriate controls is well known. Early research in the field focused largely on comparisons at the school level. This was due in part to a mis-framing of the research question: while the policy question focused on schools and their effectiveness, the school was an incorrect unit of analysis. Defenders of charter schools cited differences between charter schools and traditional schools based on the age of the school, length of student enrollment, potentially different student populations and mission differences. Any of these differences could create

bias in the sampling of controls, if true. The problem was magnified by a lack of accessible student-level data and the absence of unique student identifiers that permit students to be tracked within a state over time, even if they change schools.

Improvements in state data systems have opened the frontier of student-level analysis of learning differences by type of school. The ability to frame an analysis in terms of student learning (as well as achievement) properly frames the question of school effectiveness or impact. Having student-level data however brings all the aforementioned questions of equivalent research samples into sharp relief.

A number of responses have been used in other studies. Simple matching of students on observable characteristics (Matched pairs) is regarded as insufficient for the same reasons cited above in favor of random assignment: there may be structural but unmeasured differences among students with similar measureable attributes – some may have more concerned families, others may have developed aversion to school and so on.

Propensity matching has also been employed. Propensity matching attempts to instill some of the strength of experimental designs into quasi-experimental studies. This approach selects control matches based on an array of match factors that are used to create a probability of inclusion in the experimental or control groups. The probabilities are then used as match factors or covariates in analysis. (References 12, 13) The approach is considered by many to address some of the limitation of simple matching, especially when the overlap of the two potential samples is small. For example, if only the worst performing students are permitted to enroll in charter schools (as is the case in some states), drawing controls from the full distribution of traditional school students would create bias in the estimation of program effects for charter schools.

All forms of propensity matching rest on the assumption that, in addition to different treatment paths (attending charter schools or traditional schools), there are non-observable differences between experimental and control subjects that are associated with their path. In the current context, the assumption is that parents who enroll their student in a charter school exercise their right to choose but that parents of students in traditional

schools do not. It is difficult to support such an assumption; the evidence from open enrollment districts strongly suggests that traditional school parents make active choices regularly. It is reasonable for parents to weigh the alternatives and decide to have their child remain in the traditional school. It is reasonable for parents to want a new charter school to mature and demonstrate its quality before considering it for their child. The student or parent may assess factors such as competitive athletic opportunities (that could lead to college scholarships) as a requirement in their school of choice, which would eliminate many charter schools from consideration. The point is that the absence of a change in schools does not assure the absence of choice.

For this reason, CREDO pursued another approach to sample selection.

Case Selection: Charter schools are identified in the statewide dataset either through an indicator variable included in the dataset from Illinois or through a merge with an external dataset. The students with test records associated with the subset of charter schools for a given year become the case students for that test year cohort. The process is repeated for subsequent test years. Records are linked across years if multiple records exist.

This case selection approach combines both school switchers (those that moved from Traditional Public Schools to Charters) and Charter-only students. Each is tagged for bias control purposes later in the analysis phase. In this way, we balance the potential downward bias of only sampling school switchers noted by Ballou, Teasley and Zeidner (2007) in their analysis of Idaho charter schools. (Reference 14)

Control Selection: We use a Virtual Control Record (VCR) matched to each charter school student. The approach was developed by CREDO as an accommodation to the re-disclosure restrictions of FERPA, which even under the new guidelines of 2008 are considered in force by many states. In 2007, FERPA issued CREDO a Memorandum of Determination that verified that the VCR methodology was in compliance with the FERPA legislation and regulations.

The VCR approach builds on work by Abadie, Diamond and Hainmueller at Harvard University and the Northwest Evaluation Association (NWEA). Both groups have explored the use of synthetic control groups in comparative research. Both sets of researchers create an aggregate control record by drawing on the available records that match (in varying degrees of precision) with the case record. In Abadie et al, the available control records are weighted by their goodness of fit to the case on every available match factor. NWEA sets the condition of a successful match in advance so that only "true" or "near true" records are selected. (Reference 15)

CREDO's methodology follows more closely to NWEA than Abadie et al. We identify all the TPS schools that have at least 1 student transfer to the charter school ("Feeder schools"). Using the records of the students in those schools in the year prior to the test year of interest, CREDO selects all the available records that match each charter school student. Match factors include:

- Grade
- Gender
- Race/Ethnicity
- Free or Reduced Price Lunch Status
- English Language Learner Status
- Special Education Status
- Grade Repeater
- Prior test score on state achievement tests

The scores from the test year of interest are then averaged and a Virtual Control Record is produced. That record is completely masked, because there is no record of where the students who make up the VCR came from. The VCR produces a score for the test year of interest that corresponds to the expected value results of propensity matching.

VCRs are re-examined for every test period to ensure that the conditions of match still apply – namely that the students included in the VCR record are still enrolled in traditional public schools and have not left the state. Where the conditions are violated, the VCR is reconstructed to delete the disqualified student records.

The result of the case and control sampling strategies described above are matched pairs which are followed over as many years as are supported by available data. In addition to being used to study charter schools, this same method will be utilized to study emerging education policies and initiatives.

The VCR method has been favorably compared with a more widely accepted quasi-experimental method, student fixed effects. In a 2012 journal article written by CREDO staff, results from the two methods were similar when the charter student samples were identical. However, the VCR method is able to include a much larger proportion of charter school students than the student fixed effects method, which requires students to switch between the charter and traditional public school sectors to be included. (Reference 16) Another research group, Mathematica, recently published a report that compared results from the VCR method with those obtained from an experimental design and found statistically similar results. (Reference 17) These recent methods comparisons make us confident that the VCR method yields statistically sound and generalizable results.

Dataset Development: A dataset of all matched records will be compiled for Illinois to prepare it for analysis. This dataset will conform to those created for the other states that have agreed to share their student-level data for use in our research.

Respondent Burden: Our research involves administrative datasets only; as such, we will not interact directly with any students, parents, teachers, or school administrators in the course of this research.

j. Timeline for the Research Project

Date of Receipt through July 31, 2013: Process data, perform analysis, prepare reports.

August 1, 2013 – January 31, 2014: Examine researcher critiques, re-analyze data if necessary and respond.

Februrary 1, 2014 – February 31, 2016: Hold data for potential related scholarly journal publications.

March 1, 2016: Data destruction.

References:

1. Zimmer, Ron, Brian Gill, Kevin Booker, Stephane Lavertu, Tim R. Sass, and John Witte. *Are Charter Schools Making a Difference?: A Study of Student Outcomes in Eight States - 2009.* Santa Monica, CA: RAND Corporation, 2009.

2. Rivkin, Steven and Finis Welch. "Has School Desegregation Improved Academic and Economic Outcomes for Blacks?" in Eric Hanushek and Finis Welch (Eds.) *Handbook of the Economics of Education*, Elsevier Press, 2006.

2. Jencks, Christopher and Merideth Phillips (Eds.) *The Black White Test Score Gap*. Washington DC: Brookings Press, 1998.

3. Collins, William J. and Robert A. Margo. "Historical Perspectives on Race Differences in Schooling in the United States" in Eric Hanushek and Finis Welch (Eds.) *Handbook of the Economics of Education*, Elsevier Press, 2006.

4. Molnar, Alex. *Educational Vouchers: A Review of the Research.* Center for Education Research, Analysis, and Innovation, School of Education, University of Wisconsin-Milwaukee, 1999.

5. McEwan, Patrick. "The Potential Impact of Large-Scale Voucher Programs" *Review of Educational Research,* Vol. 70, No. 2.

6. Legler, Ray. *Alternative Certification: A Review of Theory and Research.* Learning Point Associates, 2002.

7. O'Connell, Donald W. Political Science Quarterly, Vol. 64, No. 1 (March, 1949), pp. 122-125. See also Levin, Harvey. "Competition Among Mass Media and the Public Interest." *The Public Opinion Quarterly,* Vol. 18, No. 1 (Spring, 1954), pp. 62-79

8. Moe, Terry M. Schools, *Vouchers and the American Public.* Washington, DC: Brookings Institution Press, 2001.

9. Nechyba, T. J. *Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects, and Vouchers,* NBER Working Paper No. 5642, June 1996.

10. Hoxby, Caroline M. and Sonali Murarka. *New York City's Charter Schools Overall Report* Cambridge, MA: New York City Charter Schools Evaluation Project, 2007.

11. Haviland, Amelia M. and Daniel S. Nagin. "Using Group-Based Trajectory Modeling in Conjunction with Propensity Scores to Improve Balance." *Journal of Experimental Criminology,* 2007, no 3, pp. 65-82.

12. Dehejia, Rajeev H. and Sadek Wahba. "Propensity Score Matching Methods for Nonexperimental Causal Studies." *The Review of Economics and Statistics,* February 2002, Vol. 84, No. 1, pp. 151-161.

13. Ballou, Dale; Bettie Teasley and Tim Zeidner. "Charter Schools in Idaho" in Mark Berends, Matthew G. Springer and Herbert J. Walbert, (Eds.), *Charter School Outcomes*, New York: L. Erlbaum Associates, 2007, pp. 221-241.

14. Abadie, Alberto, Alexis Diamond and Jens Hainmueller. *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program.* NBER Working Papers #12831, September, 2007.

15. Davis, Devora H. and Margaret E. Raymond. "Choices for Studying Choice: Assessing Charter School Effectiveness using Two Quasi-Experimental Methods." *Economics of Education Review*, 2012, Vol. 31, No. 12, pp. 225-236.

16. Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Kopa, and Philip Gleason. *Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates.* NCEE 2012-4019 U.S. Department of Education, 2012.

Eric A. Hanushek

Paul and Jean Hanna Senior Fellow

Stanford University

Contact

Hoover Institution
Stanford University
Stanford, CA 94305-6010
(650) 736-0942

fax: (650) 723-1687
e-mail: hanushek@stanford.edu
home page: http://www.hanushek.net

Education

1965 B.S. (Distinguished Graduate) U.S. Air Force Academy

1968 Ph.D. (Economics) Massachusetts Institute of Technology
Thesis: "The Education of Negroes and Whites"

Learned Societies

American Economic Association
American Education Finance Association
(Board of Directors, 2006-2009)
Association for Public Policy Analysis and Management
(Policy Council, 1981-85; vice president, 1986-87; president, 1988-89)
Econometric Society
Society of Labor Economists
(Fellow, 2006-)
American Educational Research Association
(Fellow, 2008-)
International Institute of Public Finance

Honors

Fellow, International Academy of Education, 1997
(Board of Directors, 2002- 2008)
Fellow, Society of Labor Economists, 2006
Member, National Academy of Education, 2006
Fellow, American Educational Research Association, 2008
Fordham Prize for Excellence in Education (distinguished scholarship), 2004

Military Service

U.S. Air Force, 1965-74

Academic Experience

2000-	Paul and Jean Hanna Senior Fellow, Hoover Institution, Stanford University Professor (by courtesy) of Education (2001-) Senior Fellow (by courtesy), Stanford Center for International Development [SCID], Stanford Institute for Economic Policy Research [SIEPR], (2003-) Professor (by courtesy) of Economics (2004-)
2003-	Chairman, Executive Board, Texas Schools Project, University of Texas at Dallas
2000-	Senior Research Fellow, Cecil and Ida Green Center for the Study of Science and Society, University of Texas at Dallas
1999-	Member, Koret Task Force on K-12 Education, Hoover Institution, Stanford University
1995-	Research Associate, National Bureau of Economic Research
2006-	Research Professor, Ifo Institute for Economic Research (University of Munich)
2008-	Member, Management Team, Center for Analysis of Longitudinal Data in Education Research (CALDER)
2006-	Area coordinator, Economics of Education, CESifo Research Network
2006-2008	Member, Scientific Advisory Council, Ifo Institute for Economic Research
1978-2000	Professor of Economics and Political Science, University of Rochester Director, W. Allen Wallis Institute of Political Economy (1991-99) Professor of Public Policy (1992-2000) Senior Research Associate, Rochester Center for Economic Research (1984- ; Director, 1994-99) Chairman, Department of Economics (1982-87; 1988-90; 1991-93)
1999-2000	Distinguished Visiting Fellow, Hoover Institution, Stanford University
1994	Visiting Fellow, Australian National University
1975-78	Associate Professor, Department of Economics and Institution for Social and Policy Studies, Yale University
1974	Lecturer, Virginia Polytechnic Institute (Reston Campus)
1968-73	Associate Professor of Economics, U.S. Air Force Academy (Assistant Professor, 1969-71; Instructor, 1968-69)
1970-71	Research Associate, J.F. Kennedy School of Government, Harvard University

Government Experience

2011-	Commissioner, Equity and Excellence Commission, U.S. Department of Education
2004-2010	Member, Board of Directors, National Board for Education Sciences Chair, 2008- 2010
2008-2010	Member, Council of Economic Advisors, California
2007	Member, Governor's Commission for a College Ready Texas, Texas
2006	Chair, NCLB Growth Model Pilot Peer Review, U.S. Department of Education
2005-2008	Member, Governor's Committee on Education Excellence, California
2002-2006	Member, Independent Review Panel, National Assessment of Title I, U.S. Department of Education
2001-05	Member, NCES Finance Technical Review Committee, U.S. Department of Education
2002	Member, Advisory Council on Education Statistics, U.S. Department of Education
1994-98	Member, Board of Economic Advisors, New York State Assembly
1994-95	Member, Technical Panel on Trends and Issues in Retirement Savings, Advisory Council on Social Security
1987-95	Consultant, U.S. Department of Education
1986-89	Consultant, U.S. Commission on Civil Rights
1987-89	Chairman, Technical Advisory Panel, Congressional Budget Office
1985-87	Member, Panel of Economic Advisers, Congressional Budget Office
1983-85	Deputy Director, Congressional Budget Office
1974	Systems Analyst, Military Airlift Command, U.S. Air Force
1973-74	Senior Economist, Cost of Living Council
1971-72	Senior Staff Economist, Council of Economic Advisers

Other Experience

2011-	Director, CollegeSpring (formerly SEE College Prep)
2010-	Member, Education Reform Advisory Group, George W. Bush Institute
2010-	Director, GreatSchools
2007-	Member, Review Board for Broad Prize for Urban Education
2000-01	Member, Committee on Scientific Principles of Education Research, National Academy of Sciences/National Research Council
2000	Member, Historic Preservation Commission, Town of Brighton, NY
1998-2001	Member, Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs, National Academy of Sciences/National Research Council
1992-98	Member, Committee on National Statistics, National Academy of Sciences/National Research Council
1993-97	Chairman, Panel on Retirement Income Modeling, National Academy of

	Sciences/National Research Council
1990-94	Chairman, Panel on the Economics of Educational Reform (PEER)
1984-95	Consultant, The World Bank
1992	Chairman, Blue Ribbon Commission on Monroe County Finances, Monroe County, NY
1988-91	Chairman, Panel to Evaluate Microsimulation Models for Social Welfare Programs, National Academy of Sciences/National Research Council
1977-83	Consultant, Mathematica Policy Research
1976-78	Member, Mayor's Task Force on Education, New Haven, CT
1975-77	Senior Research Associate, Institute for Demographic and Economic Studies
1975-77	Consultant, Abt Associates
1972-74	Member, RFF-Academy for Contemporary Problems, Metropolitan Governance Research Committee
1969-73	Consultant, The Rand Corporation

Invited Lectures

Waino Pihl Lecture, Wayne State University, 2012

Distinguished Scholar Lecture, Martin School of Public Policy, University of Kentucky, 2010

Giblin Lecture, University of Tasmania, 2009

Hannah Lecture, Michigan State University, 2009

Gilbert Memorial Lecture, University of Rochester, 2008

Spencer Foundation Distinguished Lecture , Association for Public Policy Analysis and Management, 2005

Sweat Lecture, Georgia State University, 2005

Birger Lecture, Tufts University, 2005

Lee Hysan Lecture, Chinese University of Hong Kong, 2004

Askwith Lecture, Harvard University, 2003

Reilly Lecture, Louisiana State University, 2002

Mullen Lecture, University of Maryland, Baltimore County, 1999

Saks Memorial Lecture, Vanderbilt University, 1996

Editorial Activities

2007-	Associate Editor, *Journal of Human Capital*
2005-	Editorial Board, *Education Finance and Policy*
2004-	Co-editor, *Education Policy Series*, International Academy of Education/International Institute for Educational Planning, UNESCO
2000-	Editorial Board, *Education Next*
1982-	Editorial Board, *Economics of Education Review*

1978-	Advisory Editor, *Social Science Research*
2002-2008	Editorial Board, *Fundamentals of Educational Planning*, UNESCO
2003-2007	Associate Editor, *Economic Bulletin*
1995-2002	Associate Editor, *Review of Economics and Statistics*
1997- 2001	Editorial Board, *Educational Evaluation and Policy Analysis*
1994-2001	Editorial Board, *Journal of Policy Analysis and Management*
1994-96	Editorial Board, *Socio-Economic Planning Sciences*
1991-97	Associate Editor, *Regional Science and Urban Economics*
1990-95	Editorial Board, *Journal of Economic Education*
1992-95	Advisory Board, *American Journal of Education*
1990-94	Co-editor, *Journal of Human Resources*
1987-1989	Associate Editor, *Evaluation Review*

PUBLICATIONS

Books

Handbook of the Economics of Education, Volume 4 (co-editor with Stephen J. Machin and Ludger Woessmann). *Handbook of the Economics of Education, Vol. 4, Amsterdam: North Holland,* 2011, 708 pages.

Handbook of the Economics of Education, Volume 3 (co-editor with Stephen J. Machin and Ludger Woessmann). *Handbook of the Economics of Education, Vol. 3, Amsterdam: North Holland,* 2010, 616 pages.

Schoolhouses, Courthouses, and Statehouses: Solving the Funding-Achievement Puzzle in America's Public Schools (with Alfred A. Lindseth). *Princeton University Press,* 2009, 432 pages.

Handbook of the Economics of Education, Volume 2 (co-editor with Finis Welch). *Handbook of the Economics of Education, Vol. 2, Amsterdam: North Holland,* 2006, 742 pages.

Handbook of the Economics of Education, Volume 1 (co-editor with Finis Welch). *Handbook of the Economics of Education, Vol. 1, Amsterdam: North Holland,* 2006, 700 pages.

Courting Failure: How School Finance Lawsuits Exploit Judges' Good Intentions and Harm our Children (editor). *Stanford: Education Next Books,* 2006, 366 pages.

Institutional Models in Education: Legal Framework and Methodological Aspects for a New Approach to the Problem of School Governance (co-editor with Enrico Gori, Daniele Vidoni and Charles Glenn). *Nijmegen, Netherlands: Wolf Legal Publishers,* 2006, 243 pages.

The Economics of Schooling and School Quality - Volume II: Efficiency, Competition, and Policy (editor). *London: Edward Elgar Publishing Ltd.,* 2003.

The Economics of Schooling and School Quality - Volume I: Labor Markets, Distribution, and Growth (editor). *London: Edward Elgar Publishing Ltd.,* 2003, 976 pages.

Assessing Policies for Retirement Income: Needs for Data, Research, and Models (co-editor with Constance F. Citro). *Washington, DC: National Academy Press,* 1997, 256 pages.

Improving America's Schools: The Role of Incentives (co-editor with Dale W. Jorgenson). *Washington, DC: National Academy Press,* 1996, 280 pages.

Assessing Knowledge of Retirement Behavior (co-editor with Nancy L. Maritato). *Washington, DC: National Academy Press,* 1996, 288 pages.

Modern Political Economy: Old Topics, New Directions (co-editor with Jeffrey S. Banks). *New York: Cambridge University Press,* 1995, 283 pages.

Making Schools Work: Improving Performance and Controlling Costs. *Washington, DC: The Brookings Institution,* 1994, 200 pages.

Educação Rural: Lições do Edurural (with João Batista F. Gomes Neto, Ralph W. Harbison, and Raimundo Hélio Leite). *São Paulo: Editora da Universidade de São Paulo,* 1994, 236 pages.

Educational Performance of the Poor: Lessons from Rural Northeast Brazil (with Ralph W. Harbison). *New York: Oxford University Press,* 1992, 362 pages.

Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling - Volume II: Technical Papers (co-editor with Constance F. Citro). *Washington, DC: National Academy Press,* 1991, 368 pages.

Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling - Volume I: Review and Recommendations (co-editor with Constance F. Citro). *Washington, DC: National Academy Press,* 1991, 360 pages.

Statistical Methods for Social Scientists (with John E. Jackson). *New York: Academic Press,* 1977, 374 pages.

Education and Race: An Analysis of the Educational Production Process. *Lexington, MA: D.C. Heath,* 1972, 176 pages.

Articles

Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA (with Susanne Link and Ludger Woessmann). *Journal of Development Economics,* Forthcoming.

Household location and schools in metropolitan areas with heterogeneous suburbs: Tiebout, Alonso, and government policy (with Kuzey Yilmaz). *Journal of Public Economic Theory,* Forthcoming.

School Leaders Matter: Measuring the impact of effective principals (with Gregory F. Branch and Steven G. Rivkin). *Education Next,* 13(1), Winter 2013, pp. 62-69.

Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation (with Ludger Woessmann). *Journal of Economic Growth,* 17(4), December 2012, 267-321.

Schooling, Educational Achievement, and the Latin American Growth Puzzle (with Ludger Woessmann). *Journal of Development Economics,* 99(2), November 2012, pp. 497-512.

The Distribution of Teacher Quality and Implications for Policy (with Stephen G. Rivkin). *Annual Review of Economics,* 4, September 2012, pp. 131-157.

Is the U.S. catching up? International and state trends in student achievement (with Paul E. Peterson and Ludger Woessmann). *Education Next,* 12(4), Fall 2012, pp. 24-33.

The Role of International Assessments of Cognitive Skills in the Analysis of Growth and Development (with Ludger Woessmann), In Matthias von Davier, Eugenio Gonzalez, Irwin Kirsch, and Kentaro Yamamoto (Ed.). *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research,* Dordrecht, Netherlands: Springer, 2012, 47-65.

Achievement growth: International and state trends in student achievement (with Paul E. Peterson and Ludger Woessmann). *PEPG Report No. 12-03,* July 2012.

A flawed analysis of unrepresentative state achievement data. *Educational Policy,* 26(3), May 2012, 360-368.

Education quality and economic growth In Brendan Miniter (ed.). *The 4 percent solution: Unleashing the economic growth America needs,* New York: Crown Business, 2012, pp. 227-239.

Grinding the Antitesting Ax: More bias than evidence behind NRC panel's conclusions. *Education Next,* Spring 2012, pp. 49-55.

The Economic Benefit of Educational Reform in the European Union (with Ludger Woessmann). *CESifo Economic Studies,* 58(1), March 2012, pp. 73–109.

Urban education, location, and opportunity in the United States (with Kuzey Yilmaz). In Nancy Brooks, Kieran Donaghy, and Gerrit-Jan Knaap (ed.). *Oxford Handbook of Urban Economics and Planning,* Oxford: Oxford University Press, 2011, pp. 583-615.

Are U.S. Students Ready to Compete? (with Paul E. Peterson, Ludger Woessmann, and Carlos Xabel Lastra-Anadón). *Education Next,* 11(4), Fall 2011, pp. 51-59.

Globally Challenged: Are U.S. Students Ready to Compete? (with Paul E. Peterson, Ludger Woessmann, and Carlos Xabel Lastra-Anadón). *PEPG Report No. 11-03,* Cambridge, MA: Program on Education Policy and Governance, Harvard University, August 2011.

Private Schools and Residential Choices: Accessibility, Mobility, and Welfare (with Sinan Sarpça and Kuzey Yilmaz). *B.E. Journal of Economic Analysis & Policy (Contributions),* 11(1) article 44, 2011, pp. 1-32.

How Much Do Educational Outcomes Matter in OECD Countries? (with Ludger Woessmann). *Economic Policy,* 26(67), July 2011, pp. 427-491.

Valuing Teachers: How Much is a Good Teacher Worth? *Education Next,* 11(3), Summer 2011, pp. 40-45.

The Economic Value of Higher Teacher Quality. *Economics of Education Review,* 30(2), June 2011, pp. 466-479.

Teaching Math to the Talented: Which Countries - and States - are Producing High-Achieving Students? (with Paul E. Peterson and Ludger Woessmann) *Education Next,* Winter 2011, pp. 10-18.

Paying Teachers Appropriately. *The American Public School Teacher: Past, Present, and Future,* (Cambridge, MA: Harvard Education Press), 2011, pp. 109-118.

Sample Selectivity and the Validity of International Student Achievement Tests in Economic Research (with Ludger Woessmann). *Economics Letters,* 110(2), February 2011, pp. 79-82.

The Economics of International Differences in Educational Achievement (with Ludger Woessmann) in Eric A. Hanushek, Stephen Machin and Ludger Woessmann (ed.). *Handbook of the Economics of Education, Vol. 3, Amsterdam: North Holland,* 2011, pp. 89-200.

How well do we understand achievement gaps? *Focus,* 27(2), Winter 2010, pp. 5-12.

U.S. Math Performance in Global Perspective: How Well Does Each State Do at Producing High-Achieving Students? (with Paul E. Peterson and Ludger Woessmann) *PEPG Report No. 10–19,* Cambridge, MA: Program on Education Policy and Governance, Harvard University, November 2010.

The Difference is Teacher Quality in Karl Weber (ed.). *Waiting for "Superman": How We Can Save America's Failing Public Schools. New York: Public Affairs,* 2010, pp. 81-100.

An Effective Teacher in Every Classroom: A lofty goal, but how to do it (with Kati Haycock). *Education Next,* 10(3), Summer 2010, pp. 46-52.

The Quality and Distribution of Teachers under the No Child Left Behind Act (with Steven G. Rivkin). *Journal of Economic Perspectives,* 24(3), Summer 2010, pp. 133-150.

Generalizations about Using Value-Added Measures of Teacher Quality (with Steven G. Rivkin). *American Economic Review,* 100(2), May 2010, pp. 267-271.

Education and Economic Growth (with Ludger Woessmann) in Dominic J. Brewer and Patrick J. McEwan (ed.). *Economics of Education (Amsterdam: Elsevier, 2010),* 2010, pp. 60-67.

Education Production Functions: Developed Countries Evidence in Dominic J. Brewer and Patrick J. McEwan (ed.). *Economics of Education (Amsterdam: Elsevier, 2010),* 2010, pp. 132-136.

The High Cost of Low Educational Performance: The Long-Run Impact of Improving PISA Outcomes (with Ludger Woessmann). *Paris: Organization for Economic Cooperation and Development,* 2010, pp. 1-55.

Teacher Deselection in Dan Goldhaber and Jane Hannaway (ed.). *Creating a New Teaching Profession,* Washington, DC: Urban Institute Press, 2009, pp. 165-180.

Does Pollution Increase School Absences? (with Janet Currie, E. Megan Kahn, Matthew Neidell, and Steve G. Rivkin). *Review of Economics and Statistics,* 91(4), November 2009, pp. 683-694.

Many Schools Are Still Inadequate: Now what? (with Alfred A. Lindseth and Michael A. Rebel). *Education Next,* 9(4), Fall 2009, pp. 49-56.

School Policy: Implications of Recent Research for Human Capital Investments in South Asia and Other Developing Countries. *Education Economics,* 17(3), September 2009, pp. 291-313.

New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement (with John F. Kain and Steven G. Rivkin). *Journal of Labor Economics,* 27(3), July 2009, pp. 349-383.

Harming the Best: How Schools Affect the Black-White Achievement Gap (with Steven G. Rivkin). *Journal of Policy Analysis and Management,* 29(3), Summer 2009, pp. 366-393.

The Economic Value of Education and Cognitive Skills. *Handbook of Education Policy Research,* New York: Routledge, 2009, pp. 39-56.

Quality-Consistent Estimates of International Schooling and Skill Gradients (with Lei Zhang). *Journal of Human Capital,* 3(2), Summer 2009, pp. 107-143.

Getting Down to Facts: School Finance and Governance in California (with Susanna Loeb and Anthony Bryk). *Education Finance and Policy,* 3(1), Winter 2008, pp. 1-19.

The Role of Cognitive Skills in Economic Development (with Ludger Woessmann). *Journal of Economic Literature 2008,* 46(3), 2008, pp. 607-668.

Do Students Care about School Quality? Determinants of Dropout Behavior in Developing Countries (with Victor Lavy and Kohtaro Hitomi). *Journal of Human Capital,* 2(1), 2008, pp. 69-105.

Education Production Functions In Steven N. Durlauf and Lawrence E. Blume (eds.). *The New Palgrave Dictionary of Economics,* Basingstoke: Palgrave Macmillan, 2008.

What Do Cost Functions Tell Us About the Cost of an Adequate Education? (with Robert Costrell and Susanna Loeb). *Peabody Journal of Education,* 83(2), 2008, pp. 198-223.

Incentives for Efficiency and Equity in the School System. *Perspektiven der Wirtschaftspolitik,* 9 (Special Issue), 2008, pp. 5-27.

Education and Economic Growth: It's not Just Going to School but Learning That Matters (with Dean T. Jamison, ELliot A. Jamison and Ludger Woessmann). *Education Next,* 8(2), Spring 2008, pp. 62-70.

The Effects of Education Quality on Mortality Decline and Income Growth (with Eliot A. Jamison and Dean T. Jamison). *Economics of Education Review,* 26(2), December 2007, pp. 772-789.

The Single Salary Schedule and Other Issues of Teacher Pay. *Peabody Journal of Education,* 82(4), October 2007, pp. 574-586.

The Alchemy of 'Costing Out' an Adequate Education In Martin R. West and Paul E. Peterson (eds.). *School Money Trials: The Legal Pursuit of Educational Adequacy,* Washington, D.C.: Brookings Institution Press, 2007, pp. 77-101.

Education Quality and Economic Growth (with Ludger Woessmann)Washington, DC: World Bank, July 2007, 27 pages.

Some U.S. Evidence on how the Distribution of Educational Outcomes can be Changed In Ludger Woessmann and Paul E. Peterson (ed.). *Schools and the Equal Opportunity Problem,* Cambridge, MA: MIT Press, 2007, pp. 159-190.

The Confidence Men: Selling Adequacy, Making Millions. *Education Next,* 7(3), Summer 2007, pp. 73-78.

Charter School Quality and Parental Decision Making with School Choice (with John F. Kain, Steven G. Rivkin, and Gregory F. Branch). *Journal of Public Economics,* 91(5-6), June 2007, pp. 823-848.

Pay, Working Conditions, and Teacher Quality (with Steven G. Rivkin). *Future of Children,* 17(1), Spring 2007, pp. 69-96.

The Economic Benefits of Improved Teacher Quality In Nils C. Soquel and Pierre Jaccard (ed.). *Governance and Performance of Education Systems,* Dordrecht, Netherlands: Springer, 2007, pp. 107-135.

Teacher Quality (with Steven G. Rivkin) in Eric A. Hanushek and Finis Welch (ed.). *Handbook of the Economics of Education, Volume 2, Amsterdam: North Holland,* 2006, pp. 1052-1078.

School Resources In Eric A. Hanushek and Finis Welch (Ed.). *Handbook of the Economics of Education, Volume 2,* Amsterdam: North Holland, 2006, pp. 865-908.

Is There Hope for Expanded School Choice? In Robert C. Enlow and Lenore T. Ealy *Liberty and Learning: Milton Friedman's Voucher Idea at Fifty,* Washington, DC: Cato Institute, 2006, pp. 67-79.

Alternative School Policies and the Benefits of General Cognitive Skills. *Economics of Education Review,* 25(4), August 2006, pp. 447-462.

The Complementarity of Tiebout and Alonso (with Kuzey Yilmaz). *Journal of Housing Economics,* 16(2), August 2006, pp. 243-261.

Science Violated: Spending Projections and the "Costing Out" of an Adequate Education in Eric A. Hanushek (ed.). *Courting Failure: How School Finance Lawsuits Exploit Judges' Good Intentions and Harm Our Children,* Stanford, CA: Education Next Books, 2006, pp. 257-311.

Introduction: Good Intentions Captured – School Funding Adequacy and the Courts in Eric A. Hanushek (ed.). *Courting Failure: How School Finance Lawsuits Exploit Judges' Good Intentions and Harm Our Children,* Stanford, CA: Education Next Books, 2006, pp. xiii-xxxii.

Early Returns from School Accountability. *Generational Change: Closing the Test Score Gap,* Lanham, MD: Rowman and Littlefield, 2006, pp. 143-166.

More Accountability or More Resources: The US Experience with NCLB. *Institutional Models in Education: Legal Framework and Methodological Aspects for a New Approach to the Problem of School Governance,* (Nijmegen, Netherlands: Wolf Legal Publishers), 2006, pg. 69-80.

School Accountability and Student Performance. *Regional Economic Development,* Federal Reserve Bank of St. Louis, 2(1) , March 2006, pp. 51-61.

Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries (with Ludger Woessmann). *The Economic Journal,* 116(150), March 2006, pp. C63-C76.

Teacher Compensation in Paul E. Peterson (ed.). *Reforming Education in Florida,* 2006, pp. 149-163.

The Economic Value of Improving Public Schools. *Proceedings of Federal Reserve Bank of Cleveland Research Conference, November 18-19, 2004,* Cleveland: Federal Reserve Bank of Cleveland, 2005, pp. 59-72.

Pseudo-Science and a Sound Basic Education: Voodoo Statistics in New York. *Education Next,* 5(4), Fall 2005.

Economic Outcomes and School Quality. *Education Policy Series,* Volume 4. Paris: International

Institute for Educational Planning and International Academy of Education}, 2005.

The Economics of School Quality. *German Economic Review,* 6(3), August 2005, pp. 269-286.

Education and Training (with Michael Mertaugth). In Nicholas Barr (ed.). *Labor Markets and Social Policy in Central and Eastern Europe: The Accession and Beyond,* Washington, DC: The World Bank, 2005, pp. 207-251.

Why Quality Matters in Education. *Finance and Development,* 42(2), June 2005, pp.15-19.

Does School Accountability Lead to Improved Student Performance? (with Margaret E. Raymond). *Journal of Policy Analysis and Management,* 24(2), Spring 2005, pp. 298-327.

Teachers, Schools, and Academic Achievement (with Steven G. Rivkin and John F. Kain). *Econometrica,* 73(2), March 2005, pp. 417-458.

Why the Federal Government Should Be Involved in School Accountability. *Journal of Policy Analysis and Management,* 24(1), Winter 2005, pp. 168-172.

United States Lessons about School Accountability. *CESifo DICE Report,* 2(4), Winter 2004, pp. 27-32.

The Revolving Door (with John F. Kain and Steven G. Rivkin). *Education Next,* 4(1), Winter 2004, pp. 77-82.

Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools (with John F. Kain and Steven G. Rivkin). *Journal of Public Economics,* 88(9), August 2004, pp. 1722-1746.

The Toughest Battleground: Schools In Mark A. Wynne, Harvey Rosenblum and Robert L. Formaini (ed.). *The legacy of Milton and Rose Friedman's Free to Choose: Economic liberalism at the turn of the twenty first century,* Dallas, TX: Federal Reserve Bank of Dallas, 2004, pp. 21-35.

What If There Are No 'Best Practices'? *Scottish Journal of Political Economy,* 51(2), May 2004, pp. 156-172.

Shopping for Evidence Against School Accountability (with Margaret E. Raymond). *Developments in School Finance: 2003,* (Washington, DC: National Center for Education Statistics), 2004, pp. 119-130.

The Effect of School Accountability Systems on the Level and Distribution of Student Achievement (with Margaret E. Raymond). *Journal of the European Economic Association,* 2(2-3), May 2004, pp. 406-415.

Why Public Schools Lose Teachers (with John F. Kain and Steven G. Rivkin). *The Journal of Human Resources,* 39(2), Spring 2004, pp. 326-354.

How to Improve the Supply of High Quality Teachers (with Steven G. Rivkin). *Brookings Papers on Education Policy: 2004,* 2004, pp. 7-44.

Opportunities, Race, and Urban Location: The Influence of John Kain (with Edward L. Glaeser and John M. Quigley). *Journal of Urban Economics,* 56(1), 2004.

Economic Analysis of School Quality. *European Economy: Quality and Efficiency in Education,* Special Report No 3. Brussels: Directorate-General for Economic and Financial Affairs, European Commission, 2004, pp. 29-48.

Redistribution through Education and Other Transfer Mechanisms (with Charles Ka Yui Leung and Kuzey Yilmaz). *Journal of Monetary Economic,* 50(8), November 2003, pp. 1719-1750.

Does Peer Ability Affect Student Achievement? (with John F. Kain, Jacob M. Markman, and Steven G. Rivkin). *Journal of Applied Econometrics,* 18(5), October 2003, pp. 527-544.

Efficiency and Equity in Schools around the World (with Javier A. Luque). *Economics of Education Review,* 22(5), October 2003, pp. 481-502.

High Stakes Research (with Margaret E. Raymond). *Education Next,* 3(3), Summer 2003, pp. 48-55.

Lessons about the Design of State Accountability Systems (with Margaret E. Raymond). In Paul E. Peterson and Martin R. West (ed.). *No Child Left Behind? The Politics and Practice of Accountability,* Washington, DC: Brookings, 2003, pp. 126-151.

Does Public School Competition Affect Teacher Quality? (with Steven G. Rivkin). In Caroline Minter Hoxby (ed.) *The Economics of School Choice,* Chicago: University of Chicago Press, 2003, pp. 23-47.

Improving Educational Quality: How Best to Evaluate Our Schools (with Margaret E. Raymond). In Yolanda Kodrzycki (ed.). *Education in the 21st Century: Meeting the Challenges of a Changing World,* Boston, MA: Federal Reserve Bank of Boston, 2003, pp. 193-224.

Lost Opportunity. *Education Next,* 3(2), Spring 2003, pp. 84-87.

The Failure of Input-based Schooling Policies. *The Economic Journal,* 113, February 2003, pp. 64-98.

Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disabilities (with John F. Kain and Steven G. Rivkin). *Review of Economics and Statistics,* 84(4), November 2002, pp. 584-599.

The Seeds of Growth. *Education Next,* 2(3), Fall 2002, pp. 10-17.

The Importance of School Quality In Paul E. Peterson (ed.). *Our Schools and Our Future: Are We Still at Risk?,* (Stanford, CA: Hoover Institution Press), 2002, pp. 141-173.

Evidence, Politics, and the Class Size Debate In Lawrence Mishel and Richard Rothstein (ed.). *The Class Size Debate,* Washington, DC: Economic Policy Institute, 2002, pp. 37-65.

Teacher Quality. *Teacher Quality,* (Stanford: Hoover Institution Press), 2002, pp. 1-12.

Will Quality of Peers Doom Those Left in the Public Schools? in Paul T. Hill (ed.). *Choice with Equity,* 2002, pp. 121-140.

