Accountability in Education around the World:

Lessons from International Achievement Tests

Ludger Woessmann

University of Munich and Ifo Institute

Conference on

Improving Education through Accountability and Evaluation: Lessons from Around the World

Rome, October 3, 2012

The International Perspective

- "If custom and law define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible."
  - Arthur W. Foshay (1962) on the first pilot study of international student achievement
- Emerging economic literature over past decade:

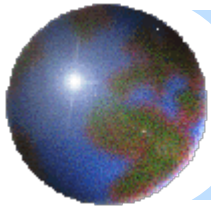
→ Use international tests of educational achievement to analyze determinants and impacts of cognitive skills

- See: E.A. Hanushek, L. Woessmann (2011), 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, Amsterdam: North Holland. (available as NBER Working Paper 15949, Cambridge, MA: National Bureau for Economic Research)

International Tests of Educational Achievement: IEA and OECD Student Achievement Tests

	Abbr.	Study	Year	Region	Subject	Age[a,b]	Countries[c]	Organiz.[d]	Scale[e]
1	FIMS	First International Mathematics Study	1964	World	Math	13,FS	11	IEA	PC
2	FISS	First International Science Study	1970-71	World	Science	10,14,FS	14,16,16	IEA	PC
3	FIRS	First International Reading Study	1970-72	World	Reading	13	12	IEA	PC
4	SIMS	Second International Mathematics Study	1980-82	World	Math	13,FS	17,12	IEA	PC
5	SISS	Second International Science Study	1983-84	World	Science	10,13,FS	15,17,13	IEA	PC
6	SIRS	Second International Reading Study	1990-91	World	Reading	9,13	26,30	IEA	IRT
7	TIMSS	Third International Mathematics and Science Study	1994-95	World	Math/Science	9(3+4), 13(7+8),FS	25,39,21	IEA	IRT
8	TIMSS-Repeat	TIMSS-Repeat	1999	World	Math/Science	13(8)	38	IEA	IRT
9	PISA 2000/02	Programme for International Student Assessment	2000+02	OECD+W.	Math/Scie./Read.	15	31+10	OECD	IRT
10	PIRLS	Progress in International Reading Literacy Study	2001	World	Reading	9(4)	34	IEA	IRT
11	TIMSS 2003	Trends in Internat. Mathematics and Science Study	2003	World	Math/Science	9(4),13(8)	24,45	IEA	IRT
12	PISA 2003	Programme for International Student Assessment	2003	World	Math/Scie./Read.	15	40	OECD	IRT
13	PIRLS 2006	Progress in International Reading Literacy Study	2006	World	Reading	>9.5(4)	39	IEA	IRT
14	PISA 2006	Programme for International Student Assessment	2006	World	Math/Scie./Read.	15	57	OECD	IRT
15	TIMSS 2007	Trends in Internat. Mathematics and Science Study	2007	World	Math/Science	>9.5(4), > 13.5(8)	35,48	IEA	IRT

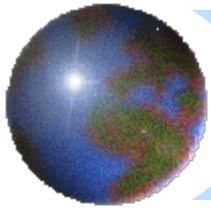
Notes:
a. Grade in parentheses where grade level was target population.
b. FS = final year of secondary education (differs across countries).
c. Number of participating countries that yielded internationally comparable performance data.
d. Conducting organization: International Association for the Evaluation 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).



Advantages of the International Perspective

- Cross-country comparative approach provides some unique advantages over national studies:

1. Use institutional variation not existent within countries
2. Draw on much larger variation than usually available within any country
3. Reveal whether result is country-specific or general
4. Test whether effects are systematically heterogeneous in different settings
5. Circumvent selection issues of within-country identification by using system-level aggregated measures
6. Uncover general-equilibrium effects that often elude single-country studies