Publicly Provided Education in Alan J. Auerbach and Martin Feldstein (ed.). *Handbook of Public Economics (Amsterdam: North-Holland, 2002),* 2002, pp. 2045-2141.

Sorting out accountability systems (with Margaret E. Raymond). In Williamson M. Evers, Herbert J. Walberg (Eds.). *School Accountability,* 2002, pp. 75-104.

The Confusing World of Educational Accountability (with Margaret E. Raymond). *National Tax Journal,* 54(2), June 2001, pp. 365-384.

Black-white Achievement Differences and Governmental Interventions. *American Economic Review,* 91(2), May 2001, pp. 24-28.

Spending on Schools In Terry Moe (ed.). *A Primer on American Education,* Stanford, CA: Hoover Institution Press, 2001, pp. 69-88.

Schooling, Inequality, and the Impact of Government (with Julie A. Somers). In Finis Welch (ed.). *The Causes and Consequences of Increasing Inequality,* Chicago: University of Chicago Press, 2001, pp. 169-199.

Deconstructing RAND. *Education Matters,* 1(1), Spring 2001, pp. 65-70.

Economics of Education In Paul B. Baltes and Neil J. Smelser (ed.). *International Encyclopedia of the Social and Behavioral Sciences,* vol. 6, Amsterdam: Elsevier Science, 2001, pp. 4200-4208.

Schooling, Labor Force Quality, and the Growth of Nations (with Dennis D. Kimko). *American Economic Review,* 90(5), December 2000, pp. 1184-1208.

Smaller Classes, Lower Salaries? The Effects of Class Size on Teacher Labor Markets (with Javier Luque). In Sabrina W.M. Laine and James G. Ward (ed.). *Using What We Know: A Review of the Research on Implementing Class-Size Reduction Initiatives for State and Local Policymakers,* Oak Brook, Ill.: North Central Regional Educational Laboratory, 2000, pp. 35-51.

The Evidence on Class Size in Susan E. Mayer and Paul E. Peterson (ed.). *Earning and learning: How schools matter,* Washington, DC: Brookings Institution, 1999, pp. 131-168.

Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects. *Educational Evaluation and Policy Analysis,* 21(2), Summer 1999, pp. 143-163.

Adjusting for Differences in the Costs of Educational Inputs in William J. Fowler, Jr. (ed.). *Selected Papers in School Finance, 1997-1999,* 1999, pp. 17-27.

Budgets, Priorities, and Investment in Human Capital In Marvin H. Kosters (ed.). *Financing College Tuition: Government Policies and Educational Priorities,* Washington, DC AEI Press, 1999, pp. 8-27.

Conclusions and Controversies about the Effectiveness of School Resources. *FRBNY Economic Policy Review,* 4(1), March 1998, pp. 11-28.

Understanding the 20th Century Growth in U.S. School Spending (with Steven G. Rivkin). *Journal of Human Resources,* 31(1), Winter 1997, pp. 34-68.

Outcomes, Incentives, and Beliefs: Reflections on Analysis of the Economics of Schools. *Educational Evaluation and Policy Analysis,* 19(4), Winter 1997, pp. 301-308.

Incentives Are Key to Improved Schools. *Forum for Applied Research and Public Policy,* 12(3), 1997, pp. 62-67.

Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis,* 19(2), 1997, pp. 141-164.

Health and Schooling: Evidence and Policy Implications for Developing Countries (with João Batista Gomes-Neto, Raimundo Hélio Leite, and Roberto Cláudio Frota-Bezzera). *Economics of Education Review,* 16(3), June 1997, pp. 271-282.

Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis,* 19(2), Summer 1997, pp. 141-164.

Applying Performance Incentives to Schools for Disadvantaged Populations. *Education and Urban Society,* 29(3), May 1997, pp. 296-316.

Applying Performance Incentives to Schools for Disadvantaged Populations. *Education and Urban Society,* 29(3), 1997, pp. 296-316.

The Productivity Collapse in Schools In William J. Fowler, Jr. (ed.). *Developments in School Finance, 1996,* Washington, DC: National Center for Educational Statistics, U.S. Department of Education, 1997, pp. 183-195.

Analytical Framework for Retirement Policy Decisions (with Constance F. Citro). In Michael S. Gordon, Olivia S. Mitchell, and Marc M. Twinney (ed.). *Positioning Pensions for the Twenty-first Century,* Philadelphia: University of Pennsylvania Press, 1997, pp. 209-219.

Aggregation and the Estimated Effects of School Resources (with Steven G. Rivkin and Lori L. Taylor). *Review of Economics and Statistics,* 78(4), December 1996, pp. 611-627.

Measuring Investment in Education. *Journal of Economic Perspectives,* 10(4), Fall 1996, pp. 9-30.

A More Complete Picture of School Resource Policies. *Review of Educational Research,* 66(3), Fall 1996, pp. 397-409.

The Identification of School Resource Effects (with Steven G. Rivkin and Lori L. Taylor). *Education Economics,* 4(2), 1996, pp. 105-125.

Rationalizing School Spending: Efficiency, Equity, and Externalities, and Their Connection to Rising Expenditure In Victor Fuchs (ed.). *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America* , University of Chicago Press/NBER, 1996, pp. 59-91.

Efficiency-Enhancing Investments in School Quality (with João Batista Gomes Neto and Ralph W. Harbison). In Nancy Birdsall and Richard H. Sabot (ed.). *Opportunity Foregone: Education in Brazil* , Washington, DC: Inter-American Development Bank, 1996, pp. 385-424.

School Resources and Student Performance In Gary Burtless (ed.). *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success,* Washington, D.C.: The Brookings Institution, 1996, pp. 43-73.

Improving School Performance While Controlling Costs in William J. Fowler, Jr. (ed.). *Developments in School Finance, 1995,* 1996, pp. 111-122.

Interpreting Recent Research on Schooling in Developing Countries. *World Bank Research Observer,* 10(2), August 1995, pp. 227-246.

Who Chooses to Teach and Why? (with Richard Pace). *Economics of Education Review,* 14(2), June 1995, pp. 101-117.

The (Dis)Incentive to Settle Personal Injury Cases (with Michael J. Wolkoff). *New York State Bar Journal,* 67(3), 1995, pp. 52-56.

Outcomes, Costs, and Incentives in Schools In Board on Science, Technology, and Economic Policy, National Research Council. *Improving the Performance of America's Schools,* Washington, DC: National Academy Press, 1995, pp. 28-51.

The Economics of Structured Judgments Under CPLR Article 50-B (with Michael J. Wolkoff). *Buffalo Law Review,* 43(2), 1995, pp. 563-582.

The Quest for Equalized Mediocrity: School Finance Reform without Consideration of School Performance In Lawrence O. Picus and James L. Wattenbarger (ed.). *Where Does the Money Go? Resource Allocation in Elementary and Secondary Schools,* Thousand Oaks, CA: Corwin Press, 1995, pp. 20-43.

Production Functions in Education In T. Husén and T. N. Postlethwaite (ed.). *The International*

Encyclopedia of Education, Oxford: Pergamon Press, 1995, pp. 4059-4070.

A Jaundiced View of "Adequacy" in School Finance Reform. *Educational Policy,* 8(4), December 1994, pp. 460-69.

The Causes and Consequences of Grade Repetition: Evidence from Brazil (with João Batista Gomes-Neto). *Economic Development and Cultural Change,* 43(1), October 1994, pp. 117-148.

Money Might Matter Somewhere: A Response to Hedges, Laine, and Greenwald. *Educational Researcher,* 23(4), May 1994, pp. 5-8.

Estimating the Effects of Proposed Legislation: The Case for Model Validation with Constance F. Citro. *Chance,* 7(2), 1994, pp. 31-40.

Understanding Entry into the Teaching Profession (with Richard R. Pace). In Ronald G. Ehrenberg (ed.). *Choices and Consequences: Contemporary Policy Issues in Education,* Ithaca, NY: ILR Press, 1994, pp. 12-28.

Education Production Functions In Torsten Husén and T. Neville Postlethwaite (ed.). *International Encyclopedia of Education,* 2nd Edition, Volume 3, Oxford: Pergamon, 1994, pp. 1756-1762.

Concepts of Educational Efficiency and Effectiveness (with Marlaine E. Lockheed). In Torsten Husén and T. Neville Postlethwaite (ed.). *International Encyclopedia of Education,* 2nd Edition, Volume 3, Oxford: Pergamon, 1994, pp. 1779-1784.

Investimentos Autofinanciáveis em Educação (with João Batista Gomes-Neto, Raimundo H. Leite, and Ralph W. Harbison). *Cadernos de Pesquisa,* 85, 1993, pp. 11-25.

Can Equity Be Separated from Efficiency in School Finance Debates? In Emily P. Hoffman (ed.) *Essays on the Economics of Education,* Kalamazoo, MI: Upjohn Institute, 1993, pp. 35-73.

Improving Educational Outcomes While Controlling Costs (with Steven G. Rivkin and Dean T. Jamison). *Carnegie-Rochester Conference Series on Public Policy 37,* December 1992, pp. 205-238.

The Trade-off Between Child Quantity and Quality. *Journal of Political Economy,* 100(1), February 1992, pp. 84-117.

When School Finance 'Reform' May Not Be Good Policy. *Harvard Journal on Legislation,* 28(2), Summer 1991, pp. 423-456.

School Performance and Educational Policy Making In Gary Libecap (ed.). *The Education and Quality of the American Labor Force,* Greenwich, CT: JAI Press, 1991, pp. 65-94.

Concepts of Educational Efficiency and Effectiveness (with Marlaine E. Lockheed). In Torsten Husén and T. Neville Postlethwaite (ed.). *International Encyclopedia of Education,* Supplementary Volume 2, Oxford: Pergamon, 1991.

Commercial Land Use Regulation and Local Government Finance (with John M. Quigley). *American Economic Review,* 19(2), May 1990, pp. 176-180.

The Policy Research Markets. *Journal of Policy Analysis and Management,* 9(2), Spring 1990, pp. 147-154.

Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement (with Lori L. Taylor). *The Journal of Human Resources,* 25(2), Spring 1990, pp. 179-201.

Reforming Educational Reform (with John E. Chubb). In Henry Aaron (ed.). *Setting National Priorities* , Washington, DC: The Brookings Institution, 1990, pp. 213-247.

Social Science Research and Policy: A Review Essay. *Journal of Human Resources,* Spring 1990, pp. 46-51.

Expenditures, Efficiency, and Equity in Education: The Federal Government's Role. *American Economic Review,* 79(2), May 1989, pp. 46-51.

The Impact of Differential Expenditures on School Performance. *Educational Researcher,* 18(4), May 1989, pp. 45-51.

Improving Educational Efficiency in Developing Countries: What Do We Know? (with Marlaine E. Lockheed) *Compare,* 18(1), 1988, pp. 21-37.

Non-Labor-Supply Responses to the Income Maintenance Experiments In Alicia H. Munnell (ed.). *Lessons from the Income Maintenance Experiments,* Boston, MA: Federal Reserve Bank of Boston and the Brookings Institution, 1987, pp. 106-121.

Formula Budgeting: The Economics and Analytics of Fiscal Policy under Rules. *Journal of Policy Analysis and Management,* 6(1), Fall 1986, pp. 3-19.

The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature,* 49(3), September 1986, pp. 1141-1177.

Alternative Poverty Measures and the Allocation of Federal Benefits (with Roberton Williams). *Conference on the Measurement of Noncash Benefits,* Proceedings, Volume I, Washington, DC: U.S. Bureau of the Census, 1986, pp. 104-125.

Life-Cycle Earning Capacity and the OJT Investment Model (with John M. Quigley). *International Economic Review,* 26(2), June 1985, pp. 365-385.

Sources of Black-White Earnings Differences. *Social Science Research,* 11(2), June 1982, pp. 103-126.

The Determinants of Housing Demand (with John M. Quigley). In J.V. Henderson (ed.). *Research in*

Urban Economics, Vol. II. (Greenwich, CT: JAI Press), 1982, pp. 221-242.

Throwing Money at Schools. *Journal of Policy Analysis and Management,* 1(1), Autumn 1981, pp. 19-41.

Alternative Models of Earnings Determination and Labor Market Structure. *Journal of Human Resources,* 16(2), Spring 1981, pp. 238-259.

Education Policy Research-An Industry Perspective. *Economics of Education Review,* 1(2), 1981, pp. 193-224.

Consumption Aspects (with John M. Quigley). In Katherine L. Bradbury and Anthony Downs (ed.). *Do Housing Allowances Work?* , Washington, D.C.: The Brookings Institution, 1981, pp. 185-240.

What is the Price Elasticity of Housing Demand? (with John M. Quigley). *Review of Economics and Statistics,* 62(3), August 1980, pp. 449-454.

Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources,* 14(3), Summer 1979, pp. 351-388.

The Dynamics of the Housing Market: A Stock Adjustment Model of Housing Consumption (with John M. Quigley). *Journal of Urban Economics,* 6(1), January 1979, pp. 90-111.

An Explicit Model of Intra-Metropolitan Mobility (with John M. Quigley). *Land Economics,* 54(4), November 1978, pp. 411-429.

The Dynamics of Postwar Industrial Location (with Byung Nak Song). *Review of Economics and Statistics,* 60(4), November 1978, pp. 515-522.

Implicit Investment Profiles and Intertemporal Adjustments of Relative Wages (with John M. Quigley). *American Economic Review,* 68(1), March 1978, pp. 67-79.

Ethnic Income Variations: Magnitudes and Explanations. *American Ethnic Groups* , Washington, D.C.: The Urban Institute, 1978, pp. 139-156.

Housing Market Disequilibrium and Residential Mobility (with John M. Quigley). In Eric A. Moore and William A.V. Clark (ed.). *Population Mobility and Residential Change,* Evanston, IL: Northwestern University Press, 1978.

The Volunteer Military and the Rest of the Iceberg. *Policy Sciences,* 8(3), September 1977, pp. 343-361.

Learning by Observing the Performance of Schools In Robert A. Wallhaus (ed.). *New Directions for Institutional Research,* Measuring and Increasing Academic Productivity, No. 8 (San Francisco: Jossey-Bass), 1975, pp. 17-38.

The Demand for Local Public Service: An Exploratory Analysis In John E. Jackson (ed.). *Public Needs and Private Behavior in Metropolitan Areas,* Cambridge, MA: Ballinger, 1975, pp. 121-138.

Model Specification, Use of Aggregate Data, and the Ecological Correlation Fallacy (with John E. Jackson and John F. Kain). *Political Methodology,* Winter 1974, pp. 89-107.

Efficient Estimators for Regressing Regression Coefficients. *The American Statistician,* 28(2), May 1974, pp. 66-67 .

The High Cost of Graduate Education in the Military. *Public Policy,* 21(4), 1973, pp. 525-552.

Regional Differences in the Structure of Earnings. *Review of Economics and Statistics,* 55(2), May 1973, pp. 204-213.

On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy (with John F. Kain). In Frederick Mosteller and Daniel P. Moynihan (ed.). *On Equality of Educational Opportunity,* (New York: Random House), 1972, pp. 116-145.

Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data. *American Economic Review,* 61(2), May 1971, pp. 280-288.

The Value of Teachers in Teaching. *RM-6362-CC/RC,* Santa Monica: The Rand Corporation, 1970.

The Production of Education, Teacher Quality, and Efficiency in U.S. Office of Education. *Do Teachers Make a Difference?* , Washington, D.C.: Government Printing Office, 1970, pp. 79-99.

Voting Behavior in the 1960 Colorado Presidential Election (with John C. Ries, John E. Coblentz, and John F. Kain). *Rocky Mountain Social Science Journal,* 3(1), 1966, pp. 82-95.

Working Papers

Land Use Controls and the Provision of Education (with Kuzey Yilmaz). *NBER Working Paper 17730,* January 2012.

Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals (with Gregory F. Branch and Steven G. Rivkin). *NBER Working Paper 17803,* January 2012.

General Education, Vocational Education, and Labor-Market Outcomes over the Life-Cycle (with Ludger Woessmann and Lei Zhang). *NBER Working Paper 17504,* October 2011.

School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010 (with Paul W. Glewwe, Sarah D. Humpage, and Renato Ravina). *NBER Working Paper 17554,* October 2011.

Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools? (with

Steven G. Rivkin) *NBER Working Paper 15816,* March 2010.

The Market for Teacher Quality (with John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin). *NBER Working Paper 11154,* February 2005.

Notes/Comments

Dual Education: Europe's Secret Recipe? *CESifo Forum,* 13(3), Autumn 2012, 29-32.

Fewer school days is the worst of budget options for California. *San Jose Mercury News,* July 20, 2012.

Education is the Key to a Healthy Economy (with George P. Shultz). *Wall Street Journal,* May 1, 2012.

Low-Performing Teachers Have High Costs. *Education Next,* 12(3), Summer 2012.

Misplaced Optimism and Weighted Funding. *Education Week,* 31(26), March 28, 2006, pp. 28,36.

Teacher ratings are a vital step forward. *New York Daily News,* February 24, 2012.

Allowing local schools to make more decisions may work in developed countries but is questionable in developing countries (with Susanne Link, and Ludger Woessmann). *Vox,* January 9, 2012.

Math Matters. *Hoover Digest,* No. 1, 2011.

Vocational education facilitates entry into the labour market but hurts employment at older ages (with Ludger Woessmann and Lei Zhang). *Vox,* November 21, 2011.

Why Can't American Students Compete with the Rest of the World? (with Paul E. Peterson) *Newsweek,* September 5, 2011, 42-45.

Overview of the Symposium on Performance Pay for Teachers (with Ludger Woessmann). *Economics of Education Review,* 30(3), June 2011, pp. 391-393.

Recognizing the Value of Good Teachers. *Education Week,* 30(27), April 6, 2011, pp. 34-35.

Education reform solves state's budget crisis (with George P. Shultz). *San Francisco Chronicle,* April 10, 2011, p. F-6.

Saving the schools: Why more money is not the answer. *New York Post,* April 1, 2011.

The "War on Teachers" Is a Myth. *Hoover Digest,* No. 1, 2011.

Why Is It So Hard To Make Teachers Better? *Defining Ideas,* January 30, 2011.

Viewpoints: Test evaluation put teachers on the spot. *Sacramento Bee,* November 12, 2010.

There is no 'War on Teachers". *Wall Street Journal,* October 19, 2010.

UFT wrong to fight Joel Klein's attempt to release teacher data, says leading education researcher. *New York Daily News,* October 27, 2010.

Cry Wolf! This Budget Crunch Is for Real. *Education Week,* 29(32), May 2010, pp. 32-40.

Total Student Load: Maybe worth a longer look, but hardly a revolution (A review of The Secret of TSL: The revolutionary discovery that raises school performance, by William G. Ouchi). *Education Next,* 10(2), Spring 2010, pp. 84-85.

The Choice Movement and the Courts (with Alfred A. Lindseth). *School Choice Advocate* , Foundation for Educational Choice, February 2010.

Building on No Child Left Behind. *Science,* 326, November 2009, pp. 802-803.

Judicial Funding Mandates Related to Education Sharply Decline (with Alfred A. Lindseth). *State Courts Project, Federalist Society,* Fall 2009.

Poor student learning explains the Latin American growth puzzle (with Ludger Woessmann). *Vox,* August 14, 2009.

An F in Effectiveness (with Alfred A. Lindseth). *Hoover Digest,* No. 3, , pp. 43-47.

California Needs to Make Wiser Use of School Funding (with Alfred A. Lindseth). *San Francisco Chronicle,* June 17, 2009, p. A-15.

Performance-Based Funding (with Alfred A. Lindseth). *Education Week,* 28(33), June 10, 2009, pp. 28-30.

Performance-Based Funding. *Defining Ideas,* 1, 2009, pp. 101-105.

The Effectiveness of Court-Ordered Funding of Schools (with Alfred A. Lindseth). *Education Outlook,* No. 6 (Washington, DC: American Enterprise Institute), May 2009.

John Forest Kain (1935-2003) In Steven N. Durlauf and Lawrence E. Blume (eds.). *The New Palgrave Dictionary of Economics,* Basingstoke: Palgrave Macmillan, 2008.

Getting down to facts: School finance and governance in California (with Susanna Loeb and Anthony S. Bryk). *Institute for Research on Education Policy and Practice, Stanford University,* September 2007.

A numbers game: Consultants sell legislatures school studies and collect millions. *Arkansas*

Democrat-Gazette, June 17, 2007, Perspective, pp. 93, 98.

Is the 'Evidence-Based Approach' a Good Guide to School Finance Policy? *Paper commissioned by Washington Learns,* March 2007.

Milton Friedman's Unfinished Business. *Hoover Digest,* No. 1, Winter 2007, pp. 42-49.

The Court's Gift to Spitzer. *The New York Sun,* November 28, 2006.

The Cost of an 'Adequate' Education. *The Wall Street Journal,* October 9, 2006, p. A19.

Choice, Charters, and Public School Competition. *Economic Commentary, Federal Reserve Bank of Cleveland,* March 15, 2006, March 15, 2006.

Comment [on Murnane, Willett, Bub, and McCartney] In Gary Burtless and Janet Rothenberg Pack(ed.). *Brookings-Wharton Papers on Urban Affairs,* Washington, DC: Brookings Institution, 2006, pp. 128-131.

Remedial Math: Rather than Spend More on Schools, We Should Spend More Wisely. *The New Democrat,* 7(6) , November/December 1995, pp. 25-27.

Education and the Economy: Our School Performance Matters. *Education Week,* 24(21), February 2, 2005.

Why the Federal Government Should be Involved in School Accountability. *Journal of Policy Analysis and Management* , 24(1), Winter, 2005, pp. 168-172.

Rewarding Teachers (with Caroline M. Hoxby). *Reforming Education in Arkansas,* (Stanford: Hoover Institution Press), 2005, pp. 155-166.

Developing Value-Added Measures for Teachers and Schools (with Caroline M. Hoxby). *Reforming Education in Arkansas,* (Stanford: Hoover Institution Press), 2005, pp. 99-104.

Our School Performance Matters. *Journal of Education,* 185(3), November 2004, pp. 1-6.

Learn Lessons of the Past. *The Australian,* August 26, 2004.

The Underrepresentation of Minority Faculty in Higher Education: Panel Discussion. *American Economic Review,* 94(2), May 2004, pp. 304-306.

It's not how much, but how you spend the money on schools. *Salt Lake Tribune,* April 18, 2004.

How to Determine Who is a Quality Teacher (with Lewis C. Solmon, Philip Bigler, Lee S. Shulman, and Herbert J. Walberg). In Lewis C. Solmon and Tamara W. Schiff (ed.). *Talented Teachers: The Essential Force for Improving Student Achievement,* Greenwich, CT: Information Age Publishing, 2004, pp. 49-85.

Same Amount of Money Should Yield Better Results. *San Jose Mercury News,* November 23, 2003, p. 1P, 3P.

Potential Gains and Losses in Education. *World and I,* September 2003, pp. 30-35.

A False Schools 'Fix'. *New York Post,* June 30, 2003.

End Class-Size Straightjacket. *Los Angeles Times,* April 27, 2003, p. M2.

An Enormous Opportunity. *Arkansas Democrat-Gazette,* April 1, 2003.

Comment [on Krueger and Heckman] In Benjamin M. Friedman (ed.). *Inequality in America: What Role for Human Capital Policies?,* Cambridge, MA: MIT Press, 2003, pp. 252-269.

Public School Finance and Urban School Policy: General versus Partial Equilibrium Analysis In William G. Gale and Janet Rothenberg Pack (ed.). *Brookings-Wharton Papers on Urban Affairs,* Washington, DC: Brookings Institution, 2003, pp. 171-176.

Teacher quality and teacher salaries. *Policy Brief,* Stanford Institute for Economic Policy Research, Stanford University, November 2002.

Introduction to the JHR's Special Issue on Designing Incentives to Promote Human Capital (with James Heckman, Derek Nea). *The Journal of Human Resources,* 37(4), Autumn, 2002, pp. 693-695.

Comment [on Galiani and Schargrodsky]. *Economia,* 2(2), Spring 2002, April 2002, pp. 302-305.

Efficiency and Equity in Education. *NBER Reporter, National Bureau of Economic Research,* 2001, p. 15-19.

The Sequel: RAND versus RAND. *Education Matters,* 1(1), Spring, 2001, pp. 68-69.

Teacher quality and school reform (with Steven G. Rivkin). *Education Finance Research Consortium, The Teaching Workforce, Symposium Proceedings,* Albany: Center for Policy Research, Rockefeller College, University of Albany, 2001, pp. 81-99.

The Truth about Teacher Salaries and Student Achievement In Williamson M. Evers, Lance T. Izumi, and Pamela A. Riley (ed.). *School Reform: The Critical Issues,* (Stanford, CA: Hoover Institution Press), 2001, pp. 174-175.

Money Alone Will Not Fix Bad Schools. *New York Daily News,* January 23, 2001.

Further Evidence on the Effects of Catholic Secondary Schooling in William G. Gale and Janet Rothenberg Pack (ed.). *Brookings-Wharton Papers on Urban Affairs,* (Washington, DC: Brookings Institution, 2000), 2000, pp. 194-197.

Class Size Reduction: Good Politics, Bad Education Policy. *High School Magazine,* 6(4), January/February 1999, p. 44.

Review of Susan E. Mayer, What Money Can't Buy: Family Income and Children's Life Chances. *Journal of Policy Analysis and Management,* 17(2), Summer 1998, pp. 535-538.

Improving Student Achievement: Is Reducing Class Size the Answer? *Policy Brief, Progressive Policy Institute,* June 1998.

Are Resources Important? *Journal of Negro Education,* 66(3), Summer, 1997, pp. 289-303.

Review of Cohn and Johnes. *Economics of Education Review,* 16(3), June 1997, pp. 346-347.

Why True Reform of Schools is so Unlikely. *Jobs and Capital,* (Milken Institute for Jobs and Capital Formation), 6, Winter, 1997, pp.23-27.

Discussion [of Kain and Singleton]. *New England Economic Review,* May/June 1996, pp. 111-114.

Comment on Chapters Two, Three, and Four in Helen Ladd (ed.). *Holding Schools Accountable: Performance-Based Reform in Education,* Washington, DC: Brookings Institution, 1996, pp. 128-136.

Making Schools Work: The Economics of School Reform in Robert Berne (ed.). *Study on Cost-Effectiveness in Education,* (New York: New York State Board of Regents), 1996.

Moving Beyond Spending Fetishes. *Educational Leadership,* 53(3), November 1995, pp. 60-64.

Making Schools Work: Spending and Student Achievement. *Heartland Policy Study No. 68,* (The Heartland Institute), September 26, 1995.

Schools Need Incentives, Not More Money. *Wall Street Journal,* October 5, 1994.

Education Investment and Education Reform. *Jobs and Capital,* (Milken Institute for Jobs and Capital Formation) 3 (Fall), 1994, pp. 36-38.

Making America's Schools Work: This Time Money is Not the Answer. *Brookings Review,* Fall, 1994, pp. 10-13.

Money Might Matter Somewhere: A Response to Hedges, Laine, and Greenwald ," , , .. *Educational Researcher,* 23(4), May 1994, pp. 5-8.

Microsimulation Models for Social Welfare Programs: An Evaluation (with Constance F. Citro). *Focus, Institute for Research on Poverty,* University of Wisconsin—Madison, Winter , December 1993, p. 13-21.

Comment [on Manski] in Charles T. Clotfelter and Michael Rothschild. *Studies of Supply and Demand in Higher Education,* (Chicago: University of Chicago Press), 1993, pp. 57-60.

Revew [of McPherson and Schapiro]. *Economics of Education Review* , 12(2), 1993, pp. 187-188.

Will More Spending Fix Unequal Schools? *Detroit News*, October 30, 1991.

Bringing Educational Measurement into the Age of Newton (with C. Eugene Steuerle and Robert H. Meyer). *Policy Bites,* The Urban Institute, October 1991.

Testing Economic Knowledge. *Journal of Economic Education,* 22(3), Summer, 1991, pp. 273-75.

Walter Y. Oi: Reflections on his Career and his Legacy. *Carnegie-Rochester Conference Series on Public Policy 33,* Autumn 1990, pp. 9-12.

Schools in the 1990s: The Opportunities and Risks Facing Texas and Other States in Gerald P. O'Driscoll, Jr. and Stephen P.A. Brown (ed.). *The Southwest Economy in the 1990s: A Different Decade,* (Boston: Kluwer Academic Publishers), , pp. 39-43.

Review of "The Green Book". *Journal of Policy Analysis and Management,* 1989 Edition, 8(4), Fall, 1989, pp. 691-694.

Overview of Federal Involvement in Education. *in National Economic Commission, Staff Papers, Background Papers, and Major Testimony,* March 1989, pp. 329-333.

American Domestic Priorities in John Quigley and Daniel Rubinfeld (ed.). *American Domestic Priorities,* (Berkeley, CA: Univ. of California Press), 1985, pp. 154-160.

The Continuing Hope: A Rejoinder. *Journal of Policy Analysis and Management,* 1(1), Autumn 1981, pp. 53-54.

Throwing Money at Schools. *Education Week,* November 2, 1981 .

Comment [on Apgar] in Gregory K. Ingram (ed.). *Residential Location and Urban Housing Markets,* NBER Studies in Income and Wealth, No. 45. (NY: National Bureau of Economic Research), 1977, pp. 173-180.

Review [of Sewell and Hauser]. *Journal of Human Resources,* 11(3), Summer, 1976, pp. 420-422.

Comment [on Levin] in Joseph T. Froomkin, Dean T. Jamison, and Roy Radner (ed.). *Education as an Industry,* (New York: National Bureau of Economic Research), 1976, pp. 191-196.

Price Data and Data Systems (with Paul G. Clifford, Eric Hanushek, and Elaine Gilde). *Office of Economic Stabilization, Historical Working Papers on the Economic Stabilization Program,* Volume II. Washington, D.C.: Government Printing Office, 1974.

General Testimony

"**Testimony for the House Finance and Appropriations Committee**," State of Ohio, December 5, 2012

"Testimony before the Senate Finance Committee," State of Ohio, December 5, 2012

"**Alternate Economic Futures for the United States**," U.S. Senate Committee on Health, Education, Labor and Pensions, March 8, 2012

"**Testimony for the House and Senate Education Committees**," Ohio Legislature, March 10, 2011

"**Improving Teacher Quality**," Senate Committee on Education, Texas Legislature, Austin, TX, July 20, 2010.

"**Testimony for the Select Committee on Public School Finance Weights, Allotments & Adjustments**," Texas Legislature, San Antonio, TX, May 19, 2010.

"**The importance of improved data access in Texas**," House Committee on Public Education, *Texas House of Representatives*, July 16, 2008

"Students First: Reforming California Schools," California Senate Education Committee, April 30, 2008

"What is Success and How do We Achieve It?" *High School Completion and Success Initiatives Council* [TX], February 15, 2008

"**Performance Incentives for Teachers and Administrators**,," Senate Select Committee on Education Reform & Public School Finance, *Texas State Senate*, February 27, 2006.

"School Finance Proposals: The Koret Task Force Recommendations," House Select Committee on Public School Finance, *Texas House of Representatives*, February 18, 2004

"**Thinking about School Finance in Texas**," Testimony before the Subcommittee on Cost Adjustments for the House Select Committee on Public School Finance, *Texas House of Representatives*, October 22, 2003. [reprinted in Chris Patterson (ed.), *Putting the Sides Together: Twelve Perspectives on Texas Public School Finance* (Texas Public Policy Foundation, December 2003)]

"**Arkansas Student Accountability and Educational Accountability Act of 2003**," Testimony before the Education Committee, *House of Representatives, State of Arkansas*, March 26, 2003.

"**The Structure and Funding of Special Education**," Testimony before The President's Commission on Excellence in Special Education, Los Angeles, California, March 21, 2002

"**Should Class Size Reduction Substitute for Innovation?**" Testimony before the Committee on Health, Education, Labor and Pensions, *United States Senate*, June 23, 1999 [Note also comments on "**ESEA Re-authorization and Experimentation**," July 8, 1999]

"**The Evidence on Class Size**," Subcommittee on Early Childhood, Youth, and Families, Committee on Education and the Workforce, *U.S. House of Representatives*, February 24, 1998

"**Educational Reform and Current Education Initiatives**," Committee on Education and the Workforce, *U.S. House of Representatives*, March 13, 1997

"**The Federal Role in Educational Reform**," Committee on the Budget, *U.S. Senate*, February

5, 1997

"Approaches to Reforming Vermont Schools," testimony before joint meeting of the Senate and House Education Committees, *Vermont Legislature*, January 11, 1996

"School Finance and Educational Reform," Subcommittee on Education, Arts, and Humanities, Committee on Labor and Human Resources, *U.S. Senate*, Washington, DC, July 26, 1993

"Overview of Federal Education Roles and Spending", *National Economic Commission*, Washington, DC, October 5, 1988

"Evaluation of the School Desegregation Study", *U.S. Commission on Civil Rights*, Washington, DC, February 8, 1986

"Long Run Budget Trends and Space Development Options," *National Commission on Space*, Washington, DC, September 17, 1985

"Future Budgetary Requirements for the 600-Ship Navy," Subcommittee on Seapower and Strategic and Critical Materials, Committee on Armed Services, *U.S. House of Representatives*, Washington, DC, September 6, 1985

"Space Shuttle Pricing Policy," Subcommittee on Science, Technology, and Space, Committee on Commerce, Science, and Transportation, *U.S. Senate*, Washington, DC, March 27, 1985

"Space Shuttle Pricing Policy", Subcommittee on Space Science and Applications, Committee on Science and Technology, *U.S. House of Representatives*, Washington, DC, March 5, 1985

"Financial Outlook for the Hospital Insurance Trust Fund," Subcommittee on Health, Committee on Ways and Means, *U.S. House of Representatives*, Washington, DC, September 13, 1984

"Steel Import Restrictions," Subcommittee on Trade, Committee on Ways and Means, *U.S. House of Representatives*, Washington, DC, June 20, 1984

"The Measurement of Poverty," Subcommittee on Census and Population, Committee on Post Office and Civil Service, and Subcommittee on Oversight, Committee on Ways and Means, *U.S. House of Representatives*, Washington, DC, May 15, 1984

"Targeted Jobs Tax Credit," Subcommittee on Select Revenue Measures, Committee on Ways and Means, *U.S. House of Representatives*, Washington, DC, April 10, 1984

"Financial Condition of the Highway Trust Fund," Committee on Ways and Means, *U.S. House of Representatives*, Washington, DC, February 23, 1984

Legal Testimony

School Finance

Gannon et al. v. State of Kansas (Kansas), 2012

Lobato et al. v. State of Colorado (Colorado), 2011

Abbott v. Burke (New Jersey), 2011

McCleary v. State of Washington (Washington), 2009

Davis v. South Dakota (South Dakota), 2008

Espinoza v. State of Arizona (Arizona), 2008

Committee for Educational Equality et al. v. State of Missouri et al. (Missouri), 2007

School Districts' Alliance v. State of Washington (Washington), 2006
Montoy et al. v. State of Kansas, et al. (Kansas), 2003
Campaign for Fiscal Equity et al. v. State of New York et al. (New York), 2000
Hoke County Board of Education et al. v. State of North Carolina (North Carolina), 1999
Claremont School District, et al. v. Merrill et al. (New Hampshire), 1996
Committee for Educational Equality v. Missouri; *Lee's Summit v. Missouri* (Missouri), 1992
Alabama Coalition for Equity v. Hunt; *Harper v. Hunt* (Alabama), 1992
Tennessee Small Schools et al. v. McWherter et al. (Tennessee), 1990
Abbott v. Burke (New Jersey), 1987
Somerset County Board of Education v. Hornbeck (Maryland), 1980
Levittown v. Nyquist (New York), 1976
Serrano v. Priest (California), 1973

Other Court Testimony

Jenkins et al. v. State of Missouri et al. (Kansas City, Mo.), 1997
Liddell et al. v. St. Louis et al., 1996

Court Submissions

Horne v. Flores, 2009

SMITH RICHARDSON FOUNDATION, INC.

60 Jesup Road
Westport, CT 06880
Tel: (203) 222-6222
Fax: (203) 222-6282

November 30, 2012

Ms. Nicole K. Pobuta
Contract & Grant Officer
Stanford University
Office of Sponsored Research
3160 Porter Drive, Suite 100
Palo Alto, CA 94304-8445

Dear Ms. Pobuta:

I am pleased to advise you that the Trustees of the Smith Richardson Foundation have approved a grant of $34,163 to Stanford University supporting Dr. Margaret E. Raymond's project entitled, *The Second National Charter School Study (Supplement)* (SRF Grant #2012-9042), according to the terms of a letter submitted on December 2, 2011. For reporting purposes, this grant will be combined with an earlier award (SRF Grant #2011-8707). This grant will be paid upon receipt of the signed copy of this letter in our Greensboro office specified below.

The accompanying Memorandum of Grant Conditions is a part of this letter by reference. **Please note paragraph 5, no reference to this Foundation (other than whatever may be required by government regulation) may be made without our approval.** Mark Steinmeyer, Senior Program Officer of the Foundation, will send a letter, which is incorporated into this grant agreement by reference, that conveys the comments of the Foundation's Governors, any conditions imposed on this grant by the Trustees, the Foundation's reporting requirements, and the product/s to be produced as a result of this grant.

Kindly evidence acceptance of these conditions by signing the copy of this letter and returning it, along with a copy of your IRS letter of tax exemption to:

Ms. Karla Frank
SMITH RICHARDSON FOUNDATION, INC.
701 Green Valley Road, Suite 300
Greensboro, NC 27408

Sincerely,

[redacted]
Peter L. Richardson
President

[redacted]
Nicole K. Pobuta
Contract & Grant Officer

Enclosures
cc: Karla Frank
cc: Dr. Margaret E. Raymond
30133634.doc

Form **W-9**
(Rev. December 2011)
Department of the Treasury
Internal Revenue Service

Request for Taxpayer Identification Number and Certification

Give Form to the requester. Do not send to the IRS.

Print or type
See Specific Instructions on page 2.

Name (as shown on your income tax return)
Board of Trustees of the Leland Stanford Junior University

Business name/disregarded entity name, if different from above

Check appropriate box for federal tax classification:
☐ Individual/sole proprietor ☐ C Corporation ☐ S Corporation ☐ Partnership ☐ Trust/estate

☐ Limited liability company. Enter the tax classification (C=C corporation, S=S corporation, P=partnership) ▶

☐ Exempt payee

☑ Other (see instructions) ▶ **Trust with Corporate Powers**

Address (number, street, and apt. or suite no.)
3145 Porter Drive

Requester's name and address (optional)

City, state, and ZIP code
Palo Alto, CA 94304

List account number(s) here (optional)

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the name given on the "Name" line to avoid backup withholding. For individuals, this is your social security number (SSN). However, for a resident alien, sole proprietor, or disregarded entity, see the Part I instructions on page 3. For other entities, it is your employer identification number (EIN). If you do not have a number, see *How to get a TIN* on page 3.

Note. If the account is in more than one name, see the chart on page 4 for guidelines on whose number to enter.

Social security number
[ ][ ][ ] - [ ][ ] - [ ][ ][ ][ ]

[redacted] **identification number**
[redacted]

Part II Certification

Under penalties of perjury, I certify that:

1. The number shown on this form is my correct taxpayer identification number (or I am waiting for a number to be issued to me), and
2. I am not subject to backup withholding because: (a) I am exempt from backup withholding, or (b) I have not been notified by the Internal Revenue Service (IRS) that I am subject to backup withholding as a result of a failure to report all interest or dividends, or (c) the IRS has notified me that I am no longer subject to backup withholding, and
3. I am a U.S. citizen or other U.S. person (defined below).

Certification instructions. You must cross out item 2 above if you have been notified by the IRS that you are currently subject to backup withholding because you have failed to report all interest and dividends on your tax return. For real estate transactions, item 2 does not apply. For mortgage interest paid, acquisition or abandonment of secured property, cancellation of debt, contributions to an individual retirement arrangement (IRA), and generally, payments other than interest and dividends, you are not required to sign the certification, but you must provide your correct TIN. See the instructions on page 4.

Sign Here Signature of U.S. person ▶ [redacted] Date ▶ 11/22/11

General Instructions

Section references are to the Internal Revenue Code unless otherwise noted.

Purpose of Form

A person who is required to file an information return with the IRS must obtain your correct taxpayer identification number (TIN) to report, for example, income paid to you, real estate transactions, mortgage interest you paid, acquisition or abandonment of secured property, cancellation of debt, or contributions you made to an IRA.

Use Form W-9 only if you are a U.S. person (including a resident alien), to provide your correct TIN to the person requesting it (the requester) and, when applicable, to:

1. Certify that the TIN you are giving is correct (or you are waiting for a number to be issued),
2. Certify that you are not subject to backup withholding, or
3. Claim exemption from backup withholding if you are a U.S. exempt payee. If applicable, you are also certifying that as a U.S. person, your allocable share of any partnership income from a U.S. trade or business is not subject to the withholding tax on foreign partners' share of effectively connected income.

Note. If a requester gives you a form other than Form W-9 to request your TIN, you must use the requester's form if it is substantially similar to this Form W-9.

Definition of a U.S. person. For federal tax purposes, you are considered a U.S. person if you are:

- An individual who is a U.S. citizen or U.S. resident alien,
- A partnership, corporation, company, or association created or organized in the United States or under the laws of the United States,
- An estate (other than a foreign estate), or
- A domestic trust (as defined in Regulations section 301.7701-7).

Special rules for partnerships. Partnerships that conduct a trade or business in the United States are generally required to pay a withholding tax on any foreign partners' share of income from such business. Further, in certain cases where a Form W-9 has not been received, a partnership is required to presume that a partner is a foreign person, and pay the withholding tax. Therefore, if you are a U.S. person that is a partner in a partnership conducting a trade or business in the United States, provide Form W-9 to the partnership to establish your U.S. status and avoid withholding on your share of partnership income.

IMPORTANT

We need to process payment before December 12, 2012, therefore, in order to expedite this grant, please:

1. Immediately fax your signed copy of this letter to:

   Karla Frank
   Smith Richardson Foundation, Inc.
   Fax: 336-379-5580
   Phone: 336-379-8600

   -and-

   Immediately e-mail a signed scanned copy to:
   Fran Boulds
   Smith Richardson Foundation, Inc.
   Fax: 203-222-6282
   Phone: 203-222-6222
   fboulds@srf.org

2. Immediately return by overnight carrier the original of your signed letter along with a copy of your IRS letter of tax exemption to:

   Karla Frank
   Smith Richardson Foundation, Inc.
   701 Green Valley Road, Suite 300
   Greensboro, NC 27408
   Phone: 336-379-8600

Thank you for your cooperation.

SMITH RICHARDSON FOUNDATION, INC.

60 Jesup Road
Westport, CT 06880
Tel: (203) 222-6222
Fax: (203) 222-6282

December 6, 2012

Dr. Margaret E. Raymond
Stanford University
Center for Research on Education Outcomes
434 Galvez Mall
Stanford, CA 94305-6010

Project Title: *The Second National Charter School Study*
Grant No.: 2012-9042 consolidated with 2011-8707
Start Date: December 1, 2012
End Date: January 2, 2014

Dear Macke:

Congratulations on receiving a grant from the Smith Richardson Foundation. As noted in the Foundation's November 30, 2012, grant agreement sent to Nicole Pobuta, today's letter, which is incorporated into the grant agreement by reference, sets forth the reporting requirements associated with this grant.

Standard Grant Conditions
The attached Memorandum of Grant Conditions was included with the grant agreement. Please be sure to keep the following conditions in mind as your project proceeds:

- No reference to the Smith Richardson Foundation may be made (other than whatever may be required by government regulation) without prior approval from a Foundation program officer.

- Advance notice must be given to the Foundation if the principal investigator/s will be changed during the course of this grant. The Foundation reserves the right to terminate this grant and seek a return of funds if the Foundation, in its sole discretion, concludes that the change in personnel will substantially affect the quality of the product of the grant.

Adherence to Proposal
I wish to emphasize that the project must adhere to the terms specified in the proposal. We understand, however, that circumstances may dictate minor revisions in the work plan. If this occurs, the Foundation requires that the grantee consult with the program officer. However, any significant departure from the plan set forth in the proposal (personnel, project subject matter, activities, budget line items or allocations, timing and schedule, or the expected product) must be approved in writing by the program officer.

Public Events
You must inform your program officer of all public meetings held in connection with the project in the event that members of the Foundation's staff or Board of Governors wish to attend.

Product
Based on your proposal, it is our understanding that the final product resulting from this grant will be a report and a series of papers. As soon as this product is available, please provide us with two copies.

Reporting Requirements
As a condition of the grant, you are required to send narrative and financial progress reports to the Foundation until we receive your final product. Because this award is a supplement to an award made earlier this year (SRF Grant #2011-8707), the reporting for the two grants will be consolidated under the number of the earlier award. Throughout the course of the project you will receive reminders from our records coordinator, Dale Stewart, letting you know that a report is due. We expect to receive these reports from you in a timely manner.

- Financial Reports: Your institution will be expected to provide us with a financial report at twelve-month intervals as well as a final financial report when all funds are expended.

- Narrative Reports: As principal investigator for the project, you will be expected to submit narrative progress reports to us every six months until the final product is submitted to the Foundation.

  Every scheduled narrative report should include a listing of any products produced or dissemination activities undertaken during that reporting period. This information should be submitted using the attached Project Product Report form. For an electronic version please email Lori Rainville lrainville@srf.org.

- Closing Narrative Report: As principal investigator for the project, you will be expected to submit a closing narrative report when you send the Foundation your final product.

- Post-Project Report: As principal investigator for the project, one year after you submit your final product, you will be required to send us a narrative report in which you will assess the impact of your project.

The attached Grant Reporting Requirements document describes all reporting requirements in detail. We value these reports highly. Any future grant payments or considerations are based in part on an assessment of these reports.

Again, congratulations on the grant, and I look forward to following the progress of your work.

Sincerely,

Mark Steinmeyer
Senior Program Officer

cc: Nicole Pobuta

Attachments: Memorandum of Grant Conditions
Grant Reporting Requirements
Project Product Report

05120119.doc

SMITH RICHARDSON FOUNDATION, INC.

Memorandum of Grant Conditions

The grants of the Foundation are subject to the following conditions:

1. All funds advanced by the Foundation under this grant shall be used by the Grantee for the purposes for which granted, and for no other purposes without the consent, in writing, of the Foundation; and any unexpended balance remaining in the hands of the Grantee upon the termination of the grant as hereinafter provided shall, at the option of the Foundation, be returned to it.

2. The use of the funds advanced by the Foundation under this grant shall be commenced with reasonable promptness. If, in the sole opinion of the Foundation, there shall not have been a bona fide commencement of the use of such funds within a reasonable time following the date of acceptance hereof by the Grantee, the Foundation reserves the right to rescind all or any part of the grant hereby made.

3. The funds hereby granted shall be advanced by the Foundation to the Grantee as mutually agreed upon by the Grantee and the Foundation.

4. All funds granted by the Foundation to the Grantee, until used or disbursed by the Grantee for the purposes of said grant, shall not be used for any other purpose and shall not be invested in any manner which would jeopardize or impair in anywise their availability for use by the Grantee for the purposes hereinabove provided, or for return to the Foundation as hereinafter provided.

5. Any public announcement of this grant shall be made pursuant to the approval of the Foundation. The use of the Foundation's name in relation to this grant in any medium of publication, whether written or oral, shall be pursuant to the approval of the Foundation.

6. Advance notice must be given to the Foundation if the principal investigator/s will be changed during the course of this grant. The Foundation reserves the right to terminate this grant and seek a return of funds if the Foundation, in its sole discretion, concludes that the change in personnel will substantially affect the quality of the product of the grant.

7. The governing body of the Grantee shall adopt a resolution accepting and agreeing to accept this grant on the terms stated herein and shall furnish to this Foundation a certified copy of said resolution. In the alternative, the chief executive officer of the Grantee, or a duly authorized representative thereof, shall advise the Foundation, in writing, that he is fully authorized to accept this grant in accordance with its terms on behalf of the Grantee and that in so advising the Foundation he thereby so accepts this grant. In addition, the chief executive officer of the Grantee shall submit to the Foundation a copy of the Grantee's Internal Revenue Service exemption letter under Section 501(c)(3) of the Internal Revenue Code of 1954, as amended (the "Code"). In addition, the chief executive officer of the

SMITH RICHARDSON FOUNDATION, INC.

Memorandum of Grant Conditions

7. (Continued)
Grantee shall advise, in writing, that the Grantee's tax-exempt status under Section 501(c)(3) of the Code is unrevoked and in full force and effect. In the event the Grantee is not a private foundation, the chief executive officer shall submit to the Foundation any communication from the Internal Revenue Service so indicating, together with written advice to the effect that such status is unrevoked and in full force and effect.

8. In the event the Grantee is a private foundation within the meaning of Section 509(a) of the Code, the Grantee shall cooperate fully with the Foundation in the exercise of its "expenditure responsibility" under Section 4945(h) of the Code, and to this end the Grantee, in addition to the requirements set forth above, shall:

   I. repay any portion of the amount granted which is not used for the purposes of the grant;

   II. submit full and complete reports, not less than annually, on the manner in which the funds are spent and the progress made in accomplishing the purposes of the grant;

   III. maintain records of receipts and expenditures and make its books and records available to the Foundation at all times; and

   IV. not use any of the funds:

      a. to carry on propaganda, or otherwise to attempt to influence legislation (within the meaning of Section 4945(d)(1) of the Code);

      b. to influence the outcome of any specific public election, or to carry on directly or indirectly any voter registration drive (within the meaning of Section 4945(d)(2) of the Code);

      c. to make any grant which does not comply with the requirements of Section 4945(d)(3) or (4) of the Code; or

      d. to undertake any activity for any purpose other than one specified in Section 170(c)(2)(B) of the Code.

SMITH RICHARDSON FOUNDATION
Grant Reporting Requirements

The Foundation requires all grant recipients to submit regular progress reports to the Foundation during the term of a grant. A grant remains open until the Foundation receives the final product and all reporting requirements are satisfied. It is the principal investigator's responsibility to see that reports are submitted in a timely manner. All reports should be addressed to the appropriate program officer and include the eight-digit grant number that appears in all official correspondence from the Foundation. It is preferred that all reports be sent to the program officer via e-mail.

Narrative Progress Reports are due from the principal investigator every six months until the project's final product is submitted to the Foundation and the dissemination plan is completed. Each narrative report should include a description of the progress that has been made on the work plan that was set forth in the grant proposal; a discussion of any preliminary findings; and a listing of any product/dissemination activity (e.g., papers, articles, or briefings) during the six-month reporting period.

Closing Narrative Report is due from the principal investigator when the final product (e.g., published book, report, monograph, article, edited volume) is submitted to the Foundation. This report should include a description of all the work that was accomplished during the grant term; a summary of the project's findings; and a listing of the product/dissemination activities related to the project. Immediately after the final product is available, two copies of the product should be sent to the program officer at the Foundation.

Post-Project Narrative Report is due from the principal investigator twelve months after the final product for the grant is submitted to the Foundation or twelve months after the principal objective of the grant is achieved. This report should describe the project's impact. It should include information on how the project's findings were disseminated and how they were received.

Financial Reports are due from the institution every twelve months until all the grant funds have been expended for the project. Reports must duplicate the format of the budget that appeared in the proposal, showing the amount awarded and expended against the line items listed in the proposal budget.

Final Financial Report is due from the institution after all the funds awarded in the grant have been expended. This report should include the original budget that was submitted with the grant proposal and a separate listing of how the funds were disbursed. Any differences between the proposal budget and the actual budget should be explained. If there are **remaining funds** after the project is complete, the principal investigator should contact his/her program officer, who will direct any further action.

Grant Modifications:

- **No Cost Extensions**: If a project's work plan is delayed or cannot be completed within the original grant term, principal investigators should submit a written request to their program officer for a no-cost extension. All requests should include a new project end date.

- **Grant Reprogramming**: If, during the term of a grant, there is a need to alter the work plan that was set forth in the original proposal, the principal investigator should submit a written request to his/her program officer.

- **Reallocations**: If it becomes necessary to alter the proposed disbursement of the grant funds within existing or to new budgetary line items, the principal investigator should submit a written request, including a new budget, to his/her program officer.

PROJECT PRODUCT REPORT

Organization:

Principal Investigator:

Project Name:

SRF Grant No:

Reporting Period:

☐ No products were produced during this reporting period
☐ Products produced during this reporting period are listed below

Principal Product [e.g., major book, commissioned paper, monograph]			
Title	Author	Publication	Date

Published Spin-off Product [e.g., op-ed, article, lecture, written testimony, conference proceedings]			
Title	Author	Publication	Date

Non-Published Spin-off Product [e.g., speech, presentation, conference appearance]			
Title	Author	Venue	Date

credo
CENTER FOR RESEARCH ON EDUCATION OUTCOMES

December 2, 2011

Mark Steinmeyer
Senior Program Officer
Smith Richardson Foundation
60 Jessup Road
Westport, CT 06880

Dear Mark,

Thank you for the exciting news about the Board's review of the CREDO proposal to conduct further research on charter school impacts. My team shares my appreciation of all your efforts to shape the proposal to earn a favorable outcome with the Board. And that effort continues even after the meeting. I am writing to respond to the three conditions that the Board placed on their approval of our submission.

1. **The Board requested that we submit the report and any subsequent articles to peer review.** We understand the board's concern about subjecting new research methods to the traditional peer review process. For that reason, we have written a methods paper that demonstrates parallel modeling using student fixed effects and the new Virtual Control Records (VCRs) for the *Economics of Education Review*. That paper underwent a full peer review process and we have responded to questions posed by the reviewers in our final submission. The article was accepted for publication and will appear in January 2012. 1

   We have always used peer reviewers as part of our quality control process before we release any of our reports, even though the reports are prepared for a policy rather than a purely scholarly audience. This is a practice that we intend to continue with the research proposed for this grant.

   We typically select two types of peer reviewers – those that are focused on the policy landscape and those that are quantitative methodologists. We engage our reviewers in the early stages of our projects and do not call upon them to provide substantive advice to the work until the review stage. (We have other colleagues who fill that role.)

TEL: (650) 725-3431 FAX: (650) 723-1687
CREDO, Stanford University, Stanford CA, 94305-6010
credo.stanford.edu

On the analytic side, we plan to produce a technical report that covers the methods employed in the various phases of the proposed study. The technical report will be drafted prior to the drafting of the full policy report, so the technical peer reviews can occur prior to the release of the full report. The draft technical report will cover:

- research design,
- description of the data
- matching methods, approach to the analysis
- analytic findings
- supporting documentation

We generally ask reviewers to return comments in a fortnight, to allow time for clarification of our methods or additional analysis if needed.

We share a draft of the full report with the policy-oriented and analytic peers and ask for a three-week response. The full report will not have as much detail about the analysis, since the orientation and audience will be more focused on the results and policy implications.

And for any policy-related articles that we prepare for publication, there will be the standard peer review process of our submissions. Those submissions will require a tight summary of the analytic approaches and will need to meet the peer-review standards of each journal.

2. **Determine the level of significance.** The question posed by the board concerns whether the results of the proposed analysis can be statistically significant. This issue was raised after the release of the 2009 report and was addressed at that time. Specifically, the issue is whether the use of VCRs violates the assumption of equal variance of VCRs and the charter school students to whom they were matched. Indeed, if these distributions were not equivalent, it would lead to bias in the estimated parameter for charter school effects. However, CREDO demonstrated empirically that the variances in the two groups – VCRs and charter students – were not significantly different. The results appear in the table below:

	Reading		Math	
	Mean	Standard Error	Mean	Standard Error
VCR	-.0327747	.0012930	-.0911137	.0012946
Charter Students	-.0328184	.0012949	-.0914350	.0012969

The CREDO team is keenly aware that new methodological approaches can raise concerns. We are firmly in favor of advancing the science of evaluation, along with our primary mission of informing public policy thinking. It is towards the end of furthering the scientific knowledge base that we have included in our research design multiple tests of the same data. Those tests, including details on the effect sizes and standard errors will be incorporated in the technical report discussed in #1 above and made available to the public.

3. **Subject the VCR results to sensitivity tests using other analytic methods.**

As you know, the proposal already contains the commitment to test the primary model results with a parallel student fixed effect model. Further, we agreed to isolate the schools that have been studied in recent lottery studies and conduct a stand-alone VCR analysis with them for comparison purposes. There are three other options, and they are discussed separately below.

a. *Conduct the elementary matches without reference to a 3rd grade test score and compare the results of a baseline-free VCR analysis with our normal one.*
I am puzzled by the suggestion of this option. Part of the justification for using academic growth as an outcome is that it provides a clean separation of the marginal contributions of schools and a student's starting endowments. Including baseline scores in the match criteria serves another function: it creates a reference that allows the differences in family background to be reflected without directly influencing the study of school effects. This is so because achievement, taken alone, is a combination of schooling and family effects. We take a prior score to reflect equivalent levels of endowments, arising from potentially different combinations of the two sources of influence. It is the inclusion of priors that justifies our assertion that selection bias is not a factor in the VCR analysis; without a prior score, we are completely open to the possibility that our charter

school sample is compared to a differently endowed student. We will perform the comparison, only in the context of "layering up" from a base model that is clearly flawed.

We have another possibility to address the concern of included treatment effects. There is an assumption that charter elementary school all grow grade by grade, starting with either Kindergarten or Kindergarten and first grade and expanding one grade per year. This assumption however is not valid in all cases: in some cases budget pressures or extreme demand prompt elementary schools to open multiple grades at once. In these exceptional cases, the students joining the school in tested grades can provide a "pure charter" effect against a VCR comparison. They are exactly the cases we studied in the *Economics of Education Review* but limited to elementary schools that meet the inclusion criterion of opening with multiple tested grades.

Based on a rough run through our data, it appears we have about 150 elementary schools that fit the "fast start" scenario. We can do a sensitivity analysis on the full complement of elementary school grades, divided into fast and slow growers, and see if the groups differ and if maturity since opening contributes to the explanation.

b. *Employ a propensity score matching approach.*
Since we last talked, the team has spent a fair amount of time reviewing some of the propensity Score Matching (PSM) literature. The use of PSM comes with a variety of possible conditions: include/exclude a prior test score, include more or fewer variables, a range of case-control ratios and the suitability of replacements for attrition in the matches. It is easy to estimate the feasibility of matches under these various combinations: with each additional match factor, the feasible set of matches declines. In a head-to-head comparison against the match criteria used in the VCR, in all cases the result is that the VCR produces tighter matches because we use the most extensive set of criteria. So we would a priori expect the results of other methods to have higher variance simply due to excluded variables. It is less clear what the potential is for bias in the charter school estimator with other PSM approaches, since it would depend on the degree of association between the omitted variable and the binary for charter school enrollment.

TEL: (650) 725-3431 FAX: (650) 723-1687
CREDO, Stanford University, Stanford CA, 94305-6010
credo.stanford.edu

But the VCR differs in another way that we expect will be favorable to PSM matching. We exploit the student-level data in a unique way and restrict the possible matches to the schools in each charter school's feeder pool, whereas most PSM methods use an unrestricted pool of matches either across a district or state. (There's nothing magical in our approach, it is merely that we do preliminary analysis of the migration of students prior to setting our parameters for matching.) With the feeder school limitations, the coupling of VCR-eligible students and charter students is direct and material – the candidates for a VCR actually attend the counterfactual schools the charter students would otherwise attend. We consider this fact to reduce the influence of unobserved factors, though we cannot say they are eliminated entirely. (We think we get closer with matching on baseline scores.) The result is that we expect PSM to be a weaker comparison method because the matches are inevitably more loosely aligned. We intend to demonstrate this important difference.

On the modeling side, we will include a PSM test mimicking the work of Mark Berends in his work on charter school impacts as part of the sensitivity analysis to appear in the Technical Report.

I have prepared a revised budget to show the additional resource requirements to accomplish this last step of the sensitivity analysis. It adds another $34,163 to the budget, for a new total of $471,800.

Please let me know if you have questions or other ideas.

Sincerely,

[redacted]

ME Raymond

TEL: (650) 725-3431 FAX: (650) 723-1687
CREDO, Stanford University, Stanford CA, 94305-6010
credo.stanford.edu

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

David D Oakes, M.D.
CHAIR, PANEL ON NON-MEDICAL HUMAN SUBJECTS (650) 725-8013

Certification of Human Subjects Approvals

Date: October 31, 2012

To: Eric A. Hanushek, PhD, Hoover Institution
Jennifer Hodges, Margaret E. Raymond PhD, Devora Davis B.S., Edward Cremata MS, James Lynn Woodworth M.Ed., Kathleen Dickey MS, Kristina Lawyer M.A., Emily Elizabeth Harris Peltason BA, Yohannes G Negassi M.A.

From: David D Oakes, M.D., Administrative Panel on Human Subjects in Medical Research

Protocol The National Charter School Study

Protocol ID: 15913 **IRB Number:** 6208 (Panel: 8)

The IRB approved human subjects involvement in your research project on 10/31/2012. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval;** a GCRC study, you must obtain GCRC approval; a VA study, you must obtain VA R and D Committee approval; and if a contract is involved, it must be signed.'

The expiration date of this approval is 10/31/2015 at Midnight. If this project is to continue beyond that date, you must submit an updated protocol in advance for the IRB's re-approval. If this protocol is used in conjunction with any other human use it must be re-approved. Proposed changes to approved research must be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at http://humansubjects.stanford.edu.)

All continuing projects and activities must be reviewed and re-approved on or before Midnight of the expiration date. The approval period will be less than one year if so determined by the IRB. It is your responsibility to resubmit the project to the IRB for continuing review and to report the completion of the protocol to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, or other entities. (See Policy on Retention of and Access to Research Data at http://stanford.edu/dept/DoR/rph/2-10.html.)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.

[redacted]

David D Oakes, M.D., Chair

Approval Period: 10/31/2012 THROUGH 10/31/2015
Review Type: EXPEDITED - CONTINUING REVIEW
Funding: Robertson Foundation - Hoover Inst , SPO: 104313
Expedited Under Category: 5
Assurance Number: FWA00000935 (SU)

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

David D Oakes, M.D.
CHAIR, PANEL ON NON-MEDICAL HUMAN SUBJECTS (650) 725-8013

Certification of Human Subjects Approvals

Date: October 31, 2012

To: Eric A. Hanushek, PhD, Hoover Institution
Jennifer Hodges, Margaret E. Raymond PhD, Devora Davis B.S., Edward Cremata MS, James Lynn Woodworth M.Ed., Kathleen Dickey MS, Kristina Lawyer M.A., Emily Elizabeth Harris Peltason BA, Yohannes G Negassi M.A.

From: David D Oakes, M.D., Administrative Panel on Human Subjects in Medical Research

Protocol The National Charter School Study

Protocol ID: 15913 **IRB Number:** 6208 (Panel: 8)

The IRB approved human subjects involvement in your research project on 10/31/2012. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval;** a GCRC study, you must obtain GCRC approval; a VA study, you must obtain VA R and D Committee approval; and if a contract is involved, it must be signed.'

The expiration date of this approval is 10/31/2015 at Midnight. If this project is to continue beyond that date, you must submit an updated protocol in advance for the IRB's re-approval. If this protocol is used in conjunction with any other human use it must be re-approved. Proposed changes to approved research must be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at http://humansubjects.stanford.edu.)

All continuing projects and activities must be reviewed and re-approved on or before Midnight of the expiration date. The approval period will be less than one year if so determined by the IRB. It is your responsibility to resubmit the project to the IRB for continuing review and to report the completion of the protocol to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, or other entities. (See Policy on Retention of and Access to Research Data at http://stanford.edu/dept/DoR/rph/2-10.html.)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.

[redacted]

David D Oakes, M.D., Chair

Approval Period: 10/31/2012 THROUGH 10/31/2015
Review Type: EXPEDITED - CONTINUING REVIEW
Funding: Smith Richardson Foundation
Expedited Under Category: 5
Assurance Number: FWA00000935 (SU)

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

David D Oakes, M.D.
CHAIR, PANEL ON NON-MEDICAL HUMAN SUBJECTS (650) 725-8013

Certification of Human Subjects Approvals

Date: October 31, 2012

To: Eric A. Hanushek, PhD, Hoover Institution
Jennifer Hodges, Margaret E. Raymond PhD, Devora Davis B.S., Edward Cremata MS, James Lynn Woodworth M.Ed., Kathleen Dickey MS, Kristina Lawyer M.A., Emily Elizabeth Harris Peltason BA, Yohannes G Negassi M.A.

From: David D Oakes, M.D., Administrative Panel on Human Subjects in Medical Research

Protocol The National Charter School Study

Protocol ID: 15913 **IRB Number:** 6208 (Panel: 8)

The IRB approved human subjects involvement in your research project on 10/31/2012. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval;** a GCRC study, you must obtain GCRC approval; a VA study, you must obtain VA R and D Committee approval; and if a contract is involved, it must be signed.'

The expiration date of this approval is 10/31/2015 at Midnight. If this project is to continue beyond that date, you must submit an updated protocol in advance for the IRB's re-approval. If this protocol is used in conjunction with any other human use it must be re-approved. Proposed changes to approved research must be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at http://humansubjects.stanford.edu.)

All continuing projects and activities must be reviewed and re-approved on or before Midnight of the expiration date. The approval period will be less than one year if so determined by the IRB. It is your responsibility to resubmit the project to the IRB for continuing review and to report the completion of the protocol to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, or other entities. (See Policy on Retention of and Access to Research Data at http://stanford.edu/dept/DoR/rph/2-10.html.)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.

David D Oakes, M.D., Chair

Approval Period: 10/31/2012 THROUGH 10/31/2015
Review Type: EXPEDITED - CONTINUING REVIEW
Funding: Emerson Collective
Expedited Under Category: 5
Assurance Number: FWA00000935 (SU)

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

David D Oakes, M.D.
CHAIR, PANEL ON NON-MEDICAL HUMAN SUBJECTS (650) 725-8013

Certification of Human Subjects Approvals

Date: October 31, 2012

To: Eric A. Hanushek, PhD, Hoover Institution
Jennifer Hodges, Margaret E. Raymond PhD, Devora Davis B.S., Edward Cremata MS, James Lynn Woodworth M.Ed., Kathleen Dickey MS, Kristina Lawyer M.A., Emily Elizabeth Harris Peltason BA, Yohannes G Negassi M.A.

From: David D Oakes, M.D., Administrative Panel on Human Subjects in Medical Research

Protocol The National Charter School Study

Protocol ID: 15913 **IRB Number:** 6208 (Panel: 8)

The IRB approved human subjects involvement in your research project on 10/31/2012. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval;** a GCRC study, you must obtain GCRC approval; a VA study, you must obtain VA R and D Committee approval; and if a contract is involved, it must be signed.'

The expiration date of this approval is 10/31/2015 at Midnight. If this project is to continue beyond that date, you must submit an updated protocol in advance for the IRB's re-approval. If this protocol is used in conjunction with any other human use it must be re-approved. Proposed changes to approved research must be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at http://humansubjects.stanford.edu.)

All continuing projects and activities must be reviewed and re-approved on or before Midnight of the expiration date. The approval period will be less than one year if so determined by the IRB. It is your responsibility to resubmit the project to the IRB for continuing review and to report the completion of the protocol to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, or other entities. (See Policy on Retention of and Access to Research Data at http://stanford.edu/dept/DoR/rph/2-10.html.)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.

David D Oakes, M.D., Chair

Approval Period: 10/31/2012 THROUGH 10/31/2015
Review Type: EXPEDITED - CONTINUING REVIEW
Funding: Fairbanks Foundation
Expedited Under Category: 5
Assurance Number: FWA00000935 (SU)

J Econ Growth (2012) 17:267–321
DOI 10.1007/s10887-012-9081-x

Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation

Eric A. Hanushek · Ludger Woessmann

Published online: 14 July 2012

Abstract We develop a new metric for the distribution of educational achievement across countries that can further track the cognitive skill distribution within countries and over time. Cross-country growth regressions generate a close relationship between educational achievement and GDP growth that is remarkably stable across extensive sensitivity analyses of specification, time period, and country samples. In a series of now-common microeconometric approaches for addressing causality, we narrow the range of plausible interpretations of this strong cognitive skills-growth relationship. These alternative estimation approaches, including instrumental variables, difference-in-differences among immigrants on the U.S. labor market, and longitudinal analysis of changes in cognitive skills and in growth rates, leave the stylized fact of a strong impact of cognitive skills unchanged. Moreover, the results indicate that school policy can be an important instrument to spur growth. The shares of basic literates and high performers have independent relationships with growth, the latter being larger in poorer countries.

Keywords Cognitive skills · Long run growth · Causation and identification · School quality · Educational achievement

E. A. Hanushek (✉)
Hoover Institution, Stanford University, Stanford, CA 94305-6010, USA
e-mail: hanushek@stanford.edu
URL: www.hanushek.net

E. A. Hanushek · L. Woessmann
CESifo, University of Munich, Munich, Germany

E. A. Hanushek
NBER, National Bureau of Economic Research (NBER), Cambridge, MA, USA

L. Woessmann
Ifo Institute for Economic Research, University of Munich, Poschingerstr. 5, 81679 Munich, Germany
e-mail: woessmann@ifo.de
URL: www.cesifo.de/woessmann

1 Introduction

Schooling and human capital investments have been a central focus of development policy, but doubts have arisen as disappointments with results grow. Nowhere is this more apparent than in the case of growth policy, where schooling investments have not appeared to return the economic outcomes promised by theoretical growth models.[1] Prior analyses into the specification of empirical cross-country growth models lead to a warranted skepticism about the identification of causal growth effects. Our analysis of newly developed measures of skill differences based on international tests of math and science suggests, however, that one of the most significant problems underlying these prior concerns is the valid measurement of human capital across countries. We find that accurately measuring differences in educational achievement, which we refer to simply as cognitive skills, dramatically improves our ability to explain variations in long-run growth across countries. Moreover, while having limitations in macroeconomic applications, a set of microeconometric approaches can be employed in the cross-country setting to rule out many of the common concerns that undermine causal interpretations.

As a simple summary observation, world policy attention today focuses on the lagging fortunes of Sub-Saharan Africa and of Latin America. Considerably less attention goes to East Asia, and, if anything, East Asia is proposed as a role model for the lagging regions. Yet to somebody contemplating development policy in the 1960s, none of this would be so obvious. Latin America had average income exceeding that in Sub-Saharan Africa and the Middle East and North Africa regions, and both of these exceeded East Asia (see Appendix Table 8).[2] Further, Latin America had schooling levels that exceeded those in the others, which were roughly equal. Thus, on the basis of observed human capital investments, one might have expected Latin America to pull even farther ahead while having no strong priors on the other regions. The unmistakable failure of such expectations, coupled with a similar set of observations for separate countries in the regions, suggests skepticism about using human capital policies to foster development. But, this skepticism appears to be more an outgrowth of imperfect measurement of human capital investments than an empirical reality.

The measurement issues become apparent when we introduce direct measures of cognitive skills from international tests of math and science into the growth picture. The entire picture changes. Figure 1 plots regional growth in real per capita GDP between 1960 and 2000 against average test scores after conditioning on initial GDP per capita in 1960.[3] Regional annual growth rates, which vary from 1.4 % in Sub-Saharan Africa to 4.5 % in East Asia, fall on a straight line with an $R^2 = 0.985$. But, school attainment, when added to this regression, is unrelated to growth-rate differences. Figure 1 suggests that, conditional on initial income levels, regional growth over the last four decades is completely described by differences in cognitive skills.

In the upsurge of empirical analyses of why some nations grow faster than others since the seminal contributions by Barro (1991, 1997) and Mankiw et al. (1992), a vast literature of cross-country growth regressions has tended to find a significant positive association between

[1] See, for example, Pritchett (2006).

[2] Japan was significantly ahead of the rest of the East Asia region, but its exclusion does not change the regional ordering (see Appendix Table 8).

[3] Regional data come from averaging all countries with available data in a region. The 50 countries are not chosen to be representative but instead represent the universe of countries that participated in international tests and had available the requisite economic data. Still, Appendix A shows that the average 1960 incomes for all countries in each region are quite similar to those for our subset of countries. The division of Europe into three regions illustrates the heterogeneity within OECD countries, but a combined Europe also falls on the line in Fig. 1.

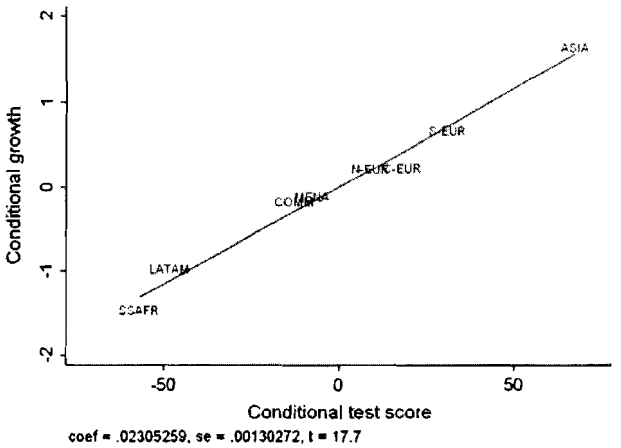


Fig. 1 Cognitive Skills and Growth across World Regions. *Notes*: Added-variable plot of a regression of the average annual rate of growth (in percent) of real GDP per capita in 1960–2000 on the initial level of real GDP per capita in 1960 and average test scores on international student achievement tests. Authors' calculations. See Table 8 for a list of the countries contained in each world region. Region codes: East Asia and India (ASIA), Central Europe (C-EUR), Commonwealth OECD members (COMM), Latin America (LATAM), Middle East and North Africa (MENA), Northern Europe (N-EUR), Southern Europe (S-EUR), Sub-Saharan Africa (SSAFR)

quantitative measures of schooling and economic growth.[4] But, all analyses using average years of schooling as the human capital measure implicitly assume that a year of schooling delivers the same increase in knowledge and skills regardless of the education system. For example, a year of schooling in Peru is assumed to create the same increase in productive human capital as a year of schooling in Japan. Equally as important, this measure assumes that formal schooling is the primary source of education and that variations in the quality of nonschool factors have a negligible effect on education outcomes.

In this paper, we concentrate directly on the role of cognitive skills. This approach was initiated by Hanushek and Kimko (2000), who related a measure of educational achievement derived from the international student achievement tests through 1991 to economic growth in 1960–1990 in a sample of 31 countries with available data. They found that the association of economic growth with cognitive skills dwarfs its association with years of schooling and raises the explanatory power of growth models substantially. Their general pattern of results has been duplicated by a series of other studies over the past 10 years that pursue different tests and specifications along with different variations of skills measurement.[5]

But should we interpret the tight relationship between cognitive skills and growth as reflecting a causal relationship that can support direct policy actions? Questions about the identification of underlying causal effects in cross-country growth models have existed for a long time and go beyond just the impact of human capital. Beginning with Levine and Renelt (1992), plentiful evidence of the general sensitivity to alternative samples and specifications has convinced many that cross-country empirical models are not fruitful policy investigations. In terms of schooling, Bils and Klenow (2000) provide convincing evidence of the

[4] For extensive reviews of the literature, see Topel (1999), Krueger and Lindahl (2001), Pritchett (2006), and Hanushek and Woessmann (2008). The robustness of the association is highlighted by the extensive analysis by Sala-i-Martin et al. (2004): Of 67 explanatory variables in growth regressions on a sample of 88 countries, primary schooling turns out to be the most robust influence factor (after an East Asian dummy) on growth in GDP per capita in 1960–1996.

[5] Detailed discussion of these studies is available in Hanushek and Woessmann (2011a).

endogeneity of school attainment in growth models. Further, it is unclear to what extent prior attempts to deal with endogeneity, such as the panel data approaches of Barro (1997) and Vandenbussche et al. (2006), have been successful in a setting where the dominant information is found in the cross-country variation.[6] Perhaps the strongest evidence on causality has been related to the importance of fundamental economic institutions using identification through historical factors (Acemoglu et al. 2001, 2005), but this has not yielded clear advice about the kinds of feasible policies that will lead to national payoffs, and it itself has been subject to question (Glaeser et al. 2004).

When estimating the effect of cognitive skills on growth, the main causality concerns relate to reverse causality and to omitted country variables such as inherent difference in nations' culture and economic institutions that are correlated both with economic growth and with cognitive skills or their determinants. We assess these issues from a number of angles with the objective of narrowing the range of threats to a causal interpretation. Of course, it is virtually impossible to identify causality in a thoroughly convincing manner given the limited observations underlying cross-country growth models. Each approach we employ deals with one or more common concerns such as the influence of cultural differences, faulty measurement of cognitive skills, or simple reverse causality. But each relies upon strong maintained hypotheses that may or may not be completely persuasive.

Our analysis, while building on Hanushek and Kimko (2000), provides new evidence about the potential causal interpretation of the cognitive skills-growth relationship. The development of a new data series on cognitive skills (Sect. 3), expanded to 64 countries for some analyses, permits approaches to estimation not previously possible. We are able to improve on the underlying measurement, to increase the country observations to a broader range of development experiences, to extend the period of observed long-term growth to 1960–2000 (and to 2007 in some specifications), and to add both a longitudinal and a within-country distributional dimension to the database. We begin by showing that the relationship between cognitive skills and economic growth is extraordinarily robust to alternative samples defined by different time periods and sets of countries and to different specifications of the skills measure and of the growth relationship (Sect. 4).

The core of the paper applies a series of approaches to identification of causal parameters now common in microeconometric studies to the macroeconomic analysis of growth, although the application to cross-country estimation remains difficult. The important new analyses include estimation with instrumental variables (Sect. 5) and consideration of intertemporal changes in growth rates within countries (Sect. 7). More recent U.S. data also permit important refinements to the analysis of cognitive skills on the labor market earnings of immigrants (Sect. 6) previously introduced in Hanushek and Kimko (2000), including the specification of full difference-in-differences models.

Each of our three approaches deals with a particular class of reverse causation or omitted variables. By identifying skill variation stemming from institutional school policies in the countries, the instrumental-variable models highlight the role of schools while addressing issues of simple reverse causality and of inherent cultural difference across nations that might be related to attitudes and performance in learning. By focusing on U.S. labor-market outcomes for immigrants, the difference-in-differences approach deals not only with reverse causality but also with the possibility that cultural differences or economic institutions of national economies may be correlated with favorable educational outcomes. By using the intertemporal dimension of our new database, our longitudinal analysis of changes in growth rates eliminates stable country-specific factors in a general way in the spirit of country fixed

[6] Aghion et al. (2009) approach causality by relying on within-country variation.

effects. In each of the three investigations, we explicitly describe the assumptions that are key to interpreting the results. Importantly, the different approaches rely on different assumptions, guard against different threats to identification, and would fail for different reasons.

A related aspect of these separate causal investigations is the pinpointing of a specific policy role for improved school quality. While variations in cognitive skills can arise from various influences—families, culture, health, and ability—the instrumental-variable results indicate that schools, and in particular institutional structures of school systems, are one way for improvement available to policy makers. This conclusion is reinforced by how country of schooling—U.S. versus home country—is important for identifying individual skills in the immigrant analysis.

A final issue addressed is that average test scores do not adequately reflect the range of policy options facing a nation. Specifically, one could institute policies chiefly directed to the lower end of the cognitive distribution, such as the Education for All initiative, or one could aim more at the top end, such as the focused technological colleges of India. In an analysis enabled by the detailed country-specific distributional dimension of our new micro database, we are able to go beyond simple mean difference in scores and provide the first estimates of how growth is affected by the distribution of skills within countries and how it might interact with the nation's technology (Sect. 8). We find improving both ends of the distribution to be beneficial and complementary. The importance of the highly skilled is even more important in developing countries that have scope for imitation than in developed countries that are innovating.

2 A simple growth model with cognitive skills

We begin with a very simple growth model: a country's growth rate (g) is a function of the skills of workers (H) and other factors (X) that include initial levels of income and technology, economic institutions, and other systematic factors. Skills are frequently referred to simply as the workers' human capital stock. For simplicity in Eq. (1), we assume that H is a one-dimensional index and that growth rates are linear in these inputs, although these are not important for our purposes.[7]

$$g = \gamma H + \beta X + \varepsilon \tag{1}$$

It is useful at this stage to understand where the skills (H) might come from. As discussed in the extensive educational production function literature (Hanushek 2002), these skills are affected by a range of factors including family inputs (F), the quantity and quality of inputs provided by schools (qS), individual ability (A), and other relevant factors (Z) which include labor market experience, health, and so forth as in:

$$H = \lambda F + \phi(qS) + \eta A + \alpha Z + \nu \tag{2}$$

The schooling term combines school attainment (S) and its quality (q).

Human capital is nonetheless a latent variable that is not directly observed. To be useful and verifiable, it is necessary to specify the measurement of H. The vast majority of existing

[7] The form of this relationship has been the subject of considerable debate and controversy. As we write it, it can be consistent with both basic endogenous growth models such as Lucas (1988), Romer (1990), and Aghion and Howitt (1998) and neoclassical growth models such as Mankiw et al. (1992). We allow for conditional convergence in the empirical specifications, and the parameters estimated suggest very long transitional periods from any perturbation off of a balanced growth path. We generally cannot adequately distinguish among alternative forms of the underlying growth process. While considering the growth implications of various policy changes, we can, however, investigate directly the sensitivity of GDP projections to the alternative models (see Hanushek and Woessmann 2011b).

theoretical and empirical work on growth begins—frequently without discussion—by taking the quantity of schooling of workers (S) as a direct measure of H.

A more compelling alternative is to focus directly on the cognitive skills component of human capital and to measure H with test-score measures of mathematics, science, and reading achievement.[8] The use of measures of educational achievement has a number of potential advantages. First, they capture variations in the knowledge and ability that schools strive to produce and thus relate the putative outputs of schooling to subsequent economic success. Second, by emphasizing total outcomes of education, they incorporate skills from any source—families, schools, and ability. Third, by allowing for differences in performance among students with differing quality of schooling (but possibly the same quantity of schooling), they open the investigation of the importance of different policies designed to affect the quality aspects of schools.[9]

3 Consistent international measures of cognitive skills

This analysis starts with the development of new measures of international differences of cognitive skills derived from educational achievement tests. We would ideally have measures of the skills for workers in the labor force, but our measures of cognitive skills come from data on testing for students who are still in school. This creates a trade-off: incorporating more recent testing has the potential advantages of improved assessments and observations on a greater number of countries but it also weights any country measures more toward students and less toward workers.[10] We begin with an expansive inclusion of more recent tests but then investigate the impact of this choice through extended robustness checks that take more restrictive choices.

The measures developed here extend those developed in Hanushek and Kimko (2000) to add new international tests, more countries, and intertemporal and within-country dimensions. They also deal with a set of problems that remained with the early calculations.[11]

[8] Some researchers have suggested that test scores should be thought of as a measure of school quality (q), leading to use of test scores times years of schooling as a measure of H, but this ignores the influence of family factors and other elements of Eq. (2) that have been shown to be very important in determining cognitive skills.

[9] Some recent work has introduced the possibility that noncognitive skills also enter into individual economic outcomes (see importantly Bowles et al. 2001; Heckman et al. 2006; Cunha et al. 2006). Hanushek and Woessmann (2008) integrate noncognitive skills into the interpretation of general models such as above and show how this affects the interpretation of the parameter on school attainment and other estimates. While there are no agreed-upon measures of noncognitive skills, at the aggregate level they might well be incorporated in "cultural differences," something that we address in the analysis below.

[10] The reliance on schooling-based measures of skills also makes it clear why it is not possible to employ panel data estimation even though tests are spread across almost four decades for some nations. Any panel study would require measuring the cognitive skills of the labor force at different points in time, something that is not possible with the sporadic measurement of student skills. Only one international test—the International Assessment of Adult Literacy—has suggested the possibility of panel estimation across countries because it has tested adults rather than students (see Coulombe and Tremblay 2006). Nonetheless, such analysis requires very strong assumptions about the mapping of observed age patterns of skills onto changes in labor force skills over time. Further, most of the variance in growth and in test scores is found across countries, not across time for individual countries—suggesting that panel data do not deal effectively with the most acute estimation issues. As shown in Appendix B, the testing has involved voluntary participation by a time-varying group of countries in tests that assess varying subject matters and grade/age ranges of students.

[11] The correlation across the common 30 countries of the new test measures developed here and those in Hanushek and Kimko (2000) is 0.83. Appendix B assesses the importance for growth modeling of the differences in their measures and those developed here.

Between 1964 and 2003, twelve different international tests of math, science, or reading were administered to a voluntarily participating group of countries (see Appendix Table 10). These include 36 different possible scores for year-age-test combinations (e.g., science for students of grade 8 in 1972 as part of the First International Science Study or math of 15-year-olds in 2000 as a part of the Programme on International Student Assessment). Only the United States participated in all possible tests.

The assessments are designed to identify a common set of expected skills, which were then tested in the local language. It is easier to do this in math and science than in reading, and a majority of the international testing has focused on math and science. Each test is newly constructed, usually with no effort to link to any of the other tests.

We wish to construct consistent measures at the national level that will allow comparing, say, math performance of 13-year-olds in 1972 to that in 2003. This would permit us to compare performance across countries, even when they did not each participate in a common assessment, as well as track performance over time. It would also provide the ability to aggregate scores across different years, ages, and even subjects as appropriate. The details of this construction along with the final data are found in Appendix B, and here we simply sketch the methodology. Because the test distribution is normal within the OECD sample, our construction of aggregate country scores focuses on transformations of the means and variances of the original country scores in order to put them each into a common distribution of outcomes.

Comparisons of the difficulty of tests across time are readily possible because the United States has participated in all assessments and because there is external information on the absolute level of performance of U.S. students of different ages and across subjects. The United States began consistent testing of a random sample of students around 1970 under the National Assessment of Educational Progress (NAEP). By using the pattern of NAEP scores for the U.S. over time, it is possible to equate the U.S. performance across each of the international tests.

The comparison of performance of other countries to the U.S. requires a distance metric for each test. Each assessment has varying country participation and has different test construction so that the variance of scores for each assessment cannot be assumed to be constant. Our approach is built on the observed variations of country means for a group of countries that have well developed and relatively stable educational systems over the time period.[12] We create the "OECD Standardization Group" (OSG) by using the thirteen OECD countries that had half or more of the relevant population attaining a secondary education in the 1960s (the time of the first tests). For each assessment, we then calibrate the variance in country mean scores for the subset of the OSG participating to the variance observed on the PISA tests in 2000 (when all countries of the OSG participate). The identifying assumption of this approach is that the *variance* in the mean performance among a group of relatively stable education systems does not change substantially over time.

By combining the adjustments in levels (based on the U.S. NAEP scores) and the adjustment in variances (based on the OECD Standardization Group), we can directly calculate standardized scores for all countries on all assessments. Each age group and subject is normalized to the PISA standard of mean 500 and individual standard deviation of 100 across OECD countries. We can then aggregate scores across time, ages, and subjects as we desire.

[12] The development of aggregate scores by Hanushek and Kimko (2000) and by Barro (2001) assumed that the test variances across assessments were constant, but there is no reason for this to be the case. Our approach is in the spirit of Gundlach et al. (2001).

The international testing protocols have evolved over time so that recent assessments employ careful sampling rules, restrictions on the extent of any student exclusions, and modern psychometric testing procedures, while earlier testing less consistently met current standards. This variation in testing quality potentially affects parts of our analysis, because the earlier (but poorer) tests relate to relevant members of the labor force during our period of observation for economic growth. The more recent testing involves students not observed to be in the labor force. As a result, most of the estimation relies upon an assumption that the average scores for a country tend to be relatively stable over time and that the differences among countries are a good index of the relative skill differences of the workforces. This assumption is partially tested below, and, while there are some observed score changes, the overall rankings of countries show considerable stability. For the 693 separate test observations in the 50 countries employed in our growth analysis, 73 % of the variance falls between countries. The remaining 27 % includes both changes over time in countries' scores and random noise from the testing. Our averaging procedure will minimize the noise component at the cost of obscuring any differences over time for each country. In Sect. 7 below, we use the intertemporal variation in scores for the subset of countries with enough observations to estimate the systematic changes as opposed to test noise. For the 15 countries employed in the analysis of score trends in that section, 85 % of the variance lies between countries, and the remaining 15 % within countries will be more heavily the result of systematic trends in scores.

The assessments give cognitive skill measures for tested students. Thus, exclusion rates (say for handicapped children) or differential student enrollment and attendance could affect the estimation. Direct investigation of these issues, at least for tests since 1995 when reporting is sufficient, indicates that the growth analysis is not affected by variations in testing.[13]

4 Stability of the cognitive skills-growth relationship

The basic growth model in Eq. (1) is estimated for the 50 countries with cognitive-skill and economic data over the period 1960–2000. Cognitive skills are measured by the simple average of all observed math and science scores between 1964 and 2003 for each country, although we test the sensitivity of the results to inclusion of varying time periods and subsets of tests. The income data come from version 6.1 of the Penn World Tables (Heston et al. 2002), while the data on years of schooling are an extended version of the Cohen and Soto (2007) data.[14]

We have concentrated on issues surrounding the measurement of cognitive skills, but other questions have recently been raised about the accuracy and reliability of both economic and schooling data. In Appendix D, we investigate the impact of using the latest Barro and Lee (2010) data on school attainment and of substituting the most recent Penn World Table data (version 7.0), which provides additional evidence confirming our basic results through the period up to 2009. Because neither of these alternatives materially affects our results, we simply combine these sensitivity studies in the Appendix.

[13] Hanushek and Woessmann (2011c) show that average tests can be affected by these: greater exclusions and higher enrollment rates are correlated with higher scores. Nonetheless, the variations caused by these factors are orthogonal to growth rates, so they do not bias our estimated skill parameters. The problems are potentially more severe with the earliest tests, but the reported information for these is insufficient for any analysis.

[14] Because we need comparable data on economic growth over the 1960–2000 period, all former communist countries are eliminated even if they have test measures. Appendix B provides details on the country sample, and Appendix C provides descriptive statistics for each of the analyses below.

Table 1 Years of schooling versus cognitive skills in growth regressions

	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[d]	(8)[e]	(9)[f]
Cognitive skills		2.015	1.980	1.975	1.933	1.666	1.265	1.239	1.985
		(10.68)	*(9.12)*	*(8.28)*	*(8.29)*	*(5.09)*	*(4.06)*	*(4.12)*	*(7.83)*
Years of schooling 1960	0.369		0.026	0.024	0.025	0.047	0.004	−0.049	−0.090
	(3.23)		*(0.34)*	*(0.78)*	*(0.29)*	*(0.54)*	*(0.05)*	*(0.66)*	*(1.02)*
GDP per capita 1960	−0.379	−0.287	−0.302	−0.298	−0.298	−0.255	−0.351	−0.310	−0.879
	(4.24)	*(9.15)*	*(5.54)*	*(6.02)*	*(5.04)*	*(3.12)*	*(6.01)*	*(5.73)*	*(3.39)*
No. of countries	50	50	50	50	52	50	47	45	50
R^2 (adj.)	0.252	0.733	0.728	0.728		0.706	0.784	0.797	0.637

Notes Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Regressions include a constant. Test scores are average of math and science, primary through end of secondary school, all years. Absolute *t*-statistics in parentheses

[a] Measure of years of schooling refers to the average between 1960 and 2000

[b] Robust regression including the two outliers of Botswana and Nigeria (using *rreg* robust estimation command implemented in Stata)

[c] Specification includes dummies for the eight world regions depicted in Fig. 1

[d] Specification includes additional controls for openness and property rights

[e] Specification includes additional controls for openness, property rights, fertility, and tropical location

[f] GDP per capita 1960 measured in logs

The central finding of the statistical analysis is the importance of cognitive skills in explaining international differences in long-run growth rates. As a comparison to prior cross-country analyses, the first column of Table 1 presents estimates of a simple growth model with school attainment.[15] While this model explains one-quarter of the variance in growth rates, adding cognitive skills increases this to three-quarters of the variance. The test score is strongly significant with a magnitude that is unchanged by whether initial school attainment in 1960 is excluded (col. 2) or included (col. 3).

School attainment is not statistically significant in the presence of the direct cognitive-skill measure of human capital. This does not change when attainment is measured as the average between 1960 and 2000 (col. 4), rather than at the beginning of the period. This finding, of course, does not mean that schooling is irrelevant. Measured skills are closely related to schooling—a point we emphasize below—but life-cycle skill accumulation depends upon the learning earlier in life. We measure achievement at various points during primary and secondary education. Even if tertiary schooling is simply additive, knowledge at earlier points in education will strongly influence the ultimate skill accumulation when students enter the labor force. But, as James Heckman and his colleagues have emphasized, there is a dynamic complementary of investments such that further schooling has a larger impact on skills if it builds on a larger base developed earlier (Cunha and Heckman 2007). The simple point is that "skill begets skill through a multiplier process" (Cunha et al. 2006, p. 698), such that additional attainment has a lessened impact if built upon lower basic skills.[16] It does suggest

[15] While not the focal point of this analysis, all specifications include GDP per capita in 1960, which provides consistent evidence for conditional convergence, i.e., countries with higher initial income tend to grow more slowly.

[16] Relatedly, a variety of people place extra weight on tertiary education (e.g., Ehrlich 2007). However, without building on strong basic skills, such investment appears to have little extra value. In analysis of growth across both developed and developing countries, tertiary education has little added value in explaining economic growth after consideration of measured cognitive skills on international tests with the exception that U.S. investments in higher education have signaled increased growth (Hanushek and Woessmann 2011b).

that simply investing in further schooling without ensuring commensurate improvements in cognitive skills does not lead to economic returns.

One standard deviation in test scores (measured at the OECD student level) is associated with a two percentage point higher average annual growth rate in GDP per capita across 40 years. (This effect is equivalent to one percentage point per country-level standard deviation, thus making it virtually identical to the more limited estimates in Hanushek and Kimko 2000). Such impacts are clearly large in substantive economic terms, and below we provide alternative perspectives on the magnitude of these effects.

The remaining columns of Table 1 provide alternative perspectives on these basic results. Estimating the model with regression techniques robust to outliers yields virtually identical coefficient estimates to those including Nigeria and Botswana, the two significant outlier countries in the growth equation (col. 5).[17] Because the robust model assigns essentially zero weight to these two observations, they are dropped from the remaining models. Including fixed effects for the eight world regions depicted in Fig. 1 (so that no between-region variation in test scores is used in the estimation) reduces the estimated test effect to 1.7 (col. 6).

Columns 7 and 8 consider economic institutions. Acemoglu et al. (2001, 2005) argue that historic factors surrounding the colonization of nations affected economic institutions and that we can thus isolate the causal impact of institutions on growth. We add institutional differences for openness of the economy and security of property rights (col. 7)[18] and for these plus fertility rates and location in the tropics (col. 8) into our growth models.[19] These reduce the estimated test-score effect to around 1.25, but the effect of cognitive skills remains strongly statistically significant. On the other hand, Glaeser et al. (2004) argue that the colonists brought human capital in addition to knowledge of good societal institutions and that it is more likely that better human capital led both to the development of good institutions and higher economic growth. This latter perspective highlights the difficulty of using measures that reflect economic institutions from near the end of the observed growth period and that might better be thought of as outcomes of growth itself. In their spirit, we interpret the reduced estimates of test scores in columns 7 and 8 as a lower bound on the true effect, since the institutional measures include any direct effects of cognitive skills on the development of good institutions.[20] Additionally (not shown), the stock of physical capital per adult in

Footnote 16 continued
The difficulty is that it is impossible to identify the impact of higher education as opposed to other unmeasured determinants of economic growth in the United States.

[17] The specific robust regression technique reported is Stata's *rreg* command, which eliminates gross outliers with Cook's distance measure greater than one and iteratively down weights observations with large absolute residuals. The OLS estimate of the test effect in the 52-country sample is 1.752 (t-statistic 5.75). Nigeria and Botswana each participated only in a single international test.

[18] The measure of openness is the Sachs and Warner (1995) index reflecting the fraction of years between 1960 and 1998 that a country was classified as having an economy open to international trade, based on five factors including tariffs, quotas, exchange rate controls, export controls, and whether or not a socialist economy. Following Acemoglu et al. (2001), the measure of security of property rights is an index of the protection against expropriation risk, averaged over 1985–1995, from Political Risk Services, a private company which assesses the risk that investments will be expropriated in different countries.

[19] Openness and security of property rights enter the model (jointly) significantly; fertility and tropical location do not.

[20] A separate analysis of institutions has evolved in the developed countries of the OECD. These countries all have strong property rights, open economies, and institutions that generally favor economic growth that explain differences between developed and developing countries. In order to explain differences within developed countries, considerable analysis has gone into how various regulations of labor and product markets within the OECD affect economic growth (e.g., Nicoletti and Scarpetta 2003 on product market regulations and Cingano et al. (2010) on labor market regulation). Nonetheless, none of the explicit measures of economic

1960 does not enter the basic growth model significantly and does not affect the test-score coefficient.

Finally, the precise specification of the growth model is the subject of considerable debate within macroeconomics. While there are many nuances of the argument, it can be framed as a simple contrast. The endogenous growth model indicates that increases in human capital can lead to permanent differences in growth rates, because a better-educated workforce leads to a larger stream of new ideas that produces technological progress at a higher rate (e.g., Lucas 1988 or Romer 1990). By contrast, in the augmented neoclassical growth model, changes in human capital lead to higher steady-state levels of income but do not affect the long-run growth path (e.g., Mankiw et al. 1992). Our estimates, which include the level of initial GDP per capita, allow for conditional convergence, but it is difficult to distinguish between temporary "catch-up" growth and long-run differences in growth in the empirical model. Those who favor the neoclassical model, however, favor estimation that includes the log of initial income. Column 9 shows these results, where the impact of cognitive skills on growth is little changed from the linear alternative. Thus, the results do not appear to be the result of the specific empirical model that is estimated.

While the estimated effect of test scores varies some across these different specifications, the cognitive-skill coefficients are always very significant and the variation is quite limited: A move of one standard deviation of individual student performance translates into 1.2–2.0 percentage points difference in annual growth rates, other things equal. How much is one standard deviation in performance? The difference between the U.S. average and the top performers on the PISA tests is approximately 0.4 standard deviations, while the difference between the average Mexican student and the rest of the OECD was approximately one standard deviation.

Two other important questions that relate to interpretation arise. The first set of issues is whether the sample of countries or years of observation heavily influences the results, thus implying that the results are potentially driven by other, unmeasured factors. The second is whether the specific measure of cognitive skills drives the estimates.

Table 2 provides the matrix of estimated cognitive-skill coefficients across different samples of observations. The columns consider sample sensitivity and concentrate on whether the overall results are driven by specific subsets of countries, which might indicate that the cognitive-skill measures simply proxy for other facets of the economies. The top row focuses on the average of all observed math and science scores—as presented previously—while, as explained below, the second row relies on just lower-secondary-school scores which may be a more reliable measure of skill differences. Each entry comes from a separate regression that includes GDP per capita in 1960 and school attainment.

The first two comparisons (col. 2–3 and col. 4–5) present evidence on whether cognitive skills are more or less important in developed countries. The first comparison divides the estimation into the 23 OECD countries and 27 non-OECD countries, while the second comparison divides countries into above and below the median level of per-capita GDP in 1960. The statistically significant difference of high-income and lower (below median)-income countries indicates that developing countries are somewhat more affected by cognitive skills than developed countries.[21] This larger impact of skills in low-income countries is consistent

Footnote 20 continued
regulations identified in this work is useful at explaining the large differences in long-run growth rates among OECD countries (Hanushek and Woessmann 2011b). Thus, at least for the developed countries, omission of this expanded set of economic institutions does not appear to bias our growth estimates.

[21] While not shown, the school attainment measures are insignificantly related to growth even among the developing countries where the levels are low and where there is considerable cross-country variance.

Table 2 Sensitivity of estimated effects of cognitive skills to the sample of countries and time periods

Country/year sample	(1) Full	(2) OECD	(3) Non-OECD	(4) High-income[a]	(5) Low-income[a]	(6) W/o East Asia	(7) 1960–1980	(8) 1980–2000	(9) 1980–2000[b]	(10) Score-schooling outliers[c]	(11) Score-schooling core[c]
Test-score specification											
All math and science	1.980	1.736	2.056	1.287	2.286	1.301	1.522	2.996	3.782	1.888	2.175
	(9.12)	*(4.17)*	*(6.10)*	*(5.37)*	*(6.98)*	*(4.90)*	*(4.29)*	*(9.42)*	*(3.11)*	*(7.81)*	*(3.47)*
Only lower secondary	1.759	1.646	1.792	1.040	2.083	1.137	1.407	2.580	4.386	1.673	1.887
	(9.22)	*(4.02)*	*(6.19)*	*(4.70)*	*(7.44)*	*(4.82)*	*(4.56)*	*(8.88)*	*(4.49)*	*(7.83)*	*(3.45)*
No. of countries	50	23	27	25	25	40	50	50	25	25	25

Notes Reported numbers are the coefficient on test scores in each model specification. Dependent variable: Unless noted otherwise, average annual growth rate in GDP per capita, 1960–2000. Control variables: Initial GDP per capita, initial years of schooling, and a constant. Test scores: Unless noted otherwise, average of math and science, primary through end of secondary school, all years. Absolute *t*-statistics in parentheses

[a] Countries above/below sample median of GDP per capita 1960

[b] Test scores refer only to tests performed until 1984

[c] Countries with largest (outliers)/smallest (core) residuals when regressing years of schooling on test scores

with the arguments by Glaeser et al. (2004) that nearly all poor countries in 1960 were dictatorships, some of which developed better societal institutions as an outcome of growth rather than a cause. The countries that did better in terms of growth were those with higher human capital, supporting the larger coefficient on human capital in the poor countries. Nonetheless, variations in math and science skills remain very important in distinguishing among growth rates of the developed countries.

A portion of the influence of cognitive skills comes from the high growth of East Asian countries. As shown in column 6, excluding the ten East Asian countries lowers the estimated impact of math and science skills to 1.3, but it remains highly significant in the remaining countries. In other words, the overall estimates, while influenced by the East Asian growth experience, are not simply identifying the high growth—high test-score position of East Asia, which would raise the possibility that the growth relationships might be driven by other factors that were simply correlated with East Asian test performance.

The growth estimates are meant to identify long-run factors, but the sample period of 1960–2000 includes sub-periods of world stagnation, fast growth, and financial crises. Some have suggested, for example, that the observed growth rates are dominated by the early-period growth explosion of East Asia and that this changed considerably with the financial crises of the late 1990s (Ramirez et al. 2006). Our results (col. 7 and 8) indicate, however, a consistent impact of cognitive skills across the period that, if anything, has grown stronger in the second half of our observations. Indeed, the estimated impact doubles in the most recent period, consistent with various arguments that, at least for the U.S. and OECD countries, the importance of skills has increased (Murnane et al. 1995; Katz and Autor 1999; Goldin and Katz 2008).

The analysis has relied on assessments given throughout the period of economic observation. This choice is made to maximize the number of countries and to include the more precise testing of recent periods, but it raises questions of reverse causality. If greater growth provides added resources that can be used to improve schools and test scores, our estimates could suffer from simultaneity bias. One direct set of estimates addresses this issue.[22] The same impact on 1980–2000 growth is found when we restrict the test scores to measures obtained before 1985 (available for only 25 countries), i.e., when we use test scores nearly fully pre-dating the growth period (col. 9). In fact, the point estimate for cognitive skills becomes substantially larger in this specification. By using pre-determined test scores, this specification excludes the possibility of simple reverse causation. The conclusion that simple reverse causation is not driving the results is reinforced in analyses using data updates that extend the economic series to 2009 (Appendix D).[23] The possibility of reverse causation from economic growth to test scores is also unlikely because additional educational spending (which might become affordable with higher growth) does not systematically relate to better test scores (e.g., Hanushek 2002).

Levels of schooling and cognitive scores are correlated across our sample ($r = 0.62$), in part because of the differences between developed and developing countries. Still, in many cases countries with similar years of schooling have very different test scores (see Appendix Fig. 6). The separation of the impact of cognitive skills from that of school attainment in our estimation relies upon information where these two diverge, and it might be a peculiar set of countries in terms of growth where the pattern of school attainment and skills varies most.

[22] A second set of analyses directly addresses the resource question and finds that international tests are not driven by differences in resources across countries. See the review in Hanushek and Woessmann (2011a).

[23] Results using only test scores that pre-date the analyzed growth period (not shown) are also robust when combined with our other robustness checks pursued in Table 1.

The final two columns divide countries based on deviations of cognitive scores from school attainment. Specifically, the "score-schooling outliers" are the 25 countries with the largest residuals when test scores are regressed on attainment, and the "score-schooling core" are the 25 with the smallest residuals. Interestingly, the relationship between cognitive skills and growth is virtually the same across these two samples, revealing that the results are not driven by "peculiar" countries in the production of cognitive skills.

The preceding results hold looking across columns, but the pattern also obtains for the alternative measures of test scores. The estimated coefficients using only lower-secondary-school math and science scores are systematically a little smaller than those from all scores, which may reflect attenuation bias when using fewer test observations in the construction of the cognitive-skill measure, but there are no changes in patterns across any of the columnar comparisons. This test-score measure excludes any test in primary schooling or in the final year of secondary education. Test scores at the end of the secondary level, which combine the knowledge accumulated over primary and secondary schooling, may be most relevant for the labor force, but, at the same time, the duration of secondary education differs across countries, so that tests performed in the final year of secondary schooling may not be as readily comparable across countries. Further, given differing school completion rates, tests in the final year of secondary schooling may produce samples with differential selectivity of test takers. Yet neither the primary-school tests nor the tests in the final secondary year are crucial for the results.[24]

Table 3 provides more detail on sensitivity to the measure of cognitive skills, comparing several additional plausible alternatives for the aggregation of scores, including using math, science, and reading scores separately. We also provide breakdowns by OECD and non-OECD countries, although this breakdown makes little qualitative difference, and we concentrate on the variations in aggregate test information found in the table rows.

Results are qualitatively the same when using only scores on tests performed since 1995 (row A). These recent tests have not been used in previously available analyses and are generally viewed as having the highest standard of sampling and quality control. Likewise, results are robust to using tests scores since 1995 for just lower secondary grades (row B).

A drawback of using only the more recent tests is that such an approach requires a strong version of the assumption that test performance is reasonably constant over time, because it relates test performance measured since 1995 to the economic data for 1960–2000. To make sure that higher previous economic growth is not driving the measured test performance, the test-score measure used in row C disregards all tests since the late 1990s. Our results turn out to be robust, with a point estimate on the test-score variable that is significantly higher (although the sample is reduced to 37 countries). Our results are also robust to using the average early test scores as an instrument for the average of all test scores in a two-stage least-squares regression, in order to utilize only that part of the total test-score measure that can be traced back to the early test scores (row D). In sum, the results do not appear to be driven by either early or late test scores alone.

The remainder of the table investigates different combinations of the math, science, and reading tests. While we were concerned about the reliability of the reading tests and thus have focused on math and science, the use of reading tests provides similar results in the growth models (rows E-G). In a specification that enters the different subjects together (panel H), the three are always jointly significant at the 1 % level and higher, even though the science effect gets smaller and the reading effect loses significance in the joint model.

[24] Hanushek and Woessmann (2011c) also provide direct analysis of potential biases from test exclusions and enrollment rates and find that they are not an important concern.

Table 3 Sensitivity of estimated effects of cognitive skills to the measurement of skills

Country sample	(1) Full	(2) OECD	(3) Non-OECD	No. of countries
Test-score specification				
(A) Only since 1995	1.814	1.473	1.850	47
	(9.91)	*(3.80)*	*(6.74)*	
(B) Only lower secondary since 1995	1.644	1.379	1.657	47
	(9.57)	*(3.49)*	*(6.48)*	
(C) Only until 1995	3.156	1.377	3.668	37
	(6.57)	*(1.93)*	*(4.44)*	
(D) Early as instrument for average[a]	2.341	1.212	2.915	37
	(7.71)	*(1.98)*	*(5.80)*	
(E) Only math	2.009	1.529	2.063	47
	(8.98)	*(4.62)*	*(5.81)*	
(F) Only science	1.576	1.769	1.556	50
	(7.00)	*(3.28)*	*(4.48)*	
(G) Only reading	2.351	1.616	2.529	46
	(6.21)	*(3.33)*	*(3.68)*	
(H) All subjects entered jointly				41
Math	1.662	2.270	1.882	
	(3.69)	*(2.97)*	*(1.97)*	
Science	1.007	−2.414	1.270	
	(2.34)	*(1.62)*	*(1.92)*	
Reading	−0.793	1.333	−1.457	
	(1.15)	*(1.44)*	*(0.94)*	

Notes Reported numbers are the coefficient on test scores in each model specification. Dependent variable: Average annual growth rate in GDP per capita, 1960–2000. Control variables: GDP per capita 1960, years of schooling 1960, and a constant. Test scores: Unless noted otherwise, average of math and science, primary through end of secondary school, all years. Absolute t-statistics in parentheses

[a] 2SLS with average of test scores until 1995 as instrument for average of all test scores

The overall picture from this sensitivity analysis is that the estimated effect of cognitive skills on growth is quite robust to a range of samples, specifications, and measurements. This finding contrasts sharply with many previous analyses that use years of schooling as the human capital measure, beginning with Levine and Renelt (1992) and continuing through Pritchett (2006). But of course the similarity of findings, while ruling out some specification and measurement issues, cannot guard against all plausible threats to the identification of causal growth relationships.

The main theme of this paper is that cross-sectional growth regressions using existing variation across countries provide stylized facts about long-term development but that their interpretation may be hampered by endogeneity biases. Endogeneity of cognitive skills could, for example, arise because nations with conditions favorable to economic growth also produce high test performance. This correlation could arise because cultural factors, historically good economic institutions, variations in health status, or any other set of factors that lead to strong economic performance might also be systematically related to high cognitive skills. Indeed, it does not matter whether such relationships are causal or purely associational. If these factors are omitted from the growth estimation, they will tend to bias the coefficient on cognitive skills. Likewise, as suggested previously, there might be reverse causality if economic growth facilitates investments in the school system or increases family resources that improve cognitive skills.

The following three sections approach the interpretation of these stylized facts about growth from different viewpoints. Each approach to dealing with potential interpretative problems is clearly inconclusive, but each does work to eliminate specific sets of concerns. These further analyses also highlight the potential for policies based upon improved school quality.

5 Variations in cognitive skills driven by schools: instrumental variable models

One general concern in cross-country investigations is that cultural features influence both economic behavior and school outcomes. Even if the cognitive skills-growth relationship is causal, the results presented so far would only be relevant for school policy if the variation in cognitive skills emanating from school policies is in fact related to economic growth. As noted, cognitive skills are likely to depend not only on formal schooling but also on non-school factors such as families, peers, and ability. Therefore, it is important to establish any links with school policy levers.

One means of addressing the set of issues is to use measures of the institutional structure of the school systems as instruments for the cognitive-skill measure, thereby focusing only that part of the international variation in cognitive skills that can be traced back to international differences in school systems. We use several institutional features—notably the existence of external exit exam systems, the share of privately operated schools, the impact of varying Catholic church history, the centralization of decision-making, and relative teacher pay—that have been shown in the literature on international educational production to be associated with student achievement (see Hanushek and Woessmann 2011a for a review and evaluation of the micro evidence).

While other school policies such as those surrounding educational spending levels may well be endogenous to the growth process, these institutional features can plausibly be assumed uncorrelated with the regression disturbances of our growth models. First, many educational institutions such as the existence and extent of private schooling reflect long-standing polices embedded in education law and thus are not outcomes of the growth process per se (see, for example, the review of private schooling across countries in Glenn and De Groof 2002).[25] Second, while there have been some trends in these institutions—such as the slow movement toward decentralizing school decision-making—there is no suggestion that this reflects either growth or other systematic differences in cultural and economic systems.[26] Third, there is empirical support from the literature on educational production that these institutional effects on student learning are robust to including regional fixed effects in cross-country analyses, to within-country analyses, and to the use of historical instruments (see Woessmann 2003a; West and Woessmann 2010; Hanushek and Woessmann 2011a). These results suggest that institutional impacts are not driven by cultural differences and do not suffer directly from reverse causality.

[25] If private school attendance rates are also related to the religious composition of countries, this instrument might still by problematic because of religious impacts on economic behavior; see Barro and McCleary (2003). However, evidence in West and Woessmann (2010) shows that, if anything, such impacts would be likely to lead to an underestimation of the effect of private school competition.

[26] Glenn and De Groof (2002, p. 267), note that "there has been in most Western democracies a slow but very marked shift in the allocation of responsibility for the organization and control of education, in the public as well as the nonpublic education sector, through decentralization of various aspects of decision-making to the local school community." The cross-country details suggest no obvious political or cultural differences in these trends.

External exit exam systems are a device to increase accountability in the school system that has been repeatedly shown to be related to better student achievement (see Bishop 2006 for a review).[27] The first specification reported in Table 4 uses the share of students in a country who are subject to external exit exams as an instrument for our measure of cognitive skills in the growth regression. The first-stage results confirm a statistically significant association between external exit exams and cognitive skills. The effect of cognitive skills on economic growth in the second stage of the instrumental variable (IV) estimation is statistically significant and close to the OLS estimate.[28] However, the relatively low F-statistic of the instrument in the first stage indicates the possibility of a weak instrument problem. Instruments that are only weakly correlated with the endogenous explanatory variable may actually increase estimation bias and compromise the reliability of the conventional asymptotic approximations used for hypothesis testing. Thus, we also report estimates based on the modification of the limited information maximum likelihood (LIML) estimator by Fuller (1977), but the results are hardly affected.[29] While the confidence band of the conditional likelihood ratio test proposed by Moreira (2003) and Andrews et al. (2007) gets large at the upper end in this specification, difference from zero still reaches significance at the 10% level.[30]

Because initial years of schooling are insignificant in the growth model once test scores are controlled for (both in the OLS and in the IV specification), another possibility is to include years of schooling as a second instrument for test scores.[31] This approach is also suggested by the prior model as long as cognitive skills are a measure of human capital in Eq. (1). Specification (2) of Table 4 reveals that years of schooling are significantly associated with test scores in the first stage, and the first-stage F-statistic increases substantially. The Sargan test does not reject the overidentifying restrictions of the model, suggesting that, if external exit exams are a valid instrument, years of schooling are also valid. Both the 2SLS and the Fuller estimates, as well as inference based on Moreira confidence bands, confirm that schooling-induced differences in cognitive skills are significantly related to economic growth.

School choice, as measured by the share of privately operated schools in a system, consistently shows a positive association with student achievement in OECD countries (see the review in Woessmann et al. (2009), along with West and Woessmann (2010)) and provides an additional instrument. In our sample, the share of private enrollment in a country is significantly positively associated with cognitive skills in the first stage of our IV model (specification (3) of Table 4).[32] The second-stage estimate of the growth model confirms our

[27] Data on external exit exams are available for 43 countries in Woessmann et al. (2009), who update Bishop (2006)'s collection from reviews of comparative-education studies, educational encyclopedia, government documents, background papers, and interviews with national representatives. The measure refers roughly to the mid-1990s, but exam regimes are relatively stable over time for countries.

[28] The Durbin–Wu–Hausman test does not reject the exogeneity of cognitive skills at conventional levels.

[29] Fuller's modification of the LIML estimator is more robust than 2SLS in the presence of weak instruments and performs relatively well in the simulations by Hahn et al. (2004). We set the user-specified constant (Fuller 1977's alpha) to a value of one, but our results are hardly affected if we set alpha to four.

[30] Likewise, the Anderson–Rubin χ^2 statistic (3.06) of this just-identified model indicates significance at the 8% level. Note that the LIML estimators, around which the Moreira bands are centered, differ from the reported 2SLS estimates only in the third digit in all our models.

[31] School attainment will also be affected by enrollment in higher education, which is not explicitly modeled. The results here suggest that international differences in cognitive skills still remain the dominant factor in growth.

[32] The data on private enrollment as percentage of total enrollment in general secondary education are from UNESCO (1998) and refer to 1985, the earliest year with consistent data. For greater consistency of the time

Table 4 From schooling institutions to cognitive skills to economic growth: instrumental variable estimates

	(1)	(2)	(3)[a]	(4)	(5)[a]	(6)
Second stage						
2SLS						
Cognitive skills	2.151	2.023	2.978	2.207	3.914	1.749
	(2.73)	*(5.81)*	*(5.84)*	*(6.54)*	*(4.17)*	*(5.77)*
Catholic share in 1970				0.003		
				(0.01)		
Fuller (1) modification of LIML						
Cognitive skills	2.121	2.022	2.969	2.197	3.797	1.753
	(3.01)	*(5.94)*	*(5.93)*	*(6.64)*	*(4.17)*	*(5.92)*
Moreira 95 % confidence band						
Cognitive skills	[−3.888, 19.871]	[1.190, 2.868]	[1.734, 4.343]	[1.465, 3.093]	[2.063, 7.006]	[0.865, 2.525]
p-value	*(0.100)*	*(0.001)*	*(0.0004)*	*(0.0001)*	*(0.0000)*	*(0.007)*
First stage (dependent variable: cognitive skills)						
External exit exam system	0.286	0.286				
	(2.01)	*(2.01)*				
Initial years of schooling		0.176	0.137	0.186	0.065	0.161
		(4.11)	*(4.19)*	*(4.32)*	*(2.06)*	*(3.05)*
Private enrollment share			0.520			
			(2.36)			
Centralization (share) of decisions on organization of instruction					−0.941	
					(3.24)	
Catholic share in 1900				2.301		
				(2.15)		
Relative teacher salary						0.188
						(2.19)
Catholic share in 1970				−2.801		
				(2.46)		

Table 4 continued

	(1)	(2)	(3)[a]	(4)	(5)[a]	(6)
No. of countries	43	43	20	50	18	34
Centered R^2	0.752	0.753	0.791	0.743	0.590	0.819
First-stage F-statistic	4.04	10.28	12.15	10.60	13.35	6.94
Sargan statistic		0.033	0.158	0.193	0.011	0.377
p-value		*(0.856)*	*(0.691)*	*(0.661)*	*(0.917)*	*(0.540)*
Durbin–Wu–Hausman χ^2 test	0.034	0.003	0.113	0.479	4.744	0.081
p-value	*(0.855)*	*(0.957)*	*(0.737)*	*(0.489)*	*(0.029)*	*(0.776)*

Notes Dependent variable (of the second stage): average annual growth rate in GDP per capita, 1960–2000. Control variables: Initial GDP per capita, initial years of schooling, and a constant. Test scores are average of math and science, primary through end of secondary school, all years. t-statistics in parentheses unless otherwise noted

[a] Dependent variable: average annual growth rate in GDP per capita, 1980–2000; sample of OECD countries

previous results—schooling-induced differences in cognitive skills are significantly related to economic growth. Again, the Sargan test does not reject the validity of the overidentifying restrictions, and the Durbin–Wu–Hausman test presents no evidence of endogeneity of the cognitive-skill measure. Results are also very similar without years of schooling as a second instrument.

An additional way to exploit the effect of private competition is to use the historical origins of international variation in the size of the private school sector.[33] In particular, West and Woessmann (2010) show that the opposition of 19th century Catholic church doctrine to state schooling provides a natural experiment in that the share of Catholics in a country's population in 1900 is associated with the share of privately operated schools in its current school system, even after controlling for current Catholic shares. In this spirit, specification (4) of Table 4 uses the Catholic share in 1900 as an instrument for the cognitive skill measure and controls for the Catholic share in 1970 in both stages of the IV model to ensure that results are not driven by effects of religious affiliation per se on cognitive skills and on economic growth (see Hanushek and Woessmann (2012) for details). The first stage shows that historical Catholic shares are indeed positively related to cognitive skills, whereas the opposite is true for modern Catholic shares (which are insignificant in the second stage). This IV specification, which we can estimate for our full sample of 50 countries, confirms a significant effect of cognitive skills on economic growth that is very close to the OLS estimate. The F-statistic of the instruments in the first stage is just above 10, and LIML estimates and Moreira bands confirm that the result is not driven by weak instruments problems.[34]

A further institutional feature regularly shown to be positively associated with student achievement is the extent to which schools (or at least local decision-makers) are autonomous to make their own decisions about the organization of instruction (see Woessmann 2003b). Specification (5) of Table 4 shows that the share of decisions on the organization of instruction that are made at the central government level is significantly negatively associated with our cognitive-skill measure. The second-stage estimators, robust to potentially weak instruments, confirm the significantly positive effect of cognitive skills on economic growth.[35]

Finally, given the crucial importance of teacher quality in educational production, our final IV specification uses the *relative* position of teacher salaries in the income distribution of a country as an instrument for cognitive skills (see Hanushek and Woessmann 2012 for details). Following Dolton and Marcenaro-Gutierrez (2011), we use teacher salaries relative to per-capita income as a proxy for the overall quality of the teaching force in a cross-country

Footnote 32 continued
spans, the dependent variable in this specification is economic growth in 1980–2000; results are robust to using growth in 1960–2000. Given that the results from the educational production literature mostly refer to the sample of OECD countries, we restrict the analysis to the OECD sample, for which 20 observations are available.

[33] The idea for this IV specification, as well as the IV specification using relative teacher salaries, was first presented by Hanushek and Woessmann (2012) in the context of analyzing the Latin American growth puzzle; that reference also provides additional detail on these two IV specifications.

[34] While the F-statistic of the instrument in the first stage is lower (at 4.61) in a model that does not use years of schooling as a second control, such a specification also confirms the significant main result and is also robust to LIML estimation and to the use of Moreira bands.

[35] Data on the percentage of decisions on the organization of instruction in public lower secondary education taken at the central level of government are available in Organisation for Economic Co-operation and Development (1998), available only for 1998. The IV results are very similar without using years of schooling as a second instrument, and the F-statistic of the excluded instrument is already above 10. In this specification, the estimated growth effect is even larger than the OLS estimate. Note, though, that the Fuller estimate is already closer to the OLS estimate, and the Moreira confidence bands include the OLS and other IV estimates.

perspective, which targets the point in the overall "ability" distribution from which a country draws its teacher population and circumvents issues of salary *levels* being themselves related to growth.[36] Again, the results confirm the significant growth effect of cognitive skills, in the same order of magnitude as the OLS estimates, as well as the robustness to LIML estimation and Moreira bands (specification (6) of Table 4).

One potential worry about the exogeneity of our instruments is that the institutional features of school systems may be correlated with economic institutions, which are themselves correlated with economic growth. To test whether this affects our identification, we add the two measures of differences in economic institutions that tend to enter most robustly in growth regressions—openness and security of property rights—to our IV models (remembering, however, our prior reservations about the distinct possibility that these economic institutions capture part of the human capital effect). Our basic result is unaffected. In fact, the measures of economic institutions do not enter significantly (individually or jointly) in any of the IV models except column 2, and the effect of cognitive skills remains significant in all specification except for the just-identified model (1). The point estimates for cognitive skills are hardly affected except for model (2), where—in line with the OLS results of Table 1—it is reduced to 1.1 (and to 1.3 in model (4)), similar to our OLS estimate of the lower bound for the effect.

The results suggest that improvements in cognitive skills generated in the school system—through institutional features affecting school quality—lead to higher long-run growth of economies. There are obvious limits of cross-country regressions with small data samples, and these are particularly salient in IV specifications. Caution is appropriate in interpreting IV results for our relatively small samples of countries and employing the aggregate nature of the institutional measures, but these cautions also make the statistical significance, reasonable precision, and quantitative robustness of the results based on quite different instruments even more striking.[37]

A significant concern remains, however. The institutional characteristics of the school system might still be related to important unmeasured aspects of economic institutions (either causally or correlationally). Nonetheless, any such problems must go beyond the traditional measures of differences in economic institutions that are commonly employed.

6 Comparing the impacts of U.S. and home-country education on the U.S. labor market

An alternative approach for assessing the causal importance of schools and of our measured skill differences on economic outcomes relies on microdata on earnings differences within a single labor market—the U.S. labor market. This strategy, first proposed by Hanushek and Kimko (2000), looks within a single labor market, thereby explicitly holding constant the quality of economic and cultural factors affecting the operations of the economy and focusing on whether measured cognitive skills directly relate to productivity.[38] Following a difference-

[36] The teacher salary data, based on OECD and UNESCO surveys, refer to teacher salaries at the top of the experience scale in 2003.

[37] The IV results hold when employing several of the institutional instruments jointly in the specification, but only one of them tends to capture statistical significance in the joint specifications.

[38] This analysis extends the original work in Hanushek and Kimko (2000) in several ways. By placing the analysis in the framework of a difference-in-differences model, it compares the earnings of late immigrants just to early immigrants from the same country; it dramatically expands both the sample of workers and the number of countries of origin for immigrants; it uses better test information for the comparisons; and it

in-differences strategy, we can compare the returns to skills of immigrants schooled in their country of origin to those of immigrants from the same country schooled within the United States. If it is the measured differences in cognitive skills and not other economically relevant attributes of the families and economies that are important, the impact of skills can be derived from the different earnings of immigrants who received their schooling at home and in the United States.

The structure of the estimation is derived from a standard Mincer (1974) wage equation augmented by measured cognitive skills such as:

$$\ln y_{ic} = \alpha_0 + \alpha_1 S_{ic} + \alpha_2 PE_{ic} + \alpha_3 PE_{ic}^2 + \gamma_y H_{ic} + \upsilon_{ic} \tag{3}$$

where y is annual earnings for immigrant i from country c, S is years of school attainment, PE (=age-S-6) is potential experience, H is cognitive skills, and υ is a random error.[39]

We look at immigrants to the U.S. who were either educated entirely in their country of origin or entirely in the United States.[40] (This excludes any individuals partially educated in both the U.S. and their home countries in order to obtain a clear separation of treatment and control groups.) We assign the average cognitive-skill score of the home country ($\overline{T}_c$) for each immigrant and estimate the Mincer earnings Eq. (3) as:

$$\begin{aligned}\ln y_{ic} = {} & \alpha_0 + \alpha_1 S_{ic} + \alpha_2 PE_{ic} + \alpha_3 PE_{ic}^2 \\ & + [\alpha_4 ORIGIN_i + \delta \overline{T}_c + \delta_O (\overline{T}_c \times ORIGIN_i)] + \upsilon_{ic}\end{aligned} \tag{4}$$

where *ORIGIN* is an indicator that is one if immigrant i was educated entirely in schools in the country of origin and zero otherwise and the combined terms in brackets indicate the skills of individuals from country c. The parameter δ_O is the relevant contrast in skills between home-country schooling and U.S. schooling. We interpret δ_O as a difference-in-differences estimate of the effect of home-country test scores on earnings, where the first difference is between home-country educated immigrants (the "treatment group") and U.S.-educated immigrants (the "control group") from the same country, and the second difference is in the average cognitive-skill score of the home country.[41] The parameter δ captures the bias that would emerge in standard cross-sectional estimates from omitted variables like cultural traits that are correlated with home-country test scores in the same way for all immigrants from

Footnote 38 continued
considers a range of sensitivity analyses such as excluding Mexican immigrants and including only immigrants from English-speaking countries.

[39] Given that our growth specifications are most closely related to an endogenous-growth formulation, one cannot directly go from the estimated effects on individual productivity to the impact on growth rates. For that reason, in order to validate our micro estimates below, we concentrate on comparisons with other micro estimates of the impact of cognitive skills on productivity. While our analysis uses skill differences by country of origin to infer earnings differences among immigration in the U.S., Hendricks (2002) and Schoellman (2012) go the opposite way of using earnings differences of U.S. immigrations to infer cross-country differences in human capital.

[40] Immigrants are individuals born in a foreign country. The sample includes all individuals age 25 or older currently in the labor force with wage and salary earnings of at least $1,000 and not enrolled in school. Included immigrants had to have been born in a country with international test data (see Appendix B). The number of included countries is larger than in the previous growth regressions because of the lack of need to have internationally comparable GDP data for country of origin. Descriptive statistics are found in Appendix Table 13.

[41] Immigrants educated in their home country necessarily come to the U.S. at an older age than comparable immigrants educated in the U.S., suggesting that there might be differential selectivity and motivation of these two groups. But the key issue for identifying the impact of cognitive skills is that any selectivity in migration is the same across countries (which would then be captured by α_4), or at least is not correlated with differences in home-country cognitive skills.

the same country of origin (independent of where they were educated); in our more elaborate specifications with country-of-origin fixed effects, this parameter is not identified.

The first two columns of Table 5 report the estimates of the impact of cognitive skills from stratified samples for the two groups of immigrants. Test scores are normalized to mean zero and a standard deviation of one, so that the estimates indicate the proportionate increase in earnings from a one standard deviation increase in scores. Other things equal, there is essentially no relationship of U.S. earnings to scores of their country of origin, either quantitatively or statistically, for the 50,597 immigrants educated entirely in the U.S. On the other hand, one standard deviation greater performance in country-specific average test scores translates into a statistically significant earnings increase of approximately 16% for the 258,977 immigrants educated in their country of origin.

This estimate is surprisingly close to recent estimates for cognitive skills of U.S. workers, which indicate 10-15% returns to a standard deviation of test scores for young workers and 19% across the full age range of workers.[42] The closeness to the various estimates is surprising given that just average country scores as opposed to individual specific scores are used in the estimation here, although the averaging of scores does eliminate the measurement error found in individual test data.

Column 3 combines the samples and fully estimates Eq. (4). These estimates indicate a significant impact of test scores emanating from schooling in the immigrant's country of origin (δ_O). In contrast, the estimate of (home-country) test score for U.S.-educated immigrants (δ) is statistically insignificant, although the point estimate is noticeably greater than zero. Column 4 demonstrates that this latter effect comes entirely from the influence of immigrants from Mexico (who constitute 37% of all immigrants to the U.S.). The estimation for immigrants from Mexico is prone to classification error, because many Mexican families tend to move back and forth from Mexico—thus making assignment to U.S. or Mexican schooling prone to error.[43] Excluding Mexican immigrants, $\widehat{\delta_O}$ is highly significant with a point estimate of 0.13, while the coefficient for U.S.-educated immigrants falls to −0.026 and remains statistically insignificant.

The prior estimates indicate that the estimation strategy might be sensitive to variations in immigration patterns across the 64 sampled countries. For example, in addition to the complications for Mexican immigrants, the immigrants from other countries might vary by where they come in the ability distribution of the home country and the like. For this reason, the remaining columns of Table 5 contain country-of-origin fixed effects. Thus, immigrants educated entirely abroad in their home country are compared directly to immigrants *from the same country* educated entirely in the U.S. This eliminates any potential bias emanating from features specific to the country of origin, be they specific selectivity of the immigrant population or country-specific cultural traits. The only remaining assumption required for identification of our parameter of interest is that any potential difference between the early-

[42] Murnane et al. (2000) provide evidence from the High School and Beyond and the National Longitudinal Survey of the High School Class of 1972. Their estimates suggest some variation with males obtaining a 15% increase and females a 10% increase per standard deviation of test performance. Lazear (2003), relying on a somewhat younger sample from NELS88, provides a single estimate of 12%. These estimates are also very close to those in Mulligan (1999), who finds 11% for the normalized AFQT score in the NLSY data. Hanushek and Zhang (2009) estimate a return of 19% from the International Adult Literacy Survey, which samples workers aged 16–65.

[43] The assignment of individuals to U.S. schooling is based on census data indicating immigration before age 6. The assignment of individuals to schooling all in country of origin is based on age of immigration greater than years of schooling plus six. A person who moves back and forth during the schooling years could be erroneously classified as all U.S. or no U.S. schooling, even though they are really in the partial treatment category (which is excluded from the difference-in-differences estimation).

Table 5 Difference-in-differences estimates of returns to country-of-origin cognitive skills for U.S. immigrants

Sample:	U.S. educated[a]	Educated in country of origin[b]	All immigrants	W/o Mexico	All immigrants	W/o Mexico	Growth sample[c]	Only English speaking countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cognitive skills × educated in country of origin			0.0873 *(2.02)*	0.1324 *(3.31)*	0.1375 *(3.16)*	0.1398 *(4.13)*	0.1670 *(3.77)*	0.1616 *(3.57)*
Cognitive skills	0.0050 *(0.14)*	0.1582 *(2.37)*	0.0634 *(1.06)*	−0.0258 *(1.42)*	Not identified	Not identified	Not identified	Not identified
Educated in country of origin			−0.1385 *(3.95)*	−0.1011 *(3.03)*	−0.1298 *(2.98)*	−0.0626 *(2.07)*	−0.1309 *(2.58)*	−0.0206 *(0.83)*
Years of schooling	0.1155 *(14.08)*	0.0673 *(7.35)*	0.0700 *(7.43)*	0.0863 *(13.47)*	0.0579 *(4.14)*	0.0856 *(17.40)*	0.0553 *(4.06)*	0.0992 *(15.41)*

Table 5 continued

Sample:	U.S. educated[a] (1)	Educated in country of origin[b] (2)	All immigrants (3)	W/o Mexico (4)	All immigrants (5)	W/o Mexico (6)	Growth sample[c] (7)	Only English speaking countries (8)
Potential experience	0.0372	0.0235	0.0243	0.0215	0.0241	0.0227	0.0233	0.0205
	(19.71)	*(4.12)*	*(5.67)*	*(4.75)*	*(7.68)*	*(5.57)*	*(6.50)*	*(2.80)*
Potential experience squared	−0.00064	−0.00035	−0.00036	−0.0004	−0.00039	−0.0004	−0.0004	−0.0004
	(13.02)	*(4.79)*	*(6.10)*	*(4.90)*	*(7.40)*	*(5.87)*	*(6.46)*	*(3.25)*
Fixed effects for country of origin	no	no	no	no	yes	yes	yes	yes
Observations	50,597	258,977	309,574	187,506	309,574	187,506	273,213	72,091
No. of countries	64	64	64	63	64	63	47	12
R^2	0.157	0.170	0.180	0.132	0.196	0.150	0.202	0.156

Notes Dependent variable: log(annual earnings). Cognitive skills refer to average test score of country of origin (centered at zero). Sample: All immigrants identified by country of birth not in school whose age is greater than 25, who are employed, and who earned more than $1,000 in 1999. Immigrants who had obtained some but not all of their education in the U.S. were excluded from the sample. Immigrants from all countries of origin for which there are cognitive-skill scores, except for the following countries (areas) which could not be identified because of census restrictions on release of data for small cells: Swaziland, Slovenia, Macau-China, Luxembourg, Liechtenstein, Estonia, Botswana, Bahrain, Tunisia, and Iceland. Israel could not be identified separately from Palestine; both were assigned the Israeli score. Robust absolute values of t-statistics in parentheses with clustering by country of origin. Source: Authors' calculations from 2000 Census IPUMS data

[a] U.S. educated immigrants are identified as immigrating to the U.S. before the beginning year of schooling

[b] Immigrants educated in their country of origin are identified as immigrating to the U.S. after the final year of schooling

[c] The economic growth sample relies on the data for immigrants from the 50 countries in the basic growth regressions

immigrated U-S.-educated and the late-immigrated home-educated group of immigrants from each country (as captured by the *ORIGIN* indicator) does not vary across countries in a way associated with country-of-origin test scores.

Column 5 displays the primary estimation across all sampled countries with country-specific fixed effects. The estimated impact of cognitive skills is a 14 % increase in earnings from each standard deviation increase in origin-country test scores (when educated there). This estimate is highly significant. Further, the point estimate is virtually unchanged by excluding the Mexican immigrants (col. 6). The standard error is reduced by clearer assignment to treatment category (when Mexicans excluded), even though the sample is substantially reduced.

The final two columns investigate the sensitivity of these estimates to sample definition. First, our estimation of growth models used the 50 countries for which we could obtain the relevant economic data for GDP growth. Restricting this analysis to that smaller sample yields a slight increase in the magnitude of $\widehat{\delta_O}$ to 17 %, while it remains statistically significant. Second, because immigrants from non-English speaking countries may have lower earnings because of language difficulties, the final column shows estimates that come entirely from countries where English is the primary or official language.[44] Again, even for this sample of just 12 countries, variations in cognitive skills across countries have a strongly significant impact on earnings of 16 %.

The remaining rows of Table 5 provide estimates of the complete set of parameters. While there is some variance across samples in the estimate of the effect on earnings of being educated entirely in the country of origin, this appears to reduce average earnings by 6–13 % with the exception of English-speaking immigrants, who appear to suffer no significant average earnings loss compared to people educated entirely in the U.S. The estimated "Mincer" parameters (α_1, α_2, and α_3) appear within the range of typical estimates for the general population (see Heckman et al. 2008). Results remain qualitatively the same when indicators for decade of immigration and for gender are added to the model.[45] These last specifications address in part concerns that unmeasured differences between late immigrants (those receiving schooling in their home country) and early immigrants (those receiving schooling in the U.S.) might be driving the results.

These difference-in-differences estimates provide support for two conclusions about the potential causal impacts of cognitive skills. First, they contrast individuals receiving the treatment of home-country schooling to immigrants from the same country, all within the same labor market. Thus, they cannot be driven by differences in underlying economic institutions around the globe that are correlated with differences in cognitive skills. Second, they pinpoint the impact of schooling differences across countries, as distinct from family or cultural differences in attitudes, motivation, child rearing, and the like. In sum, the estimates, which are highly stable across different estimation samples, provide evidence that the economic impact is a causal one, and not purely associational.

It is very difficult, however, to compare magnitudes of coefficients from the immigrant earnings models to the growth regressions. These estimates are restricted to the private returns and leave out any of the externalities implicit in the estimated growth models. Thus, while

[44] Data on English language come from the CIA World Factbook. Countries were coded as English speaking if the CIA World Factbook listed English as an official language or as the most widely spoken language in the country. See https://www.cia.gov/library/publications/the-world-factbook/.

[45] When analyzed separately by gender, the results hold strongly for males whereas results for females—while pointing in the same direction—mostly do not reach statistical significance, as is common in labor-market analyses.

the estimated earnings impacts of cognitive skills are remarkably close to existing panel-data estimates, they do not translate easily into aggregate growth estimates.

The estimates do, however, provide direct support for the production view of schooling as contrasted with the signaling or screening view. Since the analysis of Spence (1973), questions have been raised about whether schools simply act as devices to select more able students as opposed to providing them with new knowledge and skills. Thus, an individual may get more schooling simply to signal high ability to the labor market. The difficulty has been that the labor-market implications for the returns to school attainment are the same in the alternative models—those with more schooling have more skills and thus receive higher wages. This fact has led to a variety of alternative ways to identify production versus signaling (cf. Weiss 1995; Riley 2001). One salient approach is reliance on what happens during schooling as opposed to the market returns to school attainment to identify the differences. The previous results provide just such evidence, because they show that the quality of different schools and the cognitive skills related to different schooling have direct payoffs within the same labor market. Therefore, these estimates yield even stronger evidence that policies to improve schools have social payoffs, as opposed to being limited to private payoffs as would be implied in the screening model.

7 Skill improvement and improved growth

The prior analyses have relied upon the average test score for each nation in order to characterize differences in skills of their labor forces. As noted, most of the variation in test scores occurs between countries, but the existence of some systematic change for countries suggests the possibility of an alternative difference-in-differences approach that uses the time-series evidence on performance within each country to identify the impact of skills on growth. Specifically, countries that improve the skills of their population—no matter how it is done—by the underlying model should see commensurate improvements in their rate of growth. This estimation removes any country-specific fixed effects affecting growth rates—such as basic economic institutions, cultural factors, political environment, and the like—and focuses on whether a country that alters the cognitive skills of its population is observed to receive an economic return.

While others have investigated turning points in growth, our focus is low-frequency changes such as those that might result from evolutionary schooling policies and that alter the path of economic growth.[46] Policies affecting the skill composition of the labor force necessarily unfold over lengthy periods and are not seen as sharp changes in outcomes.

To characterize the longitudinal patterns of test scores, we regress separate test scores by year, age group, and subject on a time variable (as well as age-group and subject indicators) and use the time coefficient as the measure of change in cognitive skills for each nation (see Appendix B for details). The amount of noise in each test observation, particularly with our common scaling, implies that such trends are also estimated with considerable noise. We therefore trust the rough cross-country pattern more than the specific point estimates of changes in each country. To put limits on the amount of noise affecting our analyses, we rely

[46] Relevant studies include Hausmann et al. (2005) who look at episodes of "growth accelerations"; Jones and Olken (2008) who consider patterns of 10-year periods of acceleration and collapse; and Barro and Ursúa (2008) who identify events of major declines in consumption that have potential implications for long-run growth. The identified periods are generally characterized by financial crisis, political instability, or war.

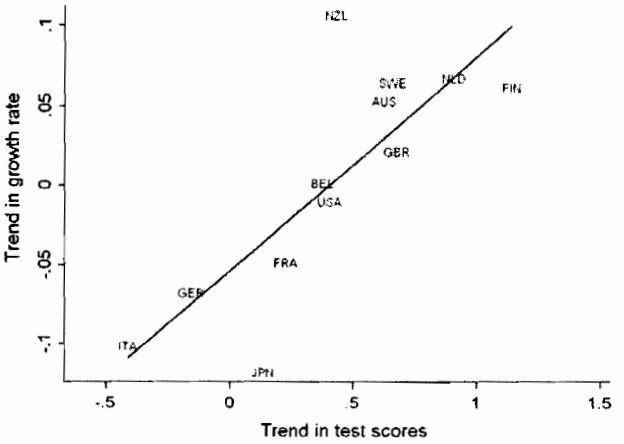


Fig. 2 Trends in Growth Rates vs. Trends in Test Scores. *Notes*: Scatter plot of trend in the growth rate of GDP per capita from 1975 to 2000 against trend in test scores, which is equivalent to the first column of Table 6. Three countries without test scores before 1972 in light gray; regression line refers to the remaining twelve countries. See Appendix B for details

on the sample of OECD countries that have test observations both before 1985 and up to 2003.[47]

As is evident from Fig. 2 (see also Appendix Fig. 5), substantial changes in test performance—both positive and negative—have occurred for OECD countries.[48] The rapid growth in performance of such countries as Canada, Finland, and the Netherlands contrasts sharply with the declining scores in Germany, Italy, and Norway. For our purposes, however, we are not interested in test scores for the school-aged population but instead in the skills of the relevant portions of the labor force. Thus, we need to assume that the currently observed trends in performance reflect long-run patterns of skill change and specifically those holding during the earlier time periods.

In a parallel manner, we estimate a time trend for annual growth rates in each country using the Penn World Tables data. The annual growth rate series for each country contains considerable noise, largely reflecting short-run business cycle phenomena or financial crises, and the trend estimation is designed to extract long-run changes in growth.[49]

The consistency of changes in test performance and changes in growth rates is evident in Fig. 2. When we split countries by being above or below the median change in growth rates and above or below the median change in cognitive skills, all countries fall into either the positive or negative quadrants on both measures. The largest outliers from the trend line are

[47] In fact, all countries except Canada, Korea, and Norway have test scores dating back at least to 1971.

[48] A comparison of the country rankings of projected skill levels for 1975 and 2000 yields a Spearman rank correlation is 0.78—again reinforcing the validity of country average scores for the main growth analysis.

[49] Descriptive statistics are found in Appendix Table 14. We also tried alternative measures of growth-rate changes, including the difference between the average growth rate in the first five years and in the last five years; trend growth using IMF data in national currencies; and IMF national currency data for the period 1975–2004. Using IMF national currency data is consistent with Nuxoll (1994) and Hanousek et al. (2008) who argue that using national accounts data is superior to relying on the price and exchange-rate adjustments in the basic Penn World Tables data when looking at growth rates. In our investigation of these options, the estimates of the impact of changes in test scores remain statistically significant and quantitatively very similar across alternatives and compared to the estimates reported in Table 6.

Table 6 Changes in cognitive skills and changes in growth paths

	(1)	(2)	(3)	(4)	(5)[a]	(6)[b]	(7)	(8)	(9)
Trend in cognitive skills	0.084 *(3.10)*	0.073 *(3.21)*	0.074 *(3.07)*	0.074 *(3.04)*	0.080 *(3.34)*	0.117 *(6.90)*			0.073 *(2.97)*
Dummy for cognitive-skill trend above median							0.117 *(5.98)*	0.103 *(4.87)*	
Average annual growth rate in GDP per capita 1975–2000		−0.030 *(2.73)*	−0.035 *(1.61)*	−0.028 *(1.69)*	−0.039 *(2.32)*	−0.085 *(5.26)*		−0.004 *(0.21)*	−0.031 *(2.55)*
Initial GDP per capita			−0.002 *(0.27)*					0.005 *(0.80)*	
Change in years of schooling 1975–2000				−0.004 *(0.21)*					
Trend in cognitive skills in 1999–2009									0.0004 *(0.03)*
No. of countries	15	15	15	15	15	12	15	15	15
R^2 (adj.)	0.380	0.586	0.551	0.550	0.582	0.891	0.713	0.735	0.548

Notes Dependent variable: trend in the growth rate of GDP per capita from 1975 to 2000. Regressions include a constant. Sample: OECD countries with test-score data both before 1985 and up to 2003. Test scores are average of math and science, primary through end of secondary school. Absolute *t*-statistics in parentheses

[a] WLS with inverse of standard error with which the trend in test scores was estimated as weights

[b] Excluding countries without test scores before 1972 (Canada, Korea, and Norway)

precisely the countries that have less historical test score data (Canada, Korea, and Norway) and that thus have poorer trend data.

We provide estimates of simple models of the change in growth rates over the 1975–2000 period in Table 6. By focusing on changes in test scores and in growth rates, these specifications are essentially equivalent to panel estimates with country fixed effects that eliminate time-invariant factors of cultural, institutions, or other potential influences. For the 15 OECD countries, 38% of the variance in growth-rate changes can be explained by test-score changes.[50] If we add measures for the average growth rate in each country and the initial GDP per capita (col. 2–3), the change in achievement scores remains statistically significant at near the same level as found in the simple regressions of column 1. The same is true when the change in quantitative educational attainment is added to the model (col. 4). Importantly, the change in educational attainment is orthogonal to the change in growth rates (either with controls for the test-score trend or without), reinforcing the introductory skepticism about the efficacy of past reliance on school attainment measures of human capital. Likewise, results are hardly affected if we weigh each observation by the inverse of the standard error with

[50] Results are fully consistent when adding six non-OECD countries with available data to the analysis (disregarding three non-OECD countries that prove extreme outliers on the estimated test-score trend at more than two and a half standard deviations above/below the OECD sample mean). While in the extended 21-country sample, OECD countries on average have lower growth trends, the effect of test-score trends on growth trends does not differ significantly between the OECD and non-OECD countries, and the main effect of the test-score trend is unaltered from the OECD-sample analysis (detailed results available from the authors upon request).

which the trend in test scores was estimated, in order to down weight those that are more noisily estimated (col. 5).

If, however, we restrict the analysis to those countries with test scores spanning a range of more than three decades, from at least 1971 to 2003, both the coefficient estimate and the explained variance grow in size (col. 6), as suggested in Fig. 2. In the sample without the three countries with limited time series information (Canada, Korea, and Norway), the test-score trend alone accounts for 64 % of the variation in growth trends. Alternative specifications look simply at whether the test-score trend is above or below the OECD median (col. 7–8). In all cases, the impact of changes in test scores on changes in growth rates remains very stable and is always statistically significant.

The underlying identifying assumption of these analyses is that the observed test-score trend captures a prior trend and is not affected by the partly overlapping growth trend. One way to test the validity of this identifying assumption is to use the most recently available test score data to estimate the trend in test scores subsequent to the period over which the growth trend is estimated. To do so, we estimate the trend in test scores from the 24 available test observations in 1999–2009 (details available on request). When entering the test-score trend in 1999–2009 to our regression, it is totally unrelated to the prior growth trend and does not affect the result on the prior test-score trend (col. 9). In fact, the test-score trend in 1999–2009 is not correlated with the prior long-term test-score trend (correlation coefficient −0.302, p-value 0.274) or with the growth trend in 1975–2000 (correlation coefficient −0.293, p-value 0.289).[51] This result corroborates the assumption that the identifying test-score variation is not itself caused by variation in growth. Furthermore, analyses presented in Appendix D (Table 16) use the most recent Penn World Table data (version 7.0) to show that results are strengthened when relating the initial test-score trend to the growth trend in 1985–2007 (rather than 1975–2000).

Still, this analysis requires backward extrapolation of the test-score data in order to capture changes for workers in the labor force. Thus, it cannot be considered definitive. We can, however, relate these estimates to the prior growth models. If we assumed that the observed trend in test scores had been going on since the oldest person in the current labor force went to school, an annual increase in test scores by 1 % of a standard deviation would translate into an annual increase in the growth rate by 0.07–0.12 percentage points. However, if we more realistically thought that any change in test scores began at the beginning of our observation period, then the impact of student improvements on the average labor force is much less, and the projected change in growth rates would be commensurately reduced. Back-of-the-envelope calculations suggest that in such a setting, the estimates based on the trend analysis in Table 6 are close to the steady-state estimates in Table 1.

In conclusion, the positive relationship between improving cognitive skills and improving growth rates provides another set of surprisingly consistent results based on a different approach to identifying the causal impact of cognitive skills—a focus on changes within each country that removes country-specific fixed effects. While it requires large extrapolations of changes to cover existing workers, the results are remarkably compatible with the underlying growth model—showing growth rates changing in a manner consistent with changes in cognitive skills.

[51] Results are qualitatively the same when using just the PISA tests in 2000–2009.

8 Rocket scientists or basic education for all?

While addressing the range of potential schooling policy options is clearly beyond the scope of this paper, our new data series allows us to extend the growth analysis to illuminate one important issue—whether to concentrate attention at the lowest or at the highest achievers. Some argue in favor of elitist school systems which focus on the top performers as potential future managers of the economy and drivers of innovation. Others favor more egalitarian school systems to ensure well-educated masses that will be capable of implementing established technologies. In other words, should education policy focus on forming a small group of "rocket scientists," or are approaches such as the Education for All initiative (UNESCO 2005) more promising in spurring growth?

To capture these differences in the distributional patterns of the test-score performance in different countries, we use the microdata from each of the international assessments to calculate measures of the share of students in each country who reach at least basic skills as well as those who reach superior performance levels (see Appendix B). We use performance of at least 400 test-score points on our transformed international scale—one standard deviation below the OECD mean—as our threshold of basic literacy and numeracy.[52] The international median of this share of students is 86 % in our sample, ranging from 18 % in Peru to 97 % in the Netherlands and Japan. As our threshold for superior performance, we take 600 points or one standard deviation above the OECD mean.[53] This level is reached by an international median of only 5 %, although it ranges from below 0.1 % in Colombia and Morocco to 18 % in Singapore and Korea and 22 % in Taiwan.[54] (As shown in Appendix Fig. 4, these differences represent more than simple mean displacement.)

As seen in the first three columns of Table 7, both measures of the test-score distribution are significantly related to economic growth, either when entered individually or jointly.[55] Both the basic-skill and the top-performing dimensions of educational performance appear separately important for growth. From the estimates in column 3, a ten percentage point increase in the share of students reaching basic literacy is associated with 0.3 percentage points higher annual growth, and a ten percentage point increase in the share of top-performing students is associated with 1.3 percentage points higher annual growth. However, it may be much more feasible to increase the basic-literacy share than to increase the top-performing share by the same amount, as suggested by the fact that the international standard deviations of

[52] The PISA 2003 science test uses the threshold of 400 points as the lowest bound for a basic level of science literacy (Organisation for Economic Co-operation and Development 2004, p. 292), and on the math test this corresponds to the middle of the level 1 range (358 to 420 test-score points), which denotes that 15-year-old students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. For example, given the exchange rate between dollars and euros, a student at level 1 can calculate the euro-equivalent of a given number of dollars.

[53] A score of 600 points is near the threshold of the level 5 range of performance on the PISA 2003 math test, which denotes that 15-year-old students can develop and work with models for complex situations, identifying constraints and specifying assumptions; they can reflect on their answers and can formulate and communicate their interpretations and reasoning.

[54] The distributions depicted in Fig. 4 reveal that such distributional measures capture much more of the overall distribution than a simple measure such as the standard deviation in national test scores. The standard deviation in test scores does not enter our basic model significantly (see Castelló and Doménech 2002 for related analyses using measures of educational inequality based on years of schooling).

[55] In the joint model, the two measures are separately significant even though they are highly correlated across countries with a simple correlation of 0.73. The mean test score used in previous models is more highly correlated with the basic literacy share ($r = 0.96$) than with the top-performing share ($r = 0.85$). If the mean test score is added to column 3, the basic-literacy share becomes insignificant, but in a specification with just the mean, mean and top-performing shares both remain significant.

Table 7 Rocket scientists or basic education for all?

	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[b]	(7)[b]
Share of students reaching basic literacy	4.717		2.732	1.002	3.460	5.150	5.869
	(6.64)		*(3.61)*	*(1.33)*	*(3.81)*	*(2.87)*	*(3.33)*
Share of top-performing students		19.347	12.880	11.735	8.460	4.226	−1.530
		(2.653)	*(4.35)*	*(4.18)*	*(2.37)*	*(0.65)*	*(0.22)*
Share of students reaching basic literacy × initial GDP per capita					0.376		
					(1.25)		
Share of top-performing students × Initial GDP per capita					−2.148		−1.649
					(2.11)		*(2.07)*
Share of students reaching basic literacy × Share of top-performing students						42.357	53.538
						(1.48)	*(1.91)*
No. of countries	50	50	50	45	50	50	50
R^2 (adj.)	0.610	0.646	0.719	0.823	0.734	0.727	0.746

Notes Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Control variables: GDP per capita 1960, years of schooling 1960, and a constant. Shares are based on average test scores in math and science, primary through end of secondary school, all years. Absolute t-statistics in parentheses

[a] Specification includes additional controls for openness, property rights, fertility, and tropical location

[b] All interacted variables are centered on zero

these two shares are 0.215 and 0.054, respectively. Thus, increasing each share by roughly half a standard deviation (10 percentage points basic-literacy share and 2.5 percentage points top-performing share) yields a similar growth effect of roughly 0.3 percentage points.

The impact of having more top performers is only slightly reduced by introducing the measures of economic institutions, fertility, and tropical geography (col. 4). On the other hand, the separate influence of basic literacy levels falls quantitatively and becomes statistically insignificant in the expanded model (for the 45 countries with complete data), in line with an interpretation where part of the effect of basic literacy comes through improved institutions (Glaeser et al. 2004).

The effect of the basic-literacy share does not vary significantly with the initial level of development, but the effect of the top-performing share is significantly larger in countries that have more scope to catch up to the initially most productive countries (col. 5). These results appear consistent with a mixture of the basic models of human capital and growth mentioned earlier. The accumulation of skills as a standard production factor, emphasized by augmented neoclassical growth models (e.g., Mankiw et al. 1992), is probably best captured by the basic-literacy term, which has positive effects that are similar in size across all countries. But, the larger growth effect of high-level skills in countries farther from the technological frontier is most consistent with technological diffusion models (e.g., Nelson and Phelps 1966). From this perspective, countries need high-skilled human capital for an imitation strategy, and the process of economic convergence is accelerated in countries with larger shares of high-performing students.[56] Obvious cases are East Asian countries such as Taiwan, Singapore, and Korea that all have particularly large shares of high-performers, started from relatively low levels, and have shown outstanding growth performances, but the results of column 5 are nonetheless robust to the inclusion of an East Asian dummy, or a full set of regional dummies.

[56] For an alternative model of imitation and innovation that emphasizes the innovation margin, see Vandenbussche et al. (2006) and Aghion et al. (2009). These studies, however, focus just on developed countries and miss the role of rocket scientists in the transmission of technologies to developing countries. Moreover, Hanushek and Woessmann (2011b) show that differences in basic skills are more important than differences in advanced skills in explaining growth differences among just the OECD countries.

A particularly informative extension considers the interaction of the top-performing and basic-literacy shares (col. 6 and 7). This complementarity between basic skills and top-level skills suggests that in order to be able to implement the imitation and innovation strategies developed by scientists, countries need a workforce with at least basic skills.[57]

Many countries have focused on either basic skills or engineers and scientists. In terms of growth, our estimates suggest that developing basic skills and highly talented people reinforce each other. Moreover, achieving basic literacy for all may well be a precondition for identifying those who can reach "rocket scientist" status. In other words, tournaments among a large pool of students with basic skills may be an efficient way to obtain a large share of high-performers.

9 Conclusions

A myriad of empirical estimates of cross-country growth models exist. The general criticism of these is that they provide little confidence that the models satisfactorily identify the causal impact of their included determinants of growth. And, a related criticism is that they then cannot provide any real policy guidance.

We have focused on the role of educational achievement, or cognitive skills, in determining economic growth and have taken the quest for policy guidance seriously. We have investigated a set of models that approach identification from different vantage points. Individually, as discussed, these approaches do require some strong maintained hypotheses, but importantly each of the approaches is subject to different questions and would fail for very different reasons. While there remain other threats to identification that cannot be ruled out in our samples, the alternative analytical perspectives narrow the range of possible opposing explanations for the stylized facts based on reverse causation, economic and social institutions, and cultural influences.

The clearest result here is the consistency of the alternative estimates of the cognitive skills-growth relationship—both in terms of quantitative impacts and statistical significance. The remarkable stability of the models in the face of alternative specifications, varying samples, and alternative measures of cognitive skills implies a robustness uncommon to most cross-country growth modeling. In terms of previous questions about the fragility of any estimates of human capital and growth, these estimates underscore a simple finding that prior results suffered from critical measurement issues.

The stylized fact of a very strong relationship between cognitive skills and growth does not address all concerns. For policy advice, it is important to know whether the estimated relationship is causal or a mere association reflecting omitted variables, poor achievement measurement, or restricted models of growth. With the limited sample of nations—each with alternative political, cultural, and economic institutions—it is clearly very difficult to rule out all hypotheses about other possible reasons for the association between cognitive skills and growth. Our approach involves adapting common microeconometric approaches to the very different circumstances of long-term economic growth. We estimate instrumental-variable models using institutional characteristics of each country's school system to rule out that the skill-growth connection just reflects cultural differences across countries. Difference-in-differences approaches for immigrants from different countries on the U.S. labor market to rule out the possibility that test scores simply reflect cultural factors or economic institutions of

[57] The issue of skill complementarity in production has been addressed in explaining the pattern of earnings inequality. The U.S. analysis of Autor et al. (2006, 2008) suggests that high-skilled workers and low-skilled workers are complements, a result that helps explain income variations across the educational spectrum.

the home countries. And longitudinal models to eliminate the level effects which may be interrelated with country-specific institutions and cultures. Conclusions about the impact of cognitive skills are remarkably consistent across the very different estimation methods.

Our analysis is reinforced by Ciccone and Papaioannou (2009) who find that countries with a more skilled labor force (using the Hanushek and Kimko test measures) experienced faster growth in skill-intensive industries during the 1980s and 1990s. This evidence, which is derived from within-country analysis of development outcomes, strengthens our interpretation that more skilled people contribute to a more rapid adoption of new technologies and production processes—a central element of both endogenous growth models that stress innovation and ideas (Romer 1990) and of models of technological diffusion and growth (Nelson and Phelps 1966). By using country and industry fixed effects, it also excludes a variety of concerns about endogeneity due to variations in country institutions and cultures which tend to affect sectors uniformly.

The simple conclusion from the combined evidence is that differences in cognitive skills lead to economically significant differences in economic growth. Moreover, since the tests concentrate on the impact of schools, the evidence suggests that school policy can, if effective in raising cognitive skills, be an important force in economic development. Thus, it would be inappropriate to interpret the test differences as a simple reflection of ability or family differences, factors that might be very impervious to policy changes.

Finally, what are the economic rewards that are suggested by the models for improving skills? Almost all of the alternative specifications and modeling approaches suggest that one standard deviation higher cognitive skills of a country's workforce is associated with approximately two percentage points higher annual growth in per capita GDP. This magnitude is clearly substantial, particularly when compared to regional growth rates that average between 1.4 and 4.5 % over the 1960-2000 period (Appendix Table 8). On the other hand, it is implausible to expect a country to improve by one standard deviation—bringing, say, Mexico up to the OECD average—over any reasonable time horizon. It is plausible to think of getting schooling improvements that would lift a country's average by 1/4 standard deviation (25 points on a PISA scale). This kind of improvement has, for example, been observed by Poland during the past decade and by Finland over the past two to three decades. In an endogenous growth world, where long-run growth rates can change permanently with improved skills and human capital, the estimates would imply a boast in growth rates of 1/2 %. Such a change would lead to enormous changes in the economic outcomes for a country. Hanushek and Woessmann (2011b) simulate the impact of such a change on GDP and find that the present value of improvements over the next 80 years would be almost three times the value of current GDP for the country.[58] Put in terms of the present value of future GDP, this amounts to a 6.2 % increase in GDP over the 80-year period.

An alternative perspective is that of the neoclassical model where there would be a period of higher growth because of the improved skills but the economy would then return to the prior steady-state rate of growth at a higher level of income. The prior growth model estimates provide a means of estimating the impacts of this.[59] In the same scenario as above, the present value would be slightly over two times current GDP or 4.3 % of the present value of GDP over the 80 year period. The relative closeness of the endogenous and neoclassical estimates reflects the fact that convergence is, similar to that in Mankiw et al. (1992), relatively slow

[58] These simulations assume that the past growth patterns hold into the future; that base growth in per capita GDP is 1.5 %; that the reform of schools takes 20 years to complete; that the impact depends on the average quality of the labor force; and that future values are discounted at 3 %.

[59] For these simulations we rely upon the estimates in Table 1, column 9, that have the log of initial GDP per capita as an explanatory variable.

so that it takes a substantial time before the economy reverts to the old steady-state path. These simulations also underscore that the estimated impacts are still large even if part of our estimates were produced by omitted variables that are correlated with cognitive skills. If, for example, the true causal impact of skill differences were only half as large as suggested by our estimates, we would still be left with extraordinarily large policy opportunities.

By itself, finding a potential role for schools does not point to any clear policies. Indeed, that discussion would enter into a variety of controversial areas and would lead us far afield. Nonetheless, our aggregate data provide direct evidence that both providing broad basic education—education for all—*and* pushing significant numbers to very high achievement levels have economic payoffs.

Appendix A: Regional data

See Table 8.

Appendix B: Measures of cognitive skills

A key element of our work is developing a measure that can equate knowledge of people across countries. In many ways this is an extension of notions of human capital that have been developed over the past half century. But it is a specific refinement that, while important in a variety of applications within nations, becomes a necessity when comparing different countries. Within a country, human capital is often proxied by quantity of schooling. This is partly necessitated by commonly available data but partly justified on the idea that differences in knowledge between levels of schooling are greater than those within levels of schooling.

Until recent publicity, most people were unaware of the international student testing that could provide direct comparisons of student knowledge across countries. In fact, international assessments of student achievement, aimed largely at math and science, were begun over four decades ago. Although national participation has been voluntary, recent expansions to all OECD countries and more have led to increasingly valid and reliable indicators of cognitive skills.

Internationally comparable student achievement tests have been conducted since the First International Mathematics Study (FIMS), which tested in 1964. The latest international studies used in our analyses are the 2003 cycles of the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). From FIMS to the latest TIMSS and PISA, a total of 12 international student achievement tests (ISATs) were conducted.[60] Although varying across the individual assessments, testing covers math, science, and reading for three age/grade groups: primary education (age 9/10), lower secondary education (age 13 to 15), and the final year of secondary education (generally grade 12 or 13).

Given this 3×3 grade-by-subject matrix, Table 9 summarizes the specific ISATs that have been conducted in three periods of time: late 1960s/early 1970s (1964-72), 1980s (1982-91), and late 1990s/early 2000s (1995-2003). Several features of the emerging pattern are worth

[60] In this study, we do not include the two tests conducted by the International Assessment of Educational Progress (IAEP) in 1988 and 1991, because they used the U.S. NAEP test as their testing instrument, which is geared to the U.S. curriculum and may thus introduce bias to the international testing. By contrast, the tests included here are not associated with the curriculum in any particular country, but have been devised in an international cooperative process between all participating countries.

Table 8 Income, education, and growth across world regions

Region[a]	No.countries[b]	GDP per capita 1960 (US$)	Growth of GDP per capita 1960–2000 (%)	GDP per capita 2000 (US$)	Years of schooling 1960	Test score	All Penn World Tables Countries	
							No. countries[c]	GDP per capita 1960 (US$)
Asia	11	1,891	4.5	13,571	4.0	479.8	15	1,642
Sub-Saharan Africa	3	2,304	1.4	3,792	3.3	360.0	40	1,482
Middle East and North Africa	8	2,599	2.7	8,415	2.7	412.4	10	2,487
Southern Europe	5	4,030	3.4	14,943	5.6	466.4	5	4,030
Latin America	7	4,152	1.8	8,063	4.7	388.3	24	3,276
Central Europe	7	8,859	2.6	24,163	8.3	505.3	7	8,859
Northern Europe	5	8,962	2.6	25,185	8.0	497.3	5	8,962
Commonwealth OECD	4	11,251	2.1	26,147	9.5	500.3	4	11,251
Note: Asia w/oJapan	*10*	*1,614*	*4.5*	*12,460*	*3.5*	*474.7*	*14*	*1,427*

[a] The country observations contained in the eight regions are: Asia (11): China, Hong Kong, India, Indonesia, Japan, Rep. of Korea, Malaysia, Philippines, Singapore, Taiwan, Thailand; Sub-Saharan Africa (3): Ghana, South Africa, Zimbabwe; Middle East and North Africa (8): Cyprus, Egypt, Iran, Israel, Jordan, Morocco, Tunisia, Turkey; Southern Europe (5): Greece, Italy, Portugal, Romania, Spain; Latin America (7): Argentina, Brazil, Chile, Colombia, Mexico, Peru, Uruguay; Central Europe (7): Austria, Belgium, France, Ireland, Netherlands, Switzerland, United Kingdom; Northern Europe (5): Denmark, Finland, Iceland, Norway, Sweden; Commonwealth OECD members (4): Australia, Canada, New Zealand, USA.

[b] Sample of all countries by region with internationally comparable data on GDP that ever participated in an international student achievement test; see Appendix B for details.

[c] Sample of all countries in Penn World Tables with data on GDP per capita in 1960 by region.

Sources GDP: own calculations based on Penn World Tables (Heston et al. (2002)); years of schooling: own calculations based on Cohen and Soto (2007); test score: own calculations based on international student achievement tests; see Appendix B for details.

Table 9 International tests by period, subject, and age group

	Math	Science	Reading
1964–72			
Primary		FISS	
Lower secondary	FIMS	FISS	FIRS
Final secondary	FIMS	FISS	
1982–91			
Primary		SISS	SIRS
Lower secondary	SIMS	SISS	SIRS
Final secondary	SIMS	SISS	
1995–2003			
Primary	TIMSS	TIMSS	PIRLS
	TIMSS 2003	TIMSS 2003	
Lower secondary	TIMSS	TIMSS	
	TIMSS-Repeat	TIMSS-Repeat	
	PISA 2000/02	PISA 2000/02	
	TIMSS 2003	TIMSS 2003	PISA 2000/02
	PISA 2003	PISA 2003	PISA 2003
Final secondary	TIMSS	TIMSS	

Notes For abbreviations and details, see Table 10

noting. First, math and science have been tested at all three grade levels, while reading has never been tested in the final grade of secondary school. Second, all subjects are available in all periods, although coverage is more extensive in math and science than in reading. Third, in each period, the lower secondary level has been tested in all three subjects; thus, there is no primary or final-secondary study that would add a subject not already tested at lower secondary in the period. Fourth, each cell available in 1964-91 has at least one counterpart in 1995-2003.

Table 10 provides additional detail on each ISAT. A total of 77 countries have participated in at least one of the ISATs in math or science, but several of the countries have participated at only one or a few points in time. Even within the same assessment, countries do not always participate at all grade levels. The largest number of countries tends to have participated at the lower secondary level.

To obtain a common measure of cognitive skills, we want to draw on as much internationally comparable information as possible. This raises the issue whether the different ISATs with their different participating countries, student samples, and perspectives on what should be tested (see Neidorf et al. 2006) are measuring a common dimension of cognitive skills. For example, the TIMSS tests are related to elements of the school curricula common to participating countries, while the PISA tests are designed to be applied assessments of real-world problems, irrespective of specific curricula. However, the fact is that the TIMSS tests with their curricular focus and the PISA tests with their applied focus are highly correlated at the country level. For example, the correlation between the TIMSS 2003 tests of 8th graders and the PISA 2003 tests of 15-year-olds across the 19 countries participating in both is as high as 0.87 in math and 0.97 in science. It is also 0.86 in both math and science across the 21 countries participating both in the TIMSS 1999 tests and the PISA

Table 10 The international student achievement tests

	Abbreviation	Study	Year	Subject	Age[a,b]	Countries[c]	Organization[d]	Scale[e]
1	FIMS	First International Mathematics Study	1964	Math	13,FS	11	IEA	PC
2	FISS	First International Science Study	1970–71	Science	10,14,FS	14,16,16	IEA	PC
3	FIRS	First International Reading Study	1970–72	Reading	13	12	IEA	PC
4	SIMS	Second International Mathematics Study	1980–82	Math	13,FS	17,12	IEA	PC
5	SISS	Second International Science Study	1983–84	Science	10,13,FS	15,17,13	IEA	PC
6	SIRS	Second International Reading Study	1990–91	Reading	9,13	26,30	IEA	IRT
7	TIMSS	Third International Mathematics and Science Study	1994–95	Math/Science	9(3+4), 13(7+8),FS	25,39,21	IEA	IRT
8	TIMSS-Repeat	TIMSS-Repeat	1999	Math/Science	13(8)	38	IEA	IRT
9	PISA 2000/02	Programme for International Student Assessment	2000+02	Reading/ Math/Science	15	31+10	OECD	IRT
10	PIRLS	Progress in International Reading Literacy Study	2001	Reading	9(4)	34	IEA	IRT
11	TIMSS 2003	Trends in International Mathematics and Science Study	2003	Math/Science	9(4),13(8)	24,45	IEA	IRT
12	PISA 2003	Programme for International Student Assessment	2003	Reading/ Math/Science	15	40	OECD	IRT

[a] Grade in parentheses where grade level was target population.
[b] FS = final year of secondary education (differs across countries).
[c] Number of participating countries that yielded internationally comparable performance data.
[d] Conducting organization: International Association for the Evaluation of Educational Achievement (IEA); Organisation for Economic Co-operation and Development (OECD).
[e] Test scale: percent-correct formal (PC); item-response-theory proficiency scale (IRT).

2000/02 tests. Thus, ISATs with very different foci and perspectives tend, nonetheless, to be highly related, lending support to our approach of aggregating different ISATs for each country.

The general idea behind our approach to aggregation is that of empirical calibration. We rely upon information about the overall distribution of scores on each ISAT to compare national responses. This contrasts with the psychometric approach to scaling that calibrates tests through common elements on each test. In reality, the international testing situations are separate events with no general attempt to provide common scaling across tests and across the full time period.

The fact that the scales of their test-score results are not directly equated across tests is a major drawback in comparative uses of the various ISATs. They do not use the same test questions; nor do they even use the same technique and scale of mapping answers into test scores.[61] The early tests mainly used aggregate scores in "percent correct" format, but with questions of varying difficulty in the different tests, these scores will not be comparable across tests. The later tests use a more sophisticated scale, constructed using Item Response Theory (IRT). Among other things, IRT weights different questions by their revealed difficulty and then maps answers onto a pre-set scale set to yield a given international mean and standard deviation among the participating countries. However, the questions on which the mapping is based are not the same in the different tests. Even more, the set of participating countries varies considerably across tests, making the separately developed scales incomparable across ISATs.

Therefore, to compare performance on the ISATs across tests and thus over time, we have to project the performance of different countries on different tests onto a common metric. For that, we have to develop a common metric both for the *level* and for the *variation* of test performance.

Comparable level. To make the level of ISATs comparable, we need information on test performance that is comparable over time. Such information is available in the United States in the form of the National Assessment of Educational Progress (NAEP), which tests the math, science, and reading performance of nationally representative samples of 9-, 13-, and 17-year-old U.S. students in an intertemporally comparable way since 1969. This is the only available information on educational performance that is consistently available for comparisons over time. The United States is also the only country that participated in every ISAT. Given the time-series evidence on the performance of U.S. students, we can thus scale the level of each ISAT relative to the known intertemporally comparable test performance of the United States. Figure 3 shows the available NAEP results in the three subjects and age groups.[62] Despite some notable changes, the performance of U.S. students has been relatively flat over the period 1969-1999.

We start by calculating the U.S. performance difference between 1999 and any earlier point in time and express it in standard deviations (s.d.) of the international PISA 2000 study:

$$U_{a,s,t}^{US} = \left(NAEP_{a,s,t}^{US} - NAEP_{a,s,1999}^{US}\right) \frac{SD_{s}^{US,PISA}}{SD_{a,s}^{US,NAEP}} \tag{A1}$$

[61] Recent testing in both TIMSS and PISA has involved overlapping test items that permit test calibration, but these do not provide any benchmarks across the two testing regimes or links with earlier testing.

[62] Note that changes in NAEP testing make it difficult to use this methodology for the more recent PISA and TIMSS assessments. For example, the science tests were revised in 2009, and the new scale that was employed makes the data incomparable to prior years. On the other hand, recent PISA and TIMSS assessments have been designed to provide comparability over time of the subject surveys.

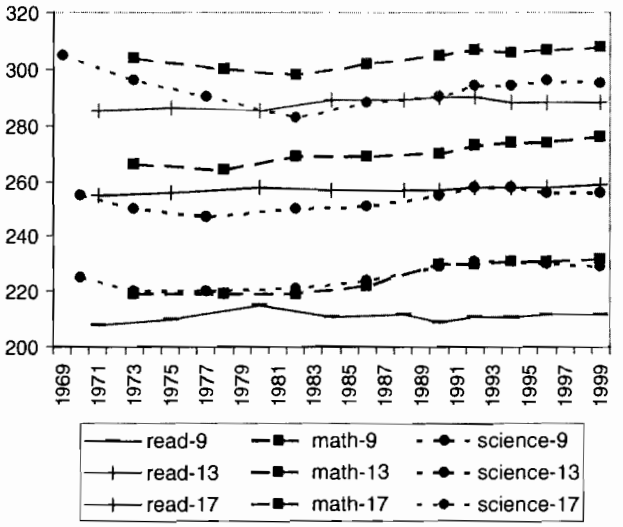


Fig. 3 Student Achievement in the United States over Time: The National Assessment of Educational Progress (NAEP) Source: U.S. Department of Education, Institute of Education Sciences (2008)

where U is the standardized performance difference of U.S. students at age a in subject s at time t relative to 1999, $NAEP$ is the age-, subject-, and time-specific NAEP test score, $SD^{US,PISA}$ is the subject-specific s.d. of U.S. students on the PISA test, and $SD^{US,NAEP}$ is the age- and subject-specific s.d. of the U.S. NAEP test.[63] NAEP scores are available at 2–4 year intervals over the period; values for non-NAEP years are obtained by linear interpolation between available years.

This alone does not yet yield a common scale for all the countries on the different tests. While we know for each participating country whether it performed above or below the respective U.S. performance on each specific test, we need to make the international variation in test scores comparable across the different ISATs to determine "how much" above or below.

Comparable variation. Developing a common metric for the variation of test scores in the different ISATs is harder to achieve than for the level. There is no explicit external information available on trends in the cross-country performance variation, and the diversity of original tests and participating countries precludes a direct comparison across tests. One way to achieve comparability, though, would be to have a group of countries across which it is reasonable to assume relative constancy in the size of the cross-country variation in test scores and whose members participated in sufficient number in the different tests. This group could only include relatively stable countries with relatively stable education systems over time, which should not have experienced major changes in overall enrollment across the ISATs.

Thus, we suggest two criteria for a group of countries to serve as a standardization benchmark for performance variation over time. First, the countries have been member states of the relatively homogenous and economically advanced group of OECD countries in the whole

[63] The s.d. of the NAEP tests in reading for 1984-1996 and in math and science in 1977/78-1996 are reported in Department of Education, Institute of Education Sciences (2008). Since no s.d. information is available for the earlier and the 1999 NAEP tests, and since the available s.d. are relatively stable over time, we take a simple mean of the available s.d. in each subject at each age level over time. PISA tested only 15-year-olds, but has the same three subjects as the NAEP test.

period of ISAT observations, that is, since 1964. Second, the countries should have had a substantial enrollment in secondary education already in 1964. Given data constraints, we implement this by dropping all countries where more than half of the 2001 population aged 45-54 (the cohort roughly in secondary school in the first ISAT) did not attain upper secondary education (OECD 2003a). There are 13 countries that meet both of these measures of stability, which we term the "OECD Standardization Group" (OSG) of countries.[64]

Under the assumption that the cross-country variation among the OSG countries did not vary substantially since 1964, we can use the OSG countries to develop a comparable scale for the variation on the different ISATs. We do so by projecting the s.d. among those of the OSG countries that participated in any particular ISAT from the subject-specific PISA test onto the particular ISAT. That is, we transform the original test score O of country i (specific for each age a and subject s) at time t into a transformed test score X according to:

$$X^i_{a,s,t} = \left(O^i_{a,s,t} - \overline{O^{OSG}_{a,s,t}}\right) \frac{SD^{OSG}_{s,PISA}}{SD^{OSG}_{a,s,t}} \tag{A2}$$

The test score X has the following distributional characteristics for each ISAT. First, it has a mean of zero among the OSG (attained by subtracting the OSG mean $\overline{O^{OSG}}$ from each country's original test score). Second, it has a between-country s.d. among the OSG that is the same as the s.d. of the very same countries on the PISA test in the specific subject (attained by dividing through the s.d. among the OSG countries in the specific test and multiplying by the s.d. of these same countries in the relevant PISA test). In effect, this rescaled test score now has a metric whose variation is comparable across tests.

Performance on a common metric. Finally, we use the time-series evidence on educational performance in the U.S. derived above to put a common level to the intertemporally comparable metric for the different ISATs. This is achieved in the standardized test score I:

$$I^i_{a,s,t} = X^i_{a,s,t} - X^{US}_{a,s,t} + O^{US}_{s,PISA} + U^{US}_{a,s,t} \tag{A3}$$

which adjusts the variation-adjusted test score X so that the U.S. performance level on each test equals the U.S. performance on the PISA test in the specific subject plus the age- and subject-specific adjustment factor U based on NAEP as derived in Eq. (A1) above.

Equation (A3) yields measures of the performance of the participating countries in each ISAT on a common scale that is comparable across ISATs. In effect, the internationally and intertemporally standardized test score I projects the PISA scale onto all other tests.

While we are reasonably confident about the comparisons of the standardized scores within the OECD countries which are fully tested in recent years, we are less certain about countries that are far from the measured OECD performance. In particular, countries far off the scale of the original test scores—e.g., two s.d. below the mean—may not be well represented because the tests may be too hard and thus not very informative for them. Our linear transformations are susceptible to considerable noise for these countries.

Our main measure of cognitive skills is a simple average of all standardized math and science test scores of the ISATs in which a country participated. Table 11 reports the basic

[64] The OSG countries are: Austria, Belgium, Canada, Denmark, France, Germany, Iceland, Japan, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The Netherlands also meets both criteria, but does not have internationally comparable PISA 2000 data which we require for our standardization.

Table 11 International data on cognitive skills

Code	Country	Growth sample[a]	Cognitive[b]	Lowsec[c]	Basic[d]	Top[e]
ALB	Albania	0	3.785	3.785	0.424	0.013
ARG	Argentina	1	3.920	3.920	0.492	0.027
ARM	Armenia	0	4.429	4.490	0.745	0.008
AUS	Australia	1	5.094	5.138	0.938	0.112
AUT	Austria	1	5.089	5.090	0.931	0.097
BEL	Belgium	1	5.041	5.072	0.931	0.094
BGR	Bulgaria	0	4.789	4.789	0.765	0.083
BHR	Bahrain	0	4.114	4.114	0.608	0.003
BRA	Brazil	1	3.638	3.638	0.338	0.011
BWA	Botswana	0	3.575	3.575	0.374	0.000
CAN	Canada	1	5.038	5.125	0.948	0.083
CHE	Switzerland	1	5.142	5.102	0.919	0.134
CHL	Chile	1	4.049	3.945	0.625	0.013
CHN	China	1	4.939	4.939	0.935	0.083
COL	Colombia	1	4.152	4.152	0.644	0.000
CYP	Cyprus	1	4.542	4.413	0.825	0.011
CZE	Czech Republic	0	5.108	5.177	0.931	0.122
DNK	Denmark	1	4.962	4.869	0.888	0.088
EGY	Egypt	1	4.030	4.030	0.577	0.010
ESP	Spain	1	4.829	4.829	0.859	0.079
EST	Estonia	0	5.192	5.192	0.973	0.095
FIN	Finland	1	5.126	5.173	0.958	0.124
FRA	France	1	5.040	4.972	0.926	0.085
GBR	United Kingdom	1	4.950	4.995	0.929	0.088
GER	Germany	0	4.956	4.959	0.906	0.105
GHA	Ghana	1	3.603	3.252	0.403	0.010
GRC	Greece	1	4.608	4.618	0.798	0.042
HKG	Hong Kong	1	5.195	5.265	0.944	0.123
HUN	Hungary	0	5.045	5.134	0.941	0.103
IDN	Indonesia	1	3.880	3.880	0.467	0.008
IND	India	1	4.281	4.165	0.922	0.013
IRL	Ireland	1	4.995	5.040	0.914	0.094
IRN	Iran	1	4.219	4.262	0.727	0.006
ISL	Iceland	1	4.936	4.945	0.908	0.074
ISR	Israel	1	4.686	4.660	0.826	0.053
ITA	Italy	1	4.758	4.693	0.875	0.054
JOR	Jordan	1	4.264	4.264	0.662	0.044
JPN	Japan	1	5.310	5.398	0.967	0.168
KOR	Korea, Republic of	1	5.338	5.401	0.962	0.178
KWT	Kuwait	0	4.046	4.223	0.575	0.000
LBN	Lebanon	0	3.950	3.950	0.595	0.002

Table 11 continued

Code	Country	Growth sample[a]	Cognitive[b]	Lowsec[c]	basic[d]	Top[e]
LIE	Liechtenstein	0	5.128	5.128	0.860	0.198
LTU	Lithuania	0	4.779	4.694	0.891	0.030
LUX	Luxembourg	0	4.641	4.641	0.776	0.067
LVA	Latvia	0	4.803	4.779	0.869	0.050
MAC	Macao-China	0	5.260	5.260	0.919	0.204
MAR	Morocco	1	3.327	3.243	0.344	0.001
MDA	Moldova	0	4.530	4.419	0.787	0.029
MEX	Mexico	1	3.998	3.998	0.489	0.009
MKD	Macedonia	0	4.151	4.151	0.609	0.028
MYS	Malaysia	1	4.838	4.838	0.864	0.065
NGA	Nigeria	0	4.154	4.163	0.671	0.001
NLD	Netherlands	1	5.115	5.149	0.965	0.092
NOR	Norway	1	4.830	4.855	0.894	0.056
NZL	New Zealand	1	4.978	5.009	0.910	0.106
PER	Peru	1	3.125	3.125	0.182	0.002
PHL	Philippines	1	3.647	3.502	0.485	0.006
POL	Poland	0	4.846	4.861	0.838	0.099
PRT	Portugal	1	4.564	4.592	0.803	0.032
PSE	Palestine	0	4.062	4.062	0.571	0.008
ROM	Romania	1	4.562	4.562	0.780	0.046
RUS	Russian Federation	0	4.922	4.906	0.884	0.081
SAU	Saudi Arabia	0	3.663	3.663	0.331	0.000
SGP	Singapore	1	5.330	5.512	0.945	0.177
SRB	Serbia	0	4.447	4.447	0.718	0.024
SVK	Slovak Rep.	0	5.052	5.052	0.906	0.112
SVN	Slovenia	0	4.993	5.076	0.939	0.061
SWE	Sweden	1	5.013	4.948	0.939	0.088
SWZ	Swaziland	0	4.398	4.398	0.801	0.004
THA	Thailand	1	4.565	4.556	0.851	0.019
TUN	Tunisia	1	3.795	3.889	0.458	0.003
TUR	Turkey	1	4.128	4.128	0.582	0.039
TWN	Taiwan (Chinese Taipei)	1	5.452	5.599	0.958	0.219
URY	Uruguay	1	4.300	4.300	0.615	0.049
USA	United States	1	4.903	4.911	0.918	0.073
ZAF	South Africa	1	3.089	2.683	0.353	0.005
ZWE	Zimbabwe	1	4.107	4.107	0.684	0.010

Notes A data file is available at www.cesifo.de/woessmann#data or http://hanushek.stanford.edu/download

[a] Indicator of whether country is in the main sample of 50 countries contained in the growth regressions, for which internationally comparable GDP data are available

[b] Average test score in math and science, primary through end of secondary school, all years (scaled to PISA scale divided by 100)

[c] Average test score in math and science, only lower secondary, all years (scaled to PISA scale divided by 100)

[d] Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school, all years)

[e] Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years)

combined measure for the 77 countries that have ever participated in any of the math and science tests.[65] The sample for our growth regressions contains 50 of these countries.[66]

Distributional measures. Apart from the mean scores, we also analyze the distribution of test scores in each country by accessing the microdata of all ISATs.[67] The kernel density plots for math achievement on the 2003 PISA in Fig. 4 show that countries vary significantly in their patterns of test-score distributions. The depicted selected examples of developed countries reveals that it is possible to achieve relatively high median performance both with a relatively equal spread (Finland) and with a relatively unequal spread (Belgium) in the test scores at the student level. The same is true for countries with low average performance such as the depicted developing countries, where Brazil has a long right tail in contrast to Indonesia which shows a much greater density around its median.

To depict both ends of the distribution, we aim to calculate both the share of students reaching a basic level of literacy in the different subjects equivalent to 400 test-score points on the PISA scale (one student-level s.d. below the OECD mean) and the share of students reaching a top performance level equivalent to 600 test-score points on the PISA scale.

To do so, we use the above transformation to translate these two thresholds into the specific metric of each ISAT. Using the microdata of each ISAT, we then calculate the share of students in each country reaching the thresholds in the overall distribution of the ISAT. The information from the different ISATs is again combined by taking a simple average of the shares across tests.

Trends over time. The standardized performance information over a long period of time also allows us to derive longitudinal patterns of test scores for countries that participated both in early and recent ISATs. Given the amplification of noise in first-differenced data and the limitations of our rescaling method for poorly performing countries mentioned above, we perform the trend estimation only for the sample of 15 OECD countries that participated both in an ISAT before 1985 (i.e., on FIMS, FISS, FIRS, SIMS, or SISS) and up to 2003, spanning a period of nearly 20 years. (Twelve countries participated in a test before 1971, spanning a period of over 30 years.)

To estimate the trend in test performance, for each country we regress performance on the different ISATs, expressed on the standardized test metric developed above, on dummies for the subjects, dummies for the age groups, and on the year the test was conducted. The unit of observation in these country-specific regressions is each subject-by-age-by-year occasion of an ISAT, using all available tests, subjects, and age groups (see Appendix Table 10). To account for heteroscedasticity and for the fact that the signal-to-noise ratio will be larger the smaller the number of OSG countries that participated in a test, we weight the regression by

[65] The sources of the underlying international test data are: Beaton et al. (1996a, 1996b), Lee and Barro (2001), Martin et al. (1997, 2000, 2004), Mullis et al. (1997, 1998, 2000, 2003, 2004), OECD (2001, 2003b, 2004), and own calculations based on the microdata of the early tests, and own calculations based on the microdata of the early tests.

[66] Twenty-five of the total of 77 countries with cognitive-skill data are not included in the growth database due to lack of data on economic output or because they drop out of the sample for a standard exclusion criterion in growth analyses (15 former communist countries, 3 countries for which oil production is the dominant industry, 2 small countries, 3 newly created countries, 2 further countries lacking early output data). In addition, two strong outliers are excluded in most models (see above). There are four countries with cognitive-skill data that have a few years of economic data missing at the beginning or end of the 1960–2000 period. Data for Tunisia start in 1961, and data for Cyprus (1996), Singapore (1996), and Taiwan (1998) end slightly earlier than in 2000. These countries were included in the growth regressions by estimating average annual growth over the available 36-to-39-year period.

[67] Unfortunately, the microdata from the FIMS test do not seem to be available in an accessible way any more, so that the distributional measures only draw on the remaining ISATs.

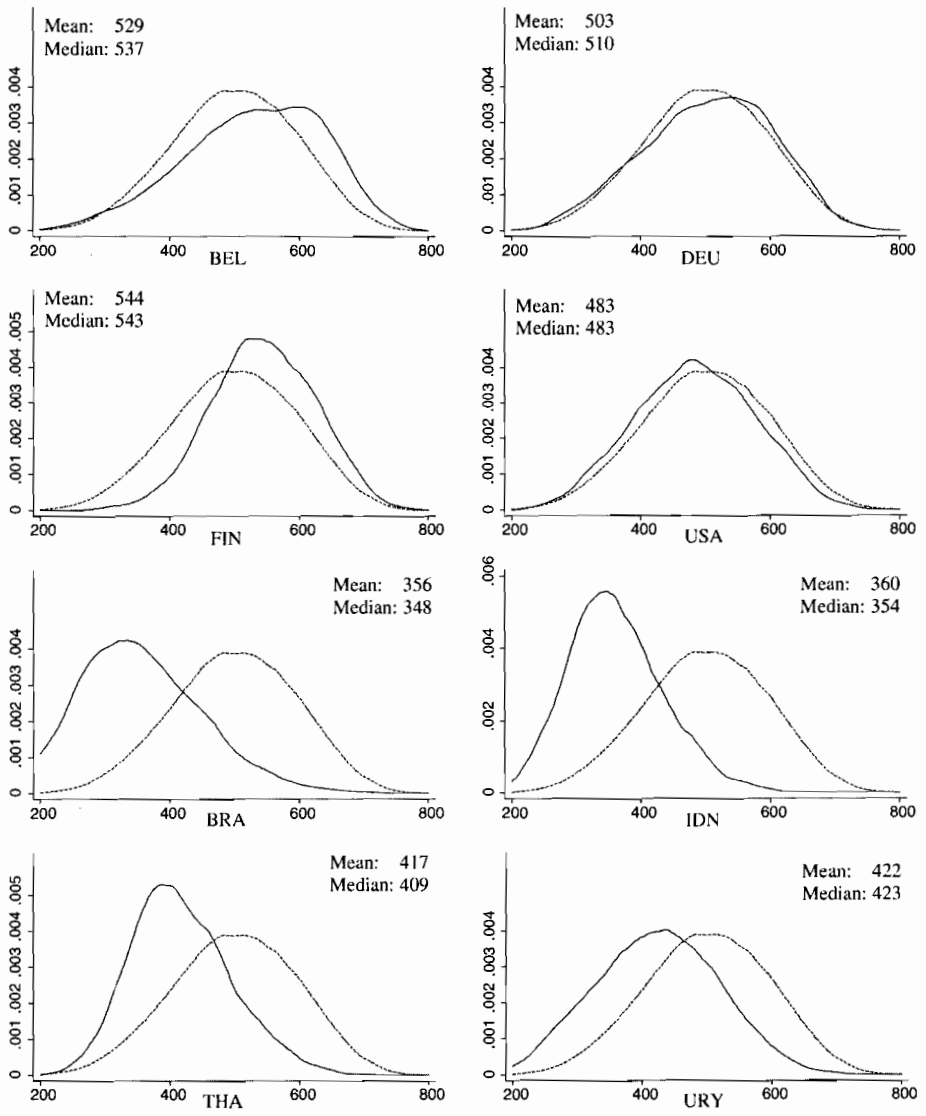


Fig. 4 Selected Examples of the Distribution of Student Performance. *Notes*: Kernel densities of student performance on the PISA 2003 math test. Bold solid line: specified country; thin dotted line: OECD countries. Codes: Belgium (BEL), Germany (DEU), Finland (FIN), United States of America (USA), Brazil (BRA), Indonesia (IDN), Thailand (THA), Uruguay (URY)

the square root of the number of OSG countries participating in each test. The coefficient on the year variable provides us with the time trend that we are interested in. The patterns captured by these country-specific regressions are shown in Fig. 5 that simply extrapolates scores for the range of 1975-2000 with scores anchored by the PISA 2000 score.

One possible worry with combining the different tests into a single measure and with estimating performance changes over time is that enrollment shares have changed to different extents over time, especially at the secondary level. To test the extent to which this affects our

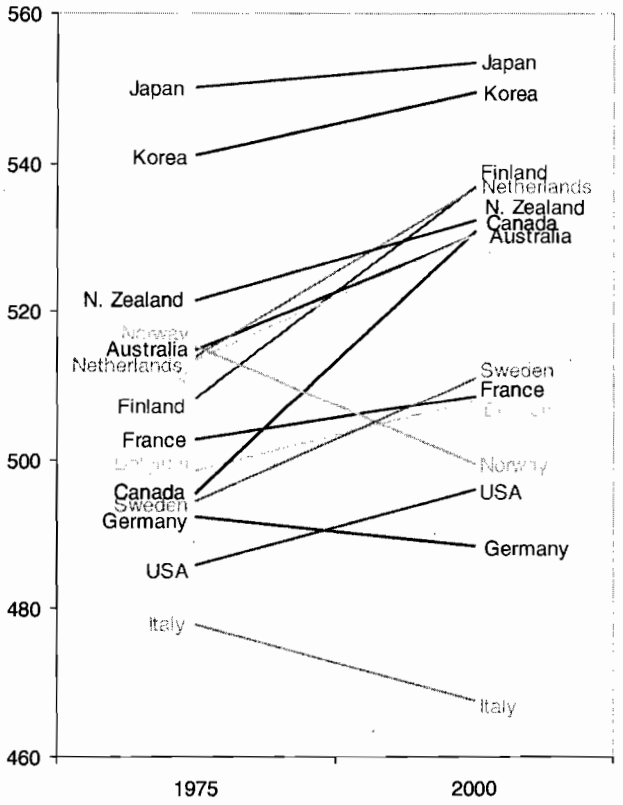


Fig. 5 Trends in Test Scores Notes: Depiction based on PISA 2000 performance and a backward induction based on the coefficient on a time variable from a regression of all available international test scores (by year, age group, and subject) on the time variable and dummies for age group and subject. See Appendix B for details

cognitive-skill measures, we calculated the correlation between our measure of trend in test scores and changes in enrollment rates. It turns out that the two are orthogonal to each other, diluting worries that differential changes in enrollment bias the results reported in this paper.[68]

Comparisons with Cognitive Skills Measure of Hanushek and Kimko (2000). The skill measures developed in Hanushek and Kimko (2000) fail to account for the unequal variances of the tests over time, but instead assume a constant variance.[69] Our measure is highly correlated with the Hanushek-Kimko measure ($r = 0.83$), but the important question is the relationship with growth. For the 30 countries in common from the two data sets, we estimate the growth models with the alternative measures of cognitive skills. While both versions of the test-score measure enter the model strongly and significantly, statistical precision is considerably higher with the new measure ($t = 7.43$ vs. $t = 4.02$), as is the explanatory power of the model (adj. $R^2 = 0.80$ vs. adj. $R^2 = 0.61$). The content of signal

[68] As noted previously, direct estimation of the impacts of school selectivity and of test exclusions on our growth models confirm that potential testing problems do not bias our growth estimation (Hanushek and Woessmann 2011c).

[69] Hanushek and Kimko (2000) actually have alternative measures. Two of their three measures assume a constant mean for all of the tests, similar to what is also done in Lee and Barro (2001).

relative to noise in the test-score measure thus seems to be considerably raised in the new measure.

Data download

The data by country may be downloaded in Excel or Stata formats at:
http://hanushek.stanford.edu/download or www.cesifo.de/woessmann#data.

Data sources

Beaton, Albert E., Ina V.S. Mullis, Michael O. Martin, Eugenio J. Gonzalez, Dana L. Kelly, and Teresa A. Smith. 1996. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Beaton, Albert E., Michael O. Martin, Ina V.S. Mullis, Eugenio J. Gonzalez, Teresa A. Smith, and Dana L. Kelly. 1996. *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Lee, Jong-Wha, and Robert J. Barro. 2001. Schooling Quality in a Cross-Section of Countries. *Economica* 68, no. 272: 465-488.

Martin, Michael O., Ina V.S. Mullis, Albert E. Beaton, Eugenio J. Gonzalez, Teresa A. Smith, and Dana L. Kelly. 1997. *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Martin, Michael O., Ina V.S. Mullis, Eugenio J. Gonzalez, Kelvin D. Gregory, Teresa A. Smith, Steven J. Chrostowski, Robert A. Garden, and Kathleen M. O'Connor. 2000. *TIMSS 1999 International Science Report:Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Chestnut Hill, MA: Boston College.

Martin, Michael O., Ina V.S. Mullis, Eugenio J. Gonzalez, and Steven J. Chrostowski. 2004. *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Chestnut Hill, MA: Boston College.

Mullis, Ina V.S., Michael O. Martin, Albert E. Beaton, Eugenio J. Gonzalez, Dana L. Kelly, and Teresa A. Smith. 1997. *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

—. 1998. *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez, Kelvin D. Gregory, Robert A. Garden, Kathleen M. O'Connor, Steven J. Chrostowski, and Teresa A. Smith. 2000. *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez, and Ann M. Kennedy. 2003. *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary School in 35 Countries*. Chestnut Hill, MA: International Study Center, Boston College.

Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez, and Steven J. Chrostowski. 2004. *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

Organisation for Economic Co-operation and Development (OECD). 2001. *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

—. 2003a. *Education at a Glance: OECD Indicators 2003*. Paris: OECD.

—. 2003b. *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris: OECD.

—. 2004. *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.

U.S. Department of Education, Institute of Education Sciences. 2008. *National Assessment of Educational Progress–The Nation's Report Card*. Website:http://nces.ed.gov/nationsreportcard/aboutnaep.asp.

Appendix C: Descriptive statistics

See Appendix Tables 12, 13, 14 and Figure 6.

Table 12 Descriptive statistics for the growth models

	Observations	Mean	Std. Dev.	Min	Max
Average annual growth rate in GDP per capita 1960–2000	50	2.903	1.387	0.967	6.871
Cognitive skills (all grades)	50	4.546	0.611	3.089	5.452
Cognitive skills (lower secondary)	50	4.535	0.671	2.683	5.599
Share of students reaching basic literacy	50	0.761	0.215	0.182	0.967
Share of top-performing students	50	0.062	0.054	0.000	0.219
GDP per capita 1960	50	4,991	3,676	685	14,877
Years of schooling 1960	50	5.447	2.877	0.611	10.963
Years of schooling, average 1960–2000	50	7.425	2.654	2.098	11.845

Table 12 continued

	Observations	Mean	Std. Dev.	Min	Max
External exit exam system	43	0.661	0.467	0	1
Private enrollment share	19	0.186	0.206	0	0.720
Centralization (share) of decisions on organization of instruction	17	0.104	0.117	0	0.380

Notes Descriptive statistics for variables used in Tables 1, 2, 3, 4, and 7. See main text for data sources

Table 13 Descriptive statistics for the U.S.-immigrant models

	Observations	Mean	Std. dev.	Min	Max
Annual earnings	309,574	33,243	40,983	1,000	385,000
Cognitive skills	309,574	4.334	0.535	3.089	5.452
Educated in country of origin	309,574	0.837	0.370	0	1
Years of schooling	309,574	11.558	5.006	0	20
Potential experience	309,574	24.841	11.966	0	87

Notes Descriptive statistics for variables used in Table 5. See main text for data sources

Table 14 Descriptive statistics for the changes-in-growth-paths models

	Observations	Mean	Std. Dev.	Min	Max
Trend in growth rate of GDP per capita 1975–2000	15	−0.007	0.071	−0.118	0.106
Trend in cognitive skills	15	0.409	0.546	−0.630	1.420
Average annual growth rate in GDP per capita 1975–2000	15	2.318	1.106	0.855	5.978
GDP per capita 1975	15	13,884	3.217	3.720	18,175
Change in years of schooling 1975–2000	15	1.994	0.895	0.899	4.376

Notes Descriptive statistics for variables used in Table 6. See main text for data sources

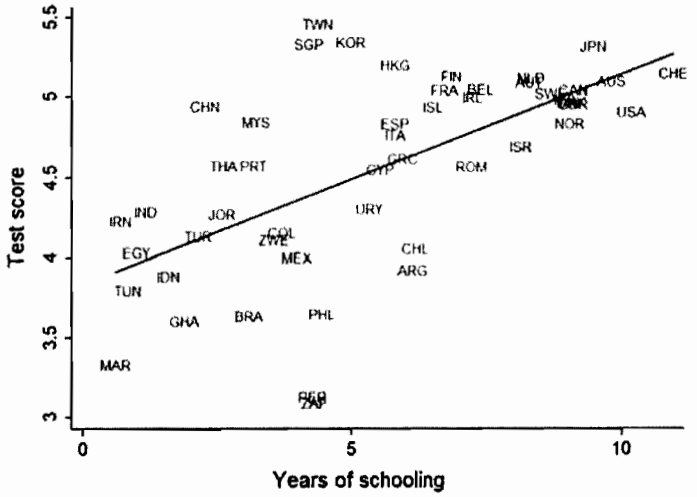


Fig. 6 Years of Schooling and Test Scores. *Notes*: Scatter plot of years of schooling in 1960 against cognitive skills (average test scores in math and science, primary through end of secondary school, all years)

Appendix D: Impact of data updates and expansion of observed growth period

The analysis in this paper concentrates on measurement issues surrounding cognitive skills. But these are not the only measurement issues that have been raised in the context of empirical growth analysis. This appendix provides an analysis of the impact of other data measurement issues in the area of data on school attainment and on economic growth. It also considers the most recent wave of cognitive skills data. These updates and revisions do not substantially change the results presented in our main analyses.

The measurement of school attainment has been discussed at various times. Barro and Lee (1993) developed the initial international database for school attainment. This was criticized by Cohen and Soto (2007), who produced the dataset that is the basis of our estimation. Most recently, Barro and Lee (2010) have produced a new dataset of their own. Using the latest series of school attainment developed by Barro and Lee (2010) (data version 1.0, 03/10, accessed on May 17, 2010) has no effect on the estimates of the impact of cognitive skills and only slightly increases the attainment coefficient. As shown in Table 15, column 2, the estimated cognitive skills coefficient is 1.92, as opposed to 1.98 in our base estimates (column 1).

A second area of concern is the underlying economic data themselves. It has recently been argued that changes in the Penn World Tables (PWT), particularly the 6.2 revision, lead to significant alternations in standard estimates of growth models (Ciccone and Jarocinski (2010); Appleton et al. (2011)).[70] To assess the impact of these data, we compare the most recent version of the PWT (version 7.0, released on June 3, 2011; Heston et al. (2011)) to our estimates based on version 6.1. As is apparent from column 3, the estimate on cognitive skills is hardly affected when estimating the same model with the same growth period (1960-2000) using the new economic data; in fact, the point estimate is slightly higher.

The new version of the PWT allows us to expand the growth period to 2007. Again, results are strongly confirmed in this 47-year growth period (column 4), with the point estimate closer to the original estimate again. We use 2007 as the endpoint of the considered growth period rather than 2009 (which is the latest year available in the new PWT) because we do not want the long-run growth analysis being affected by the global recession that started at the end of 2008 (as clearly visible in the PWT data, where average growth rates in our sample drop from around 4 % in the preceding years to 1.6 in 2007-2008 and to -2.5 in 2008-2009). Still, our result on cognitive skills is confirmed also in the 1960-2009 growth period (not shown), with a point estimate of 1.76 ($t = 7.32$).

The expanded PWT data also allow us to perform an analysis where the observation period of test scores strictly pre-dates the observation period of economic growth. For a sample of 25 countries, we have test-score data observed between 1964 and 1984. Column 5 uses these test-score data to predict economic growth in 1985-2007. Again, the significant effect of cognitive skills on growth is confirmed, with a point estimate substantially larger than in the base model. (Again, results are very similar for the growth period expanded to 2009.)

While we are reluctant to perform analyses on shorter growth periods because these are prone to country-specific shocks and business-cycle fluctuations, we can expand our sample of countries with test scores pre-dating the observed growth period to 37 countries when we use the tests conducted until 1995 to predict growth in 1995-2009 or in 2000-2009. Again, results confirm a strong estimate on cognitive skills with a point estimate larger than in

[70] As noted in section 7 above, calculation of growth rates relying on national account data rather than on purchasing power parity incomes does not significantly affect our estimates.

Table 15 Growth regressions with updated data series on school attainment and economic growth

PWT version	6.1	6.1	7.0	7.0	7.0
Years of schooling data	Cohen and Soto (2007)	Barro and Lee (2010)	Barro and Lee (2010)	Barro and Lee (2010)	Barro and Lee (2010)
Growth period	1960–2000	1960–2000	1960–2000	1960–2007	1985–2007
Test scores	All years	All years	All years	All years	Until 1984
	(1)	(2)	(3)	(4)	(5)
Cognitive skills	1.980	1.921	2.133	1.881	3.593
	(9.12)	*(9.25)*	*(9.01)*	*(7.78)*	*(2.56)*
Initial years of schooling	0.026	0.079	0.018	0.018	−0.079
	(0.34)	*(1.09)*	*(0.23)*	*(0.22)*	*(0.41)*
Initial GDP per capita	−0.302	−0.324	−0.219	−0.212	−0.123
	(5.54)	*(7.01)*	*(5.59)*	*(5.29)*	*(2.20)*
No. of countries	50	50	50	50	25
R^2 (adj.)	0.728	0.734	0.667	0.610	0.318

Notes Dependent variable: average annual growth rate in GDP per capita. Regressions include a constant. Test scores are average of math and science, primary through end of secondary school. Absolute t-statistics in parentheses

Table 16 Changes in cognitive skills and changes in growth paths with updated data series on economic growth

Period of trend in growth	1975–2000 (1)	1975–2007 (2)	1985–2007 (3)
Trend in cognitive skills	0.072	0.072	0.115
	(2.61)	*(2.92)*	*(2.94)*
Average annual growth rate	−0.017	−0.022	−0.067
in GDP per capita over period	*(1.30)*	*(1.69)*	*(3.16)*
No. of countries	15	15	15
R^2 (adj.)	0.316	0.387	0.516

Notes Dependent variable: trend in the growth rate of GDP percapita over the period shown in the header, based on PWT version 7.0. Regressions include a constant. Sample: OECD countries with test-score data both before 1985 and up to 2003. Test scores are average of math and science, primary through end of secondary school. Absolute t-statistics in parentheses

the base model (not shown), although the precise point estimate is sensitive to excluding individual countries when considering this shorter growth period.

We have also experimented with the latest wave of test-score data, the 2009 wave of the PISA study. Our test-score measure derived from the tests conducted in 1964-2003 is strongly correlated with the PISA 2009 data (correlation coefficient of 0.94 for the 37 countries available in both datasets). This corroborates the assumption of relative stability underlying our main analysis. Also, using the PISA 2009 as an alternative measure of cognitive skills in the growth regressions fully confirms our results, with a highly significant point estimate extremely close to our main analysis (1.96).

Finally, Table 16 shows that our analysis in Section 7 on trends in skills and in growth is strongly confirmed with the new 7.0 version of the PWT. Considering the revised data over the same 1975-2000 period (col. 1) hardly affects the result on cognitive skills from col. 2 of Table 6. Similarly, extending the growth period with the newly available data to 2007 (or to 2009, not shown) confirms the previous result (col. 2). Similarly, when using the growth period 1985-2007 (col. 3), so that the test-score trend partly predates the growth rate trend, even strengthens the result of a significant positive relation of trends in test scores with trends in growth rates.

Acknowledgments We thank Mark Bils, Pete Klenow, Giovanni Peri, Paul Romer, participants at seminars at Stanford University, the London School of Economics, the Universities of Stockholm, Mannheim, Heidelberg, Uppsala, and Rochester, Simon Fraser University, the European School of Management and Technology in Berlin, the City University of Hong Kong, the World Bank, the NBER growth conference in San Francisco, and the Econometric Society World Congress in Shanghai, and anonymous referees for helpful comments and suggestions. We are grateful to the International Association for the Evaluation of Educational Achievement (IEA) for providing us with the micro data for the FISS, SIMS, and SISS studies. Jason Grissom and Trey Miller provided important research assistance. Support has come from CESifo, the Packard Humanities Institute, the W. Glenn Campbell and Rita Ricardo-Campbell National Fellowship of the Hoover Institution, and the Pact for Research and Innovation of the Leibniz Association.

References

Acemoglu, D., Johnson, S., & Robinson, A. R. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review, 91*(5), 1369–1401.

Acemoglu, D., Johnson, S., & Robinson, A. R. (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (pp. 385–472). Amsterdam: North Holland.

Aghion, P., & Howitt, P. (1998). *Endogenous growth theory*. Cambridge, MA: MIT Press.

Aghion, P., Boustan, L., Hoxby, C. M., & Vandenbussche, J. (2009). The causal impact of education on economic growth: Evidence from the U.S. (mimeo) Department of Economics: Harvard University (March).

Andrews, D. R., Moreira, M. J., & Stock, J. H. (2007). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics, 139*(1), 116–132.

Appleton, S., Atherton, P., & Bleaney, M. (2011). Growth regressions and data revisions in Penn World Tables. *Journal of Economic Studies, 38*(3), 301–312.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The polarization of the U.S. labor market. *American Economic Review, 96*(2), 189–194.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics, 90*(2), 300–323.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics, 106*(2), 407–443.

Barro, R. J. (1997). *Determinants of economic growth: A cross-country empirical study*. Cambridge, MA: MIT Press.

Barro, R. J. (2001). Human capital and growth. *American Economic Review, 91*(2), 12–17.

Barro, R. J., & Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics, 32*(3), 363–394.

Barro, R. J., & Lee, J.-W. (2010). A new data set of educational attainment in the world, 1950-2010. NBER Working Paper 15902. Cambridge, MA: National Bureau of Economic Research (April).

Barro, R. J., & McCleary, R. M. (2003). Religion and economic growth across countries. *American Sociological Review, 68*(5), 760–781.

Barro, R. J., & Ursúa José, F. (2008). Macroeconomic crises since 1870. *Brookings Papers on Economic Activity, 1*, 336–350.

Bils, M., & Klenow, P. J. (2000). Does schooling cause growth? *American Economic Review, 90*(5), 1160–1183.

Bishop, J. H. (2006). Drinking from the fountain of knowledge: Student incentive to study and learn—Externalities, information problems, and peer pressure. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (pp. 909–944). Amsterdam: North Holland.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature, 39*(4), 1137–1176.

Castelló, A., & Doménech, R. (2002). Human capital inequality and economic growth: Some new evidence. *Economic Journal, 112*(478), C187–C200.

Ciccone, A., & Jarocinski, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics, 2*(4), 222–246.

Ciccone, A., & Papaioannou, E. (2009). Human capital, the structure of production, and growth. *Review of Economics and Statistics, 91*(1), 66–82.

Cingano, F., Leonardi, M., Messina, J., & Pica, G. (2010). The effects of employment protection legislation and financial market imperfections on investment: evidence from a firm-level panel of EU countries. *Economic Policy, 25*(61), 117–163.

Cohen, D., & Soto, M. (2007). Growth and human capital: Good data, good results. *Journal of Economic Growth, 12*(1), 51–76.

Coulombe, S., & Tremblay, J.-F. (2006). Literacy and growth. *Topics in Macroeconomics 6*(2), Article 4.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31–47.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education*. (pp. 697–812). Amsterdam: Elsevier.

Dolton, P., & Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross country analysis of teacher pay and pupil performance. *Economic Policy, 26*(65), 7–55.

Ehrlich, I. (2007). The mystery of human capital as engine of growth, or why the US became the economic superpower in the 20th Century. NBER Working Paper 12868. Cambridge, MA: National Bureau of Economic Research (January).

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica, 45*(4), 939–954.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth, 9*(3), 271–303.

Glenn, C. L., & De Groof, J. (2002). *Finding the right balance: Freedom, autonomy and accountability in education, Vol. II*. The Netherlands: Lemma Publishers.

Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.

Gundlach, E., Woessmann, L., & Gmelin, J. (2001). The decline of schooling productivity in OECD countries. *Economic Journal, 111*(May), C135–C147.

Hahn, J., Hausman, J. A., & Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econometrics Journal, 7*(1), 272–306.

Hanousek, J., Hajkova, D., & Filer, R. K. (2008). A rise by any other name? Sensitivity of growth regressions to data source. *Journal of Macroeconomics, 30*(3), 1188–1206.

Hanushek, E. A. (2002). Publicly provided education. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of Public Economics*. (pp. 2045–2141). Amsterdam: Elsevier.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review, 90*(5), 1184–1208.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668.

Hanushek, E. A., & Woessmann, L. (2011a). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3* (pp. 89–200). Amsterdam: North Holland.

Hanushek, E. A., & Woessmann, L. (2011b). How much do educational outcomes matter in OECD countries? *Economic Policy, 26*(67), 427–491.

Hanushek, E. A., & Woessmann, L. (2011c). Sample selectivity and the validity of international student achievement tests in economic research. *Economic Letters, 110*(2), 79–82.

Hanushek, E. A., & Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics* (forthcoming).

Hanushek, E. A., & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital, 3*(2), 107–143.

Hausmann, R., Pritchett, L., & Rodrik, D. (2005). Growth accelerations. *Journal of Economic Growth, 10*(4), 303–329.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2008). Earnings functions and rates of return. *Journal of Human Capital, 2*(1), 1–31.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.

Hendricks, L. (2002). How important is human capital for development? Evidence from immigrant earnings. *American Economic Review, 92*(1), 198–219.

Heston, A., Summers, R., & Aten, B. (2002). Penn World Table Version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP). Philadelphia: University of Pennsylvania.

Heston, A., Summers, R., & Aten, B. (2011). Penn World Table Version 7.0. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania Philadelphia: University of Pennsylvania (June 3 update).

Jones, B. F., & Olken, B. A. (2008). The anatomy of start-stop growth. *Review of Economics and Statistics, 90*(3), 582–587.

Katz, L. F., & Autor, D. H. (1999). Changes in the wage structure and earnings inequality. In A. Orley & D. Card (Eds.), *Handbook of Labor Economics*. (pp. 1463–1558). Amsterdam: Elsevier.

Krueger, A. B., & Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature, 39*(4), 1101–1136.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review, 10*(3), 179–214.

Lee, J.-W., & Barro, R. J. (2001). Schooling quality in a cross-section of countries. *Economica, 68*(272), 465–488.

Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review, 82*(4), 942–963.

Lucas, R. E., Jr. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*(1), 3–42.

Mankiw, N. G., Romer, D., & Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics, 107*(2), 407–437.

Mincer, J. (1974). *Schooling, experience, and earnings*. New York: NBER.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica, 71*(4), 1027–1048.

Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy 107*(6), pt. 2 (December): S184–S224.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management, 19*(4), 547–568.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics, 77*(2), 251–266.

Neidorf, T. S., Binkley, M., Gattis, K., & Nohara, D. (2006). *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 assessments*. Washington: National Center for Education Statistics.

Nelson, R. R., & Phelps, E. (1966). Investment in humans, technology diffusion and economic growth. *American Economic Review, 56*(2), 69–75.

Nicoletti, G., & Scarpetta, S. (2003). Regulation, productivity, and growth: OECD evidence. *Economic Policy, 18*(1), 10–72.

Nuxoll, D. A. (1994). Differences in relative prices and international differences in growth rates. *American Economic Review, 84*(5), 1423–1436.

Organisation for Economic Co-operation and Development. (1998). *Education at a glance: OECD Indicators*. Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.

Pritchett, L. (2006). Does learning to add up add up? The returns to schooling in aggregate data. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education*. (pp. 635–695). Amsterdam: North Holland.

Ramirez, F., Luo, X., Schofer, E., & Meyer, J. (2006). Student achievement and national economic growth. *American Journal of Education, 113*(1), 1–29.

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature, 39*(2), 432–478.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy 99*(5), pt. II: S71–S102.

Sachs, J. D., & Warner, A. M. (1995). Economic reform and the process of global integration. *Brookings Papers on Economic Activity* (1),1–96.

Sala-i-Martin, X., Doppelhofer, G., & Miller, R. I. (2004). Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach. *American Economic Review, 94*(4), 813–835.

Schoellman, T. (2012). Education quality and development accounting. *Review of Economic Studies, 79*(1), 388–417.

Spence, A. M (1973). Job market signalling. *Quarterly Journal of Economics, 87*(3), 355–374.

Topel, R. (1999). Labor markets and economic growth. In A. Orley & D. Card (Eds.), *Handbook of Labor Economics*. (pp. 2943–2984). Amsterdam: Elsevier.

UNESCO. (1998). *World education report, 1998: Teachers and teaching in a changing world*. Paris: UNESCO.

UNESCO. (2005). *Education for all: The quality imperative, EFA Global Monitoring Report*. Paris: UNESCO.

Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth, 11*(2), 97–127.

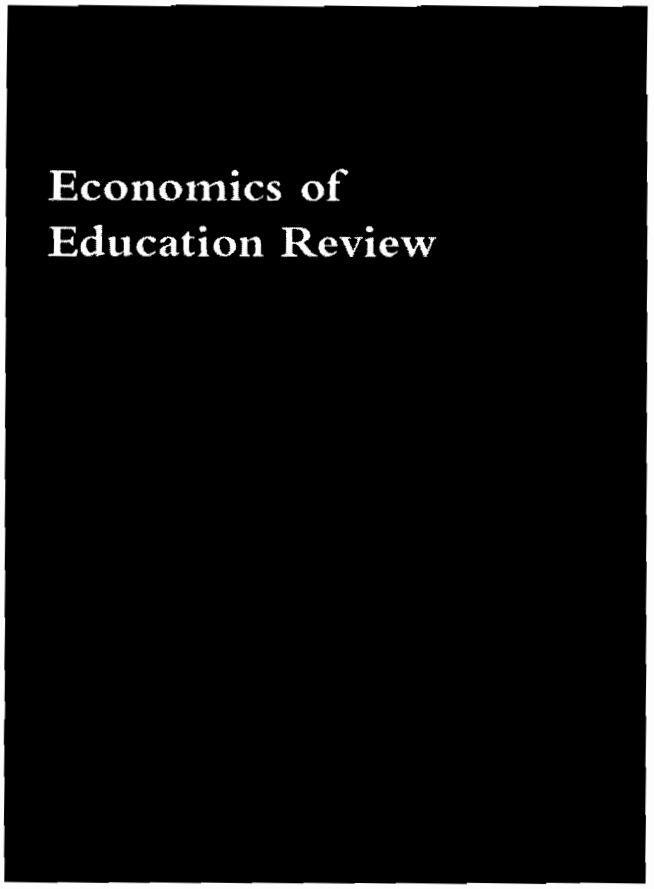
Weiss, A. (1995). Human capital vs. signalling explanations of wages. *Journal of Economic Perspectives, 9*(4), 133–154.

West, M. R., & Woessmann, L. (2010). 'Every Catholic child in a Catholic school': Historical resistance to state schooling, contemporary private competition and student achievement across countries. *Economic Journal, 120*(546), F229–F255.

Woessmann, L. (2003a). Central exit exams and student achievement: International evidence. In P. E. Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 292–323). Washington D.C.: Brookings Institution Press.

Woessmann, L. (2003b). Schooling resources, educational institutions, and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics, 65*(2), 117–170.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2009). *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.



Economics of Education Review 31 (2012) 225–236

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Choices for studying choice: Assessing charter school effectiveness using two quasi-experimental methods

Devora H. Davis, Margaret E. Raymond*

Center for Research on Education Outcomes (CREDO), Stanford University, 434 Galvez Mall, Stanford, CA 94305-6010, United States

ARTICLE INFO

Article history:
Received 6 January 2011
Received in revised form 20 August 2011
Accepted 24 August 2011

JEL classification:
I21

Keywords:
School choice
Educational economics

ABSTRACT

Two quasi-experimental methods – fixed effects (FE) and virtual control records (VCR) – were used to measure charter schooling in 14 states and two districts. The new VCR method uses all available observable charter student characteristics and prior performance to create a composite comparison record. A head-to-head comparison of the FE and VCR methods used the same charter students to test the FE control (e.g., the charter student's own traditional public school experience) and the VCR for equivalence. The comparison produced highly similar estimates; charter coefficients were identical in sign and significance and of the same general magnitudes. In an analysis of the sampling fractions included in each method using all available tested charter students, the VCR method was found to produce more generalizable results. In the policy analysis, charter school quality was found to be demographically and geographically uneven with only 19 percent of charter schools outperforming their local markets.

1. Introduction

Media attention toward charter schools tends to either demonize or canonize their practices, and data is regularly marshaled to strengthen the case. In the midst of this often highly charged and confusing environment, policymakers are tasked with important decisions ranging from the fate of individual charter schools to the overall effectiveness of charter schools as a tool for improving student achievement. To responsibly execute their duties, policymakers must have dependable – and understandable – information about charter school performance.

In attempting to provide this information, the research community has delivered an array of studies of charter schools, employing a variety of analytic techniques. When different techniques yield differing results, and the effort is not made to explain the sources of difference, the research community inadvertently contributes to the confusion. Policymakers, who view research from a practical rather than an academic standpoint, need guidance from the research community about the strengths and limitations of the various analytic techniques.

The research on charter schools has not kept up with the increasing importance of charter schooling as a policy alternative. Until recently, the research literature on charter schools was constrained to studies within single states. Accordingly, the insights likewise were constrained to within-state or even within-city variation, in which all schools would be covered by the same policies and changes to them. This led to two flawed inferences: either policy thinkers took each study result to be indicative of the entire charter school population or they rejected the results of studies conducted outside of their own state.

Policymakers, educators and researchers alike need a better understanding of the benefits and limitations of the different study designs used to measure charter school performance. This is a two-way responsibility. On the one hand, decision-makers need to be sensitive to the analytic methods used when assessing the relative strengths

* Corresponding author. Tel.: +1 650 725 3431; fax: +1 650 723 1687.
E-mail addresses: devdavis@stanford.edu (D.H. Davis), macke@stanford.edu (M.E. Raymond).

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.08.012

of results from different studies. On the other hand, the research community needs to be clearer in their discussions about the suitability and limitations of the techniques they choose to illuminate various policy questions.

The purpose of this paper is to begin such a discussion of two quasi-experimental methods that have been recently used to measure charter school effectiveness on a large scale. These two techniques are student fixed effects (FE) and virtual control records (VCR), and they are compared here using an unusually large multi-state student-level dataset. Student fixed effects is the most commonly used quasi-experimental method in education research. The virtual control record methodology, in contrast, is a newly developed technique that was designed to address some of the limitations of student fixed effects.

Two issues are central to the debate about analytic methods: selection bias and generalizability. Minimizing the former often comes at the expense of the latter, as in the case of lottery studies, which draw on the experience of a limited number of charter schools that randomly select students for admission. Quasi-experimental methods such as student fixed effects include greater proportions of students than lottery studies but introduce issues of selection bias to varying degrees. What remains an open question is whether generalizability can be improved in the quasi-experimental realm while maintaining low levels of selection bias. The research presented here compares the advantages and disadvantages of two quasi-experimental approaches, student fixed effects and virtual control records, with a focus on the issues of selection bias and generalizability.

To address both issues, two different comparisons were conducted. The first, discussed in Section 2, creates a head-to-head test of FE and VCR designs. The test involves the control conditions of the two methods, restricting the analysis to only the charter students who switch between traditional public and charter school sectors and the control specifications used in the two approaches. This computational comparison will effectively test whether the students acting as their own controls in the FE model produce equivalent results to those of the virtual controls in the VCR model, i.e., whether selection bias is comparable between the two methods.

Section 3 compares the results for FE and VCR models that are not restricted a priori. The biggest difference between the two methods is that fixed effects models are limited to students who are observed in both a traditional public school and a charter school, i.e., sector switchers. The VCR method allows controls to be found for both sector switchers and charter-only students. The policy insights gained from each model will be the main focus of the second comparison.

1.1. Methods descriptions

In recent years, the methods debate in academia has focused on the relative merits of experimental and quasi-experimental research designs. In charter school research, these are most often lottery studies and student fixed effects. Fig. 1 summarizes these two methods and their strengths and limitations along with the new quasi-experimental methodology examined in this paper, virtual control records.

1.1.1. Lottery studies

The ideal study would take the population of traditional public school (TPS) students and randomly place some in charter schools while keeping the rest in the TPS sector. Such an experimental design would eliminate selection bias from the estimates of charter school effectiveness.

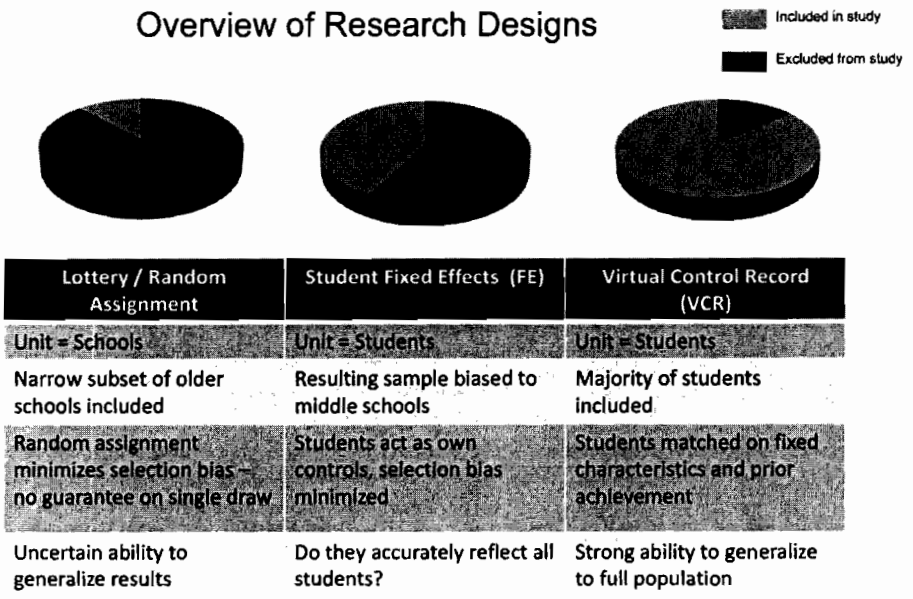


Fig. 1. Overview of research designs.

Although there are no cases of true random assignment between sectors, the rules of entry into charters mimic many aspects. In most states, if a charter school is oversubscribed (i.e., more children want to attend than there are spaces available), the charter school must select students at random by holding a lottery for entry. Thus, on average, the losers of the lottery would be expected to be a good comparison group for the winners on both observable and unobservable characteristics such as motivation, thereby giving lottery studies a high degree of internal validity. This approach was employed in recent analyses of charter school effectiveness in Boston and New York City and, in one study, in multiple locations across 15 states (Abdulkadiroglu et al., 2009; Hoxby & Murarka, 2009; Gleason, Clark, Tuttle, & Dwoyer, 2010).

However, charter school lotteries have an unfortunate limitation: The conditions that give rise to lotteries, namely over-subscription, are neither universal nor randomly distributed across schools. Research suggests that oversubscribed schools are older and academically better than the average charter school (Abdulkadiroglu et al., 2009; Hoxby & Murarka, 2007; Tuttle, Gleason, & Clark, 2012). There are also practical limitations. It only takes one student beyond a school's capacity to trigger a lottery, and many schools have subscription rates less than 125 percent of capacity (Peterson, 2003). In these cases, random selection does not produce equal numbers of students to serve as controls, and therefore can lead to problems in the estimation phase. Consequently, only a fraction of oversubscribed schools can fully support a lottery study. Thus, the results that arise from lottery studies are difficult to generalize to the full population of charter schools.

1.1.2. Student fixed effects

The alternative to lottery studies is to rely on quasi-experimental approaches, and the most commonly used approach in the charter school literature is student fixed effects (Ballou, Teasley, & Zeidner, 2008; Hanushek, Kain, Rivkin, & Branch, 2007; Zimmer & Buddin, 2008; Zimmer, Gill, Booker, Lavertu, & Witte, 2012). This method uses a panel dataset and is relatively more inclusive of schools than lottery studies. The design includes only the students who have at least one growth period in each sector: TPS and charter. Each student's TPS experience essentially serves as his or her own control condition for the charter school experience, which minimizes many potential sources of selection bias. One notable exception is the occurrence of an event that coincides with (or immediately precedes) a student's transfer between sectors and also influences the student's learning gains. In this case, estimates of the charter school effect will be biased (Hoxby & Murarka, 2008; Hanushek et al., 2007).

Like lottery studies, there are concerns about the generalizability of student fixed effects studies. First, mobility between the TPS and charter sectors is not extremely high, as will be shown below. Furthermore, as the charter sector expands, the number of students who never attend a TPS school is likely to grow. To the extent that these so-called "always charter" students look different from their sector-switching peers, either in demography or in terms of early performance, generalizing fixed effects results to the entire charter student population may become untenable (Ballou et al., 2008).

1.1.3. Virtual control records

In an attempt to produce generalizable results on national charter school effectiveness, a new quasi-experimental method was devised. The virtual control record design draws on principles used in propensity score matching and a recently developed synthetic matching technique by Abadie, Diamond, & Hainmueller (2007). At its heart is a matching methodology that creates a set of synthetic control records that closely mirrors the matched charter school students on known demographic attributes, grade in school, eligibility or participation in special support programs (subsidized lunch program eligibility, English language learner status and special education status) and a baseline test result.[1] The baseline test result reflects both earlier schooling and family and neighborhood background, including motivation, home-based learning and other inputs.

For each charter school, the virtual control record (VCR) design first identifies all the traditional public schools that had students transfer to that charter school (i.e., "feeder schools"). The students still enrolled in the feeder schools constitute the pool of potential matches. These TPS students are required to meet further criteria to be matched with a particular charter student: (1) They must attend a traditional public school for all observed time periods; (2) They must have equivalent values on all the match variables; and (3) They must have all the relevant outcome data. Matches are developed separately for reading and math. For example, a third-grade charter student will be matched to a third-grade TPS student from the feeder pool who has a similar baseline math score and the same gender, race/ethnicity, lunch program eligibility, English learner status, special education status and subsequent grade-level promotion/retention pattern. For all the TPS students who meet all the match criteria, scores from each subsequent test period are averaged across all the matched records, and a single record is produced – the VCR. The result is a composite of the expected value of experience the charter school student would have had if enrolled in the traditional public schools that compete with the charter school.[2]

The advantage of the VCR design is that it permits the inclusion of a far greater proportion of charter school students than either lottery studies or student fixed effects. Only charter students who do not have the requisite minimum of two consecutive test scores will be excluded from the sample pool. Widespread use of unique statewide student identifiers and the federal requirement that all students be tested in at least grades three through eight

[1] To be included in a synthetic control record, the baseline test result had to be no more than 0.10 of a standard deviation higher or lower than the matching charter student's baseline test score.

[2] This composite score has less noise than a nearest-neighbor matching scheme with the added benefit of survivability of the controls over time; if a student leaves the TPS sector in subsequent time periods, the VCR can be recalculated without that student.

ensures that relatively few students in elementary or middles schools will fall into this category.

The VCR method faces many of the same challenges for lower elementary and high school students as the other methods, owing to a lack of consistent test data. A concern that has been raised about the VCR design is that it could possibly underestimate the charter treatment effect for students who are exclusively enrolled in charter elementary schools. This is because the approach matches students after they have an initial test score, thereby capturing student achievement midway through the elementary charter school treatment. The fundamental question is whether performance in earlier (largely unobserved) grades at charter schools is stronger than at traditional public schools. If this is true, the VCR method would lead to an understatement of true charter school performance. The purest way to test this conjecture is to conduct a randomized trial of students in kindergarten and follow them until their first observed score. Absent that approach, we took advantage of a more limited opportunity to look at performance in earlier grades using the data amassed for this study. In Georgia student testing begins in first grade, and California tests start in second grade. In both states, the first observed achievement scores for students exclusively enrolled in charters were higher on average than those for students whose earliest score was recorded while enrolled in TPS. Whether high early achievement in the charter school sector relative to traditional public schools is due to selection bias or early charter school performance – or a combination of both – remains unclear. Two points bear noting. First, the pattern of charter student achievement in the earliest scores for both these states mirrors the patterns for growth we see in the later grades for these same students, so the same general signal is obtained even when the first score is delayed. Second, there is greater variation across states in the performance of charters relative to TPS than we see across the student groups in Georgia and California, leading to the suggestion that the VCR method does an adequate job of representing the impacts of each sector.

In the sections that follow, the FE and VCR methods are compared to determine the relative strengths and limitations of each with respect to generalizability and selection bias. Two different comparisons will be made. The first comparison, in Section 2, focuses on a comparison of the control conditions in the FE and VCR models, which necessitated using the same charter students in both models. To be included, therefore, charter students had to meet multiple criteria: (1) Be included in a regular FE model; (2) Obtain a growth score in TPS before attending a charter; and (3) Receive a VCR match. In Section 3, the second comparison of the FE and VCR models dispenses with multiple criteria for student inclusion. Instead, the FE model includes all sector switchers regardless of whether they attended a charter or TPS first or whether they have a matching virtual control. The VCR model includes all charter students with a matching virtual control regardless of whether they were included in the FE model. Throughout the rest of the paper, the models used for the head-to-head methods test discussed in Section 2 are called restricted models, while those used in the comparison discussed in Section 3 are called unrestricted models.

1.2. Data description

The laboratory for comparing the two study designs is a large multi-state student-level dataset with records that are longitudinally linked across multiple years. In addition to math and reading test scores, information on student characteristics is included: grade level for each year, race/ethnicity, gender, subsidized meal eligibility (a proxy for economic disadvantage), English language learner status and participation in special education programs.

For this analysis, fourteen states and two large urban districts were followed for four years to measure the effectiveness of charter schools on student achievement growth.[3] Information on the years and grade levels included in the analysis can be found in Appendix A. For both the FE and VCR methods, observations were pooled across states. Consolidating experience across states is not without its challenges. As noted by the National Consensus Panel on Charter School Academic Quality (2008), states vary in important ways in their treatment and support of charters. In recognition of this fact, this study sought to examine common elements of charter school performance while simultaneously recognizing that states may play a role in how these schools perform.

Others have suggested that state accountability tests differ, such that scores on a grade-level test in one state may not align with similar scores in another. The study design circumvented these potential difficulties by standardizing test results from each participating state for each grade and year and then calculating a one-year standardized growth score for each student. When scores are thus standardized into z-scores, every student is placed relative to his peers in his own state. A z-score of zero, for example, denotes a student at the 50th percentile in that state, while a z-score one standard deviation above that equates to the 84th percentile. Students who maintain their relative place from year to year would have a growth score of zero. Because both the FE and VCR methods pair control and treatment observations within states, structural problems with testing and scoring differences across states are minimized.

The pooled estimates of overall charter effectiveness, then, will be affected by the sampling fractions included in each state for each method. Table 1 displays, for each location, the total count of individual charter students who had one or more recorded growth scores at a charter school

[3] On a practical note, this study required an approach that met the multiple and conflicting interpretations across states of the Family Education Records Privacy Act (FERPA). The only realistic avenue to conduct a study of this scope was to negotiate agreements with state education agencies for permission to use administrative datasets with student-level records. In spite of changes to the implementation regulations in late 2008, the law remains unclear about the permissibility of providing independent researchers access to student-level data. Several accommodations were imposed as conditions of agreement—though, curiously, each was imposed by only one state. For example, even after all identifying information was removed, one state declined to provide gender on the theory that it prevented identification of individual students. Lacking that information in one state meant that this variable could not be included in the pooled model for any state, so this study was unable to control for gender effects.

Table 1
Unique charter students by location and sampling fraction by comparison criteria[a,b]..

State	Unique Charter Students	% Unique students meeting restricted FE criteria	% FE criteria students with VCR	% of Students by state (unrestricted models)	
				FE	VCR
AR	3184	35.65	90.13	50.16	87.88
AZ	46,372	17.23	85.37	29.91	82.29
CA	265,779	30.71	84.16	44.79	84.70
CO	38,258	21.88	87.74	39.19	90.03
DC	10,955	29.38	87.95	34.58	83.07
FL	84,922	29.38	92.00	50.09	92.64
GA	24,210	25.15	94.27	40.15	93.63
IL (Chicago)	11,842	18.59	95.00	23.93	94.04
LA	10,526	15.74	87.51	24.57	85.50
MA	19,670	22.44	87.34	33.60	80.49
MN	12,077	20.24	83.31	32.23	76.45
MO	8538	20.06	89.14	24.70	84.22
NC	24,001	19.35	81.46	39.52	80.14
NM	6873	30.51	80.73	44.06	75.34
NYC	8453	23.34	92.07	32.00	83.76
NY (upstate)	11,026	14.94	84.48	32.13	84.15
OH	31,836	14.25	80.05	24.55	78.17
TOTAL	618,522	26.02	86.32	40.44	85.53

[a] Throughout this paper, upstate New York and New York City are reported separately. This is done because the student demographics and charter policy are vastly different, which could affect results.

[b] A majority of the states were followed from the 2004–2005 through 2007–2008 school years, but a handful of areas underwent major testing changes in spring 2006 and thus were followed from the 2005–2006 through 2008–2009 school years. These are California; Washington, DC; Chicago, IL; Massachusetts; Minnesota; Missouri; and New York City and state.

in that location. Also shown are the sampling fractions for the restricted models that will form the basis of the head-to-head test of the FE and VCR methods in Section 2 and the sampling fractions for the unrestricted models to be discussed in Section 3.[4]

As noted above, the head-to-head methods test necessitated that charter students were included only when they were TPS-then-charter sector switchers and a virtual control match could be found for them. The proportion of students who met the first criterion is shown in the third column in the table. The proportion of those students who also met the second criterion is displayed in the fourth column. These statistics give some context to the methods test discussion in the following section and also show that finding a VCR match for charter students who have previously attended TPS is slightly easier than finding a VCR match for the average charter student in each state.

For the unrestricted models, the VCR method includes a much greater percentage of charter students in every location, ranging from a low of 75 percent in New Mexico to 94 percent in Chicago, IL. The FE method ranges from a 24 percent inclusion rate in Chicago, IL, to 50 percent in Arkansas. Even within states, FE inclusion rates were found to vary widely (due in part to the age of the charter law, the pace of growth in the sector and the number of consecutive grade levels included in the testing regimen). These findings suggest that the results of the FE analysis are highly temporally and location-sensitive. The findings also indicate that results from the VCR method, containing a higher proportion of all charter students, is the more generalizable of the two methods.

The state-by-state sampling fractions in the unrestricted models tell only part of the story about which students are or are not included in the model for each method, however. The whole population of unique charter students is disaggregated by characteristic below, and the proportions of charter students included in the unrestricted fixed effects and VCR models are compared to all charter students for each characteristic. The purpose of this exercise is to determine whether specific limits on generalizability for each method can be described.

1.2.1. Student traits

First, it is important to consider whether the FE and VCR models adequately represent the mix of ethnicities, income levels and program participants present in the charter student population. As seen in Table 2, the VCR population is most similar to the full charter population in terms of these

Table 2
Proportion of students by trait.

Student trait	All charter students	FE model	VCR model
Race/ethnicity			
Asian	4.28	4.22	3.56
Black	24.63	23.86	24.66
Hispanic	29.64	32.85	30.05
Native Am	1.15	1.16	0.52
White	39.05	36.83	40.46
Multi/other	1.26	1.08	0.76
Low income	47.94	50.66	48.02
English learner	8.61	8.95	7.93
Special education	9.01	8.69	6.90

[4] Percentages and counts in the tables in this section are from the math dataset unless otherwise noted. Student subgroups were generally equivalent across the two test subjects.

student traits, although the FE population is also generally similar. Both models have roughly the same proportion of black students and students who are English learners as the full charter population. In the FE models, though, white students are underrepresented and Hispanic and poor students are overrepresented by a couple percentage points each. Special education students and, to a lesser extent, Native Americans are underrepresented in the VCR data. These results suggest no particular strength or limitation for either method in demographic representativeness of the data.

1.2.2. First observed score

Another measurable difference among students is their early performance on achievement tests. Because these scores encompass a host of unmeasurable student traits, such as earlier schooling and family and neighborhood backgrounds, alignment of the means of the first observed scores of all charter students and the students included in each model is important to note. Table 3 shows how each method divides the All Charter group into two groups: Those whose observations are included in the model and all remaining students.

Table 3 provides numerous insights both within and across the models. Within the FE model, the difference in the means of the students who are and are not included is much smaller than within the VCR model. Further, the students who are not included in the FE model have a higher average than the included group, which also contrasts with the findings for the VCR model groups. However, the difference in the means of the two FE groups is statistically significant in both math and reading. The implication is that the FE models do not accurately represent students from the higher end of the performance distribution.

Within the VCR model, the average first score for students who are not included is much lower in both math and reading than that of the charter students who are included, and the difference between the two groups is statistically significant. The implication here is that the lowest students in the charter school distribution do not have counterparts in the traditional public schools that feed their charter school. The reasons for this were found to be threefold. First, lower-achieving students are more likely to repeat a grade in charter schools than in TPS. These same students are less likely to be labeled as needing special education services in charters than in TPS. Third, there are very few very low performers in each school, regardless of sector, and thus matching is excessively difficult due to demographic heterogeneity. This phenomenon is evident with the highest performers as well; the unmatched group also contains a disproportionate number of the highest achievers, as evidenced by the high variance. An analysis of the match rate by first observed score revealed lower rates at both tails of the performance distribution than in the middle of it. However, the phenomenon affected only the four percent of charter students scoring two or more standard deviations above or below the mean. The VCR method, then, appears to be more representative of students in the middle of the distribution than of the distribution as a whole.

Comparing only the included groups across the models, the average for the VCR group is significantly higher than the FE group in both subjects. Although *t*-tests of each model's included group average compared to the all-charter-student group average yielded p-values that were significant at the 0.01 level in every instance, the average for the FE model group is closer to the all-charter-student average than is the VCR group. This suggests that, in terms of first observed test scores, the FE models are more representative of the charter student population than the VCR models.

1.2.3. First observed year

One of the largest differences in the composition of the two analytic datasets is the distribution of students across years, which is seen in Table 4. As discussed earlier, charter students can be matched in the VCR method as soon as they have two consecutive test scores (one growth period), regardless of the sector in which the first score was attained. Thus, roughly 80–90 percent of students in each year are included, as was seen in Table 1 above. In contrast, students must have three consecutive test scores (two growth periods) to be included in the FE models; further, at least one growth score must be attributed to each sector. These additional requirements result in a dataset that is more heavily weighted toward the earlier years of available data. While it is possible that a longer time horizon would increase the percentages of students included in the FE models, the FE requirements will always fail to include large proportions of recent school entrants.

The inability of the FE design to obtain estimates from recent years has serious implications. Given that available data is almost always a full year out of date, FE models can only provide reasonable estimates at a significant lag, shown here to be three or four years. This represents a significant drawback for policymakers, since the models lack information about the effects of the newest charter schools or recent quality shifts in existing charter schools. Results from the FE method will also fail for several time periods

Table 3
Mean first observed score by method and inclusion.

	Math		Read	
	Mean	Variance	Mean	Variance
FE model				
Included	−0.113	0.94	−0.113	0.94
Not included	−0.093	0.98	−0.050	1.01
VCR model				
Included	−0.056	0.89	−0.023	0.90
Not included	−0.372	1.32	−0.392	1.32
All charter students	−0.101	0.97	−0.076	0.98

Table 4
Annual student count and inclusion rates.

First observed year	*N*	% of students by year	
		FE	VCR
2004–2005	185,150	49.56	89.23
2005–2006	305,048	47.45	85.22
2006–2007	90,564	15.06	81.71
2007–2008	37,734	0.00	78.98

Table 5
Attendance patterns and inclusion rates.

Attendance pattern	N	% of students by attendance pattern	
		FE	VCR
Always charter	265,370	0.00	84.80
TPS then charter	284,611	65.34	86.43
Charter then TPS	68,515	93.66	84.56

to adequately capture the most recent charter policy developments, such as the replication of successful charter schools or the proliferation of charter conversions from failing traditional public schools.

1.2.4. Attendance patterns

Results for charter student performance may also vary depending on students' attendance patterns into and out of the charter and traditional public school sectors. In the full dataset, about 43% of students attend only charter schools. About the same proportion, 46% attend traditional public schools before transferring to the charter sector. The remaining 11% attend charter schools before they transfer to TPS. Table 5 displays the percent of each group included in the VCR and FE subsets.

The VCR method includes about the same proportion of students from each attendance group. This is not the case for FE, however, since only sector switchers are eligible for this method. Interestingly, the two sector switcher groups have vastly different proportions of inclusion in the FE models. A deeper analysis of the students in the "TPS then Charter" group revealed a variety of attendance possibilities under this broad category, including in about one-third of the cases a one-year stint at a TPS, which is insufficient to obtain a TPS growth period. In contrast, the students in the "Charter then TPS" category nearly always obtain a growth score in each sector.

To fully unpack the implications of the differences in inclusion rates for each attendance pattern, it is useful to determine whether test scores differ for the three groups. Table 6 displays the results of the mean first test score for all charter students as well as each attendance pattern. Breaking the full charter group into subsets by their attendance pattern, differences in starting scores become apparent. Students who have charter experience before attending a traditional public school have higher average scores in both math and reading than students who attend TPS before transferring to a charter, and *t*-tests show that these differences are statistically significant at $p < 0.01$. In reading, the "Always Charter" and the "Charter then TPS" groups are also statistically different from each other at $p < 0.01$.

Table 6
Mean first observed score by attendance pattern.

First observed score	All charter students	FE	VCR
Math			
All students	−0.101	−0.113	−0.056
Always charter	−0.060		−0.010
TPS then charter	−0.149	−0.132	−0.104
Charter then TPS	−0.067	−0.061	−0.026
Read			
All students	−0.076	−0.113	−0.023
Always charter	0.004		0.055
TPS then charter	−0.155	−0.138	−0.098
Charter then TPS	−0.037	−0.035	0.011

As discussed above, each of the three attendance patterns is adequately represented in the VCR models, although the VCRs have statistically higher first scores than the full population in each case. The FE models represent two of the three attendance patterns and do so more closely than the VCR subset for each pattern. For FE, only the "TPS then Charter" averages are statistically different than the full average for that pattern. In this case, then, the FE subset more closely resembles the average first score of the full charter student dataset. To the extent that charter students' scores diverge along attendance patterns over time, however, the FE method may not remain representative of the full charter population's starting score.

1.2.5. Summary of data description

Both methods include similar proportions for many of the various student demographics, although each method differs on at least one dimension. The VCR method is less likely to include students with a first observed test score at the very low or very high end of the distribution. The FE method has lower rates of inclusion at each geographic location than the VCR method, which is due in large part to the inability of the FE method to include students who are observed only in the charter sector. A further limitation of the FE method is that new charter schools and new entrants are not likely to be included in sufficient proportions for the results to be representative of the most recent policy changes. As the two approaches have different strengths and weaknesses, findings in areas of agreement between the methods' analytic results should be more robust than those found with only one method.

2. Test of methods

At the heart of the debate of the FE and VCR methods is whether they produce equivalent results. Because the two methods include different charter students, it is necessary to construct a head-to-head comparison using the same charter students, i.e., TPS-then-charter sector switchers who receive a virtual control match. By using these restricted models, the control conditions of the FE and VCR methods can be directly compared.

$$\Delta z^{Ch}_{i,t} - \Delta z^{TPS}_{FEi,(t-1)} \quad \text{or} \quad \Delta z^{Ch}_{i,t} - \Delta z^{TPS}_{FEi,(t+1)} \tag{1}$$

$$\Delta z^{Ch}_{i,t} - \Delta z^{TPS}_{VCR\ i,t} \tag{2}$$

In the FE method, *z*-score growth for charter student *i* in period *t* is compared to a separate growth period in the TPS sector for the same student, as seen in Eq. (1). In a regular FE model, the TPS experience can precede (i.e., period $t-1$) or succeed the charter experience (i.e., period $t+1$). For the VCR method, however, the *z*-score growth for charter student *i* in period *t* is compared to the growth of the student's virtual control record in the same time period, as shown in Eq. (2). In other words, the VCR method uses a

Table 7
Overall and marginal charter effects by method using restricted models.

	Math		Read	
	FE $N=389{,}961$	VCR $N=351{,}616$	FE $N=428{,}018$	VCR $N=381{,}096$
All students	−0.074** (0.003)	−0.063** (0.002)	−0.034** (0.002)	−0.024** (0.002)
Marginal effects				
Race/ethnicity				
Black	−0.035** (0.005)	−0.040** (0.005)	−0.035** (0.004)	−0.035** (0.004)
Hispanic	−0.073** (0.004)	−0.062** (0.004)	−0.026** (0.004)	−0.032** (0.004)
White	−0.174** (0.004)	−0.165** (0.003)	−0.095** (0.003)	−0.075** (0.003)
Low income	0.024** (0.004)	0.051** (0.004)	0.024** (0.003)	0.034** (0.004)
English learner	0.109** (0.006)	0.048** (0.007)	0.072** (0.005)	0.042** (0.006)
Special education	0.058** (0.006)	0.024** (0.008)	0.012* (0.006)	0.017* (0.008)

Robust standard errors in parentheses; N is equal to the total number of observations included in the model.
* $p<0.05$.
** $p<0.01$.

differences-in-differences approach. To ensure time symmetry across the two methods, only charter students with preceding TPS experience were included in the models for this comparison.[5] Thus, the historical TPS experience of the charter student in the FE method will be compared with the contemporaneous TPS experience of the VCR.

For both methods, controls for the previous test score, year, grade level and an indicator if the growth score was measuring a repeated grade were included in the regression models. The use of each student as his own control obviates the need for additional demographic variables in the FE models. The VCR models include controls for time period, student demographics and state indicators. For the marginal effects model in each method, variables were added to isolate the charter effect for each subgroup. The results of the regression models are shown in Table 7. Although the estimates are not useful to a general discussion about charter school effectiveness, the FE estimates can be compared to the VCR estimates to evaluate the two methods.

Overall and for every subgroup, the FE and VCR models return estimates that are the same sign and have the same level of significance. Many of the coefficients across the two models are quite similar, such as those for black students. Further, neither method produces consistently larger or smaller coefficients. Since the charter students are exactly the same in the two methods, the differences seen in these coefficients are due to the differences in the control conditions: the charter students' historical TPS experience in the FE method versus the current TPS experience of the controls in the VCR method. These differences are particularly evident in the English Learner and Special Education coefficients. The smaller coefficients in the VCR models relative to the FE models seems to indicate that the traditional public school sector is experiencing gains with these two student groups over time, which are not captured by the FE models.

[5] About 8% of students whose first growth period is observed in TPS switch back to TPS after having growth in charter. In these cases, the average of the two TPS growth periods are used as the control condition in the FE calculations.

These results indicate that the two methods produce reasonably similar estimations of charter school effectiveness when restricted to the same charter students. Based on these results, we can infer that differences in the estimates in the following section, which use unrestricted models, will be due to differences in the charter student populations included in each method.

3. Policy analysis

The purpose of this section is to determine the policy insights that can be gained by comparing the differences in the estimations from the FE and VCR models. The models in this section are unrestricted; the FE models include all sector switchers regardless of whether they have a matching virtual control and the VCR models includes all charter students with a matching virtual control regardless of whether they were sector switchers. Both types of models take the general form

$$\Delta A_{i,t} = \theta A_{i,t-1} + \beta X_i + \gamma C_{i,t} + \varepsilon_{i,t} \quad (3)$$

where the dependent variable, growth in student achievement, is

$$\Delta A_{i,t} = A_{i,t} - A_{i,t-1} \quad (4)$$

when $A_{i,t}$ is the z-score for student i in period t; $A_{i,t-1}$ is the z-score for student i in period $t-1$; X is a set of control variables for student characteristics, states and years; C is an indicator variable for whether the student attended a charter in period t; and ε is the error term. Lagged achievement ($A_{i,t-1}$) is included as a control variable because growth is not necessarily equivalent across the spectrum of prior achievement levels, due in part to the phenomenon of regression to the mean (Ladd & Walsh, 2002). As with the prior model, the FE models include only time-variant control variables, which are a subset of those included in the VCR models. For both FE and VCR models, ordinary least squares regressions were run for pooled models and for individual locations.

Results from the pooled models are displayed in Table 8. Both the FE and VCR models show that charter students, on average, have statistically lower growth than their TPS

Table 8
Overall and marginal charter effects by method using unrestricted models.

	Math		Read	
	FE $N = 19{,}905{,}212$	VCR $N = 1{,}692{,}236$	FE $N = 19{,}780{,}459$	VCR $N = 1{,}765{,}464$
All students	−0.065** (0.001)	−0.016** (0.001)	−0.049** (0.001)	0.007** (0.001)
Marginal effects				
Race/ethnicity				
Black	−0.025** (0.003)	0.008** (0.002)	−0.037** (0.003)	0.007** (0.002)
Hispanic	−0.057** (0.003)	−0.020** (0.002)	−0.040** (0.003)	−0.001 (0.002)
White	−0.131** (0.002)	−0.062** (0.001)	−0.095** (0.002)	−0.013** (0.001)
Low income	0.016** (0.00234)	0.029** (0.002)	0.017** (0.002)	0.019** (0.002)
English learner	0.049** (0.00404)	0.034** (0.003)	0.041** (0.004)	0.028** (0.003)
Special education	0.013** (0.00379)	0.019** (0.004)	0.007 (0.004)	0.006 (0.004)

Robust standard errors in parentheses; N is equal to the total number of observations included in the model.
** $p < 0.01$.

counterparts in math. The result for reading is split, however; the FE estimate is negative and significant, while the VCR estimate is positive and significant, though small. This finding is indicative of a larger pattern between the two methods: Overall and for virtually all the subgroups in both math and reading, estimates from the VCR method are more positive than those from FE. This suggests that the students who are excluded from the FE method have higher growth than those who are included.

Both methods return positive results for charter schools for three groups: low income students, English learners, and special education students. In addition, the VCR estimates for black students in both subjects and all students in reading are small but positive and significant at $p < 0.01$. The difference between the FE and VCR coefficients in these instances may indicate a positive trajectory for these measures, since the VCR models include the experience of more recent charter school entrants.

The estimates for English learner students do not fit the pattern of VCR estimates being more positive than the estimates from FE. The TPS controls in the VCR method are contemporaneous to the charter growth observations, a phenomenon that allows the performance of the traditional public school sector to adjust over time. It is possible, then, that the lower estimates from the VCR method reflect a positive change in performance for English learners in charters at the same time that English learners in traditional public schools are also posting positive gains.

3.1.1. Variation in results by location

The real insight from each of the models is that the effectiveness of charter schooling is widely varied. To illustrate variation across jurisdictions, the unrestricted models were re-estimated by location. Individual location results are displayed in Table 9, while the observations for each

Table 9
Charter effects by location.

	Fixed Effects		VCR	
State	Math	Reading	Math	Reading
AR	0.051** (0.012)	0.018 (0.012)	0.049** (0.011)	0.017 (0.010)
AZ	−0.102** (0.004)	−0.055** (0.005)	−0.046** (0.003)	−0.013** (0.003)
CA	−0.047** (0.002)	−0.018** (0.001)	−0.026** (0.001)	0.018** (0.001)
CO	−0.077** (0.003)	−0.061** (0.004)	−0.028** (0.002)	−0.020** (0.002)
DC	0.043** (0.011)	−0.008 (0.012)	0.105** (0.007)	0.067** (0.006)
FL	−0.038** (0.002)	−0.036** (0.003)	−0.021** (0.002)	−0.021** (0.002)
GA	−0.088** (0.006)	−0.058** (0.006)	−0.017** (0.004)	0.005 (0.004)
IL (Chicago)	−0.062** (0.008)	−0.073** (0.009)	0.012** (0.004)	0.019** (0.005)
LA	0.060** (0.011)	0.089** (0.011)	0.043** (0.007)	0.055** (0.007)
MA	0.017** (0.006)	−0.021** (0.007)	0.064** (0.004)	0.027** (0.004)
MN	−0.087** (0.008)	−0.034** (0.008)	−0.028** (0.006)	−0.010 (0.005)
MO	−0.198** (0.015)	−0.145** (0.014)	0.000 (0.007)	−0.004 (0.007)
NC	−0.133** (0.005)	−0.049** (0.005)	−0.042** (0.004)	0.018** (0.004)
NM	−0.091** (0.009)	−0.044** (0.010)	−0.047** (0.007)	−0.024** (0.008)
NYC	0.072** (0.009)	−0.044** (0.010)	0.113** (0.005)	0.049** (0.005)
NY (Upstate)	0.235** (0.011)	0.019 (0.014)	0.149** (0.007)	0.025** (0.006)
OH	−0.156** (0.006)	−0.085** (0.006)	−0.058** (0.004)	−0.005 (0.004)

Robust standard errors in parentheses.
** $p < 0.01$.

Table 10
Summary of marginal effects for individual charter schools.

Subject	Number of charter schools	Positive and significant	Not significant	Negative and significant
Math	2242	19.1%	47.8%	33.1%
Reading	2234	19.1%	59.7%	21.2%

$p < 0.05$.

method at each location can be found in Appendix A. The estimates between the two methods are more similar for math than for reading in terms of sign and significance. Six of the 17 locations post positive and significant estimates in the FE method in math; all of those locations, in addition to Chicago, have positive and significant VCR estimates. In contrast, only one state has positive and significant estimates for both methods in reading. For this subject, the VCR estimates indicate that charter students in eight locations have statistically higher growth than their TPS peers. More generally for both subjects, where there is agreement between the methods on sign and significance, the estimates from the VCR method paint a more favorable picture of charter school student performance. The differences reflect the inclusion in VCR estimates of students who have only attended charter schools and those who have only a brief period in TPS.

3.1.2. Individual schools

State or city differences do not account for all the variation in performance. Some of the variation occurs at the local market level. To isolate these effects, we define highly localized markets, consisting of the students in each charter school and the public schools from which it draws students. This highly localized analysis can provide the answer to the question: do charter schools create student outcomes that are on par with their local competition?[6] Because the appropriate sampling fraction for each charter's competitors is unknowable in the FE models, that method is unable to support a school-by-school quality estimator.

The analysis employed market fixed effects to remove all the effects that are shared by a charter school and its traditional public school market, as represented by the charter's VCRs. This model takes the functional form

$$\Delta A_{i,t} = \theta A_{i,t-1} + \beta X_i + \gamma^c C^c_{i,t} + \rho^m M^m_{i,t} + \varepsilon_{i,t} \quad (5)$$

where γ^c is the marginal effect of charter c given the average performance of market m. These individual charter school coefficients convey the marginal performance of each school relative to its market. A summary of these coefficients is displayed in Table 10.

In both math and reading, about one-fifth of the charter schools have significantly higher average growth than their TPS market counterparts. The preponderance of charter schools in each subject are on par with the performance in their local market, however high or low that market performance may be.

[6] Note that this analysis is purely relative and gives no information about the overall level of each market's performance or whether the existing market performance is acceptable.

3.1.3. Summary of unrestricted models

The results of the models, when taken together, indicate that the story of charter school performance is one of uneven quality on many dimensions. The pooled result for all students is negative in math, but there is cause for cautious optimism in reading. Positive and significant results are found for multiple subgroups of charter students have while negative and significant effects are found for other subgroups. This variation is also evident when results are disaggregated by location. Charter school students outperform their TPS counterparts in one-third to one-half of the geographic locations, depending on the subject tested and method used. (The picture is much more positive for charters in reading using the VCR method.) The proportion of positive results for individual schools in math and reading is smaller still.

4. Conclusions

The present study was aimed at two purposes: to construct a rigorous test of the attributes of student fixed effects models against those employing virtual control records, and to provide a comparison of the estimates of charter school effectiveness when employed in a policy analysis. While one is hesitant to draw generalizations from a single comparison of the two techniques, the insights that the analysis provides can help guide researchers and decision makers alike in using studies featuring each approach.

For charter school research, the student fixed effects methodology is the predominant quasi-experimental technique for minimizing selection bias in regression estimates. By using a student as his/her own control, albeit temporally displaced, the fixed effect approach ensures that all time-insensitive observable and unobservable characteristics are equivalent in treatment and control conditions. The possibility of selection bias into one or the other group is thus minimized.

Alternatively, the VCR method uses all available observable characteristics and prior performance – which captures motivation and family supports – to create a composite comparison record. An important attribute of the VCR method is that the period of comparison between charter students and their controls is contemporaneous; the control condition occurs in parallel with the year of charter school study. This feature could be important when studying jurisdictions that are undergoing significant and rapid change.

The head-to-head comparison of the FE and VCR methods used the same charter students to create a computational test. Since the charter students were the same in both methods, the comparison essentially tested for

Table 11
Best method by research objective.

Research bjective	FE	VCR
Provide best estimate of students with the lowest starting scores	X	
Provide best estimate of students with the highest starting scores	X	
Provide best estimate of students with a special education designation	X	
Provide estimate of "Always Charter" students		X
Include largest proportion of charter students possible		X
Provide best estimate when sector switchers are substantially different from non-switchers		X
Provide best estimate when sector switchers are representative of overall charter population	X	
Provide estimate of head-to-head charter effect versus their feeder market		X
Provide timely estimates of changes in charter school policy		X
Provide timely estimates of new charter school effectiveness		X

equivalence between the FE control (e.g., the charter student's own TPS experience) and the VCR. Under these conditions, the two methods produced highly similar results. Coefficients for the charter effect were identical in sign and significance and of the same general magnitudes. Further, the differences in size of coefficient did not consistently favor either method. Therefore, it can be concluded that the influence of selection bias in the VCR method is about the same as in the FE method.

In addition to selection bias, a chief concern about the usefulness of an analytic method is the generalizability of the results produced by it. Each of the two methods discussed here has some limitations in this area, as does any analytic method that uses samples of a population. The fixed effects methodology appears to have more limits on generalizability than the VCR methodology. The students covered by the FE method comprise about 40 percent of the charter student population. FE is limited to sector switchers, which means its estimates can provide no insight into students who are educated solely in the charter sector. The proportion of students included in FE models is shown to be influenced by the maturity and growth rates of the charter sectors, as these factors affect the flow of students from one sector to the other. Further, the results of many of the newest entrants in the charter school market are not included in the FE estimates. In contrast, the VCR method includes about 85 percent of charter students. Students who first score in the extreme tails of the achievement distribution are difficult to match and are less likely to be included in the VCR method than those in the middle of the distribution. However, the match rates on the low end still exceed 60 percent, and the high end still has a match rate of 75 percent. On the low end, less than three percent of the students are affected; on the high end, about one percent of the students are affected. On balance, the generalizability of the VCR method is more favorable than occurs with FE methods.

This paper has shown that the strength of the FE method lies in selection bias minimization. The VCR method is shown to be equivalent with respect to selection bias and provide results that are more generalizable. However, no single analytic method can provide satisfactory answers to every research question. Therefore, the layering of evidence from multiple methods is always preferred. When that is not possible, it is important for researchers to select the most appropriate method for their research objectives and to be explicit about the strengths and limitations of the chosen technique. FE appears to be the preferred method for nontraditional students and when sector switchers are a sizable and representative proportion of the overall charter student population, while the VCR method seems more appropriate for students who attend only charter schools, charter-versus-feeder analyses and estimates of more recent charter school phenomena. Table 11 describes some possible research objectives and notes whether fixed effects or virtual controls is likely best suited to accomplish each based on the findings in this paper. This table, while useful for researchers, can also be used by policymakers to help assess the research results presented to them, which in turn will help pave the way for more-informed policy decisions.

In the area of charter school policy, more information to policymakers cannot come too soon. Although the focus of this research was to compare two quasi-experimental methods that have been used to measure charter school effectiveness, the results also serve to reinforce what has become a familiar refrain in recent charter school research: Charter school quality is demographically and geographically uneven. For the charter school movement to survive and thrive, more needs to be done to improve and tighten the quality distribution. Because it is also clear that state charter school policies influence quality, research looking at the policy similarities and differences between high- and low-quality locations could help determine the drivers of high quality and point to specific policy improvements.

Acknowledgements

The authors are grateful for helpful comments from Eric Hanushek, Susanna Loeb, Sean Reardon, Edward Cremata and anonymous reviewers as well as support from colleagues at CREDO. Financial support from the Packard Humanities Institute is also gratefully acknowledged.

Appendix A.

Table A1 displays the number of unique charter students for each location as well as the observations by method. Although the fixed effects method appears to include many more observations than the virtual control method, this is not in fact the case. In the fixed effects models, all but the sector switchers drop out of the estimation of the charter effect, thereby bringing the effective observation count much closer to that of the VCR method.

Table A1
Observations by location and method[a]..

State	School years included	Grades included	Unique charter students	FE	VCR
AR	2004–2005 to 2007–2008	3–8	3184	479,669	8124
AZ	2004–2005 to 2007–2008	3–12	46,372	989,716	116,494
CA	2005–2006 to 2008–2009	2–11	265,779	1,054,833	722,520
CO	2004–2005 to 2007–2008	3–10	38,258	1,100,752	115,498
DC	2005–2006 to 2008–2009	3–8	10,955	60,693	26,670
FL	2004–2005 to 2007–2008	3–12	84,922	3,867,055	251,274
GA	2004–2005 to 2007–2008	1–8	24,210	2,370,835	74,022
IL (Chicago)	2005–2006 to 2008–2009	3–8	11,842	383,007	36,798
LA	2004–2005 to 2007–2008	3–9	10,526	807,170	26,938
MA	2005–2006 to 2008–2009	3–8	19,670	1,033,531	54,396
MN	2005–2006 to 2008–2009	3–8	12,077	798,255	28,028
MO	2005–2006 to 2008–2009	3–8	8538	60,583	22,308
NC	2004–2005 to 2007–2008	3–12	24,001	1,992,393	65,678
NM	2004–2005 to 2007–2008	3–9	6873	365,767	14,918
NYC	2005–2006 to 2008–2009	3–8	8453	977,980	32,162
NY (upstate)	2005–2006 to 2008–2009	3–8	11,026	1,950,978	24,036
OH	2004–2005 to 2007–2008	3–8	31,836	1,612,971	72,372
TOTAL			618,522	19,906,188	1,692,236

[a] Due to the inability to acquire the full set of California data, the fixed effects models for this state included all observations for sector switchers, all observations of students educated solely in charters and VCR observations standing in for the missing observations of students educated solely in TPS.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2007). *Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control progrom* (NBER Working Paper No. T0335). Retrieved 27.06.11, from http://www.nber.org/papers/t0335.

Abdulkadiroglu, A., Angrist, J., Cohodes, S., Dynarski, S., Fullerton, J., Kane, T., et al. (2009). *Informing The Debate: Comparing Boston's Charter, Pilot And Traditional Schools*. Boston, MA: The Boston Foundation. Retrieved 27.06.11, from https://folio.iupui.edu/bitstream/handle/10244/726/InformingTheDebate_Final.pdf?sequence=2

Ballou, D., Teasley, B., & Zeidner, T. (2008). Charter schools in Idaho. In M. Berends, M. G. Springer, & H. J. Walberg (Eds.), *Charter school outcomes* (pp. 221–241). New York: Taylor & Francis Group, LLC.

Gleason, P., Clark, M., Tuttle, M.C., & Dwoyer, E. (2010). *The evaluation of Charter School Impacts: Final Report* (NCEE 2010-4029). Washington, DC: Institute for Education Sciences.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., & Branch, G. F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics, 91*, 823–848.

Hoxby, C., & Murarka, S. (2007, June). *New York City's Charter Schools Overall Report*. Cambridge, MA: New York City Charter Schools Evaluation Project. Retrieved 27.06.11, from http://www.nber.org/~schools/charterschoolseval/nyc_charter_schools_report_july2007.pdf

Hoxby, C., & Murarka, S. (2008). Methods of assessing achievement of students in charter schools. In M. Berends, M. G. Springer, & H. J. Walberg (Eds.), *Charter School Outcomes*. (pp. 7–37). New York: Taylor & Francis Group, LLC.

Hoxby, C. & Murarka, S. (2009). *Charter schools in New York City: Who enrolls and how they affect their students' enrollment* (NBER Working Paper No. 14852). Retrieved 27.06.11, from http://www.nber.org/papers/w14852.

Ladd, H. F., & Walsh, R. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education Review, 21*, 1–17.

National Consensus Panel on Charter School Academic Quality (2008). *A framework for academic quality: A report from the National Consensus Panel on Charter School Academic Quality*. Retrieved 27.06.11, from http://www.charterschoolquality.org/media/1186/FrameworkForAcademicQuality.pdf.

Peterson, P. E. (2003). *The future of school choice*. Stanford, CA: Hoover Institution Press.

Tuttle, C. C., Gleason, P., & Clark, M. (2012). Using lotteries to evaluate schools of choice: Evidence from a National Study of Charter Schools. *Economics of Education Review, 31*, 237–253.

Zimmer, R., & Buddin, R. (2008). Charter schools in California. In M. Berends, M. G. Springer, & H. J. Walberg (Eds.), *Charter School Outcomes* (pp. 163–193). New York: Taylor & Francis Group, LLC.

Zimmer, R., Gill, B., Booker, K., Lavertu, S., & Witte, J. (2012). Examining charter student achievement effects across seven states. *Economics of Education Review, 31*, 213–224.