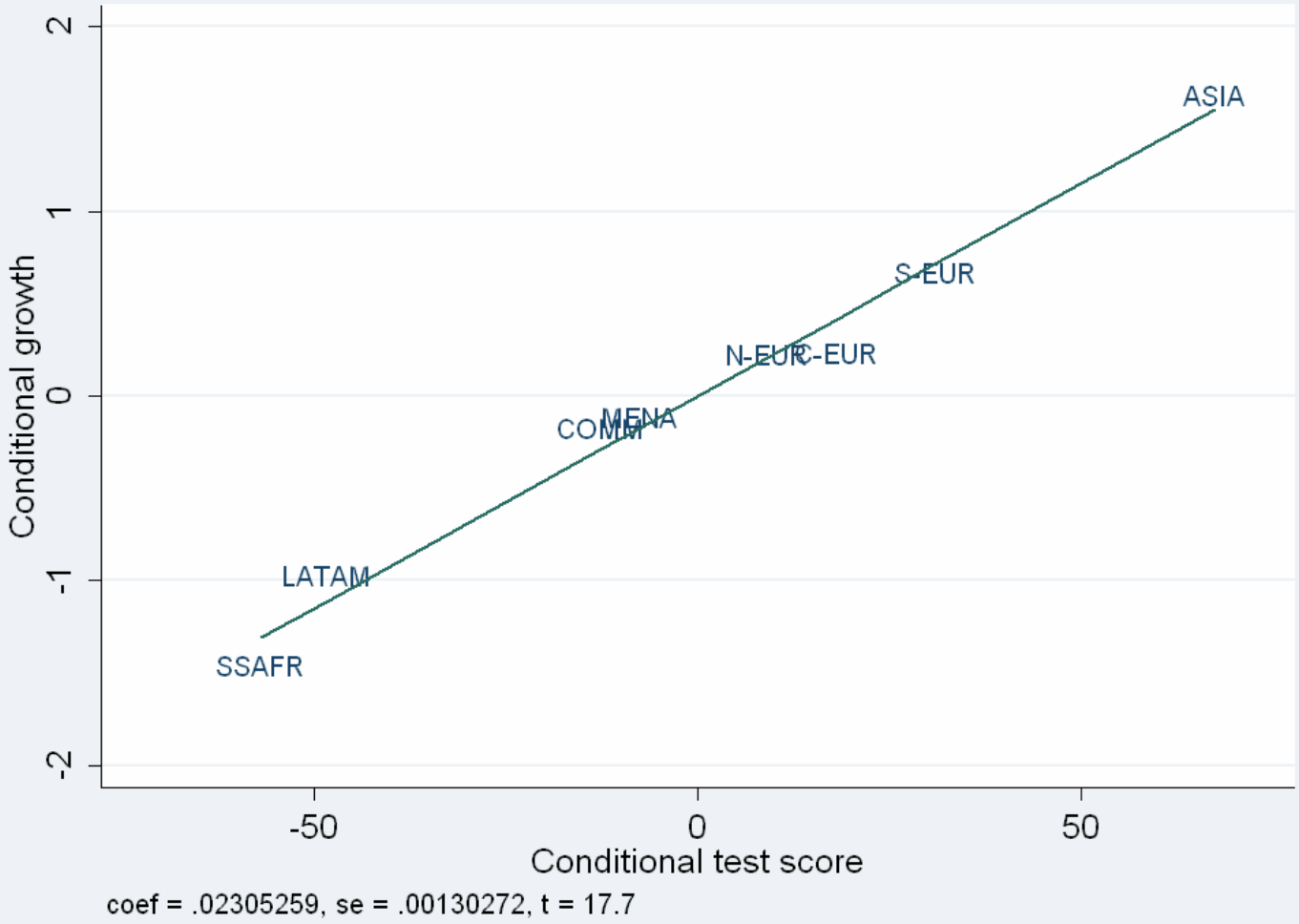


Limitations of the International Perspective

- Advantages come at the price of concerns about:

1. Limited number of country observations
2. Cross-sectional character of most available achievement data
3. Possible bias from unobserved country factors like culture
4. Broad patterns rather than details of specific implementation issues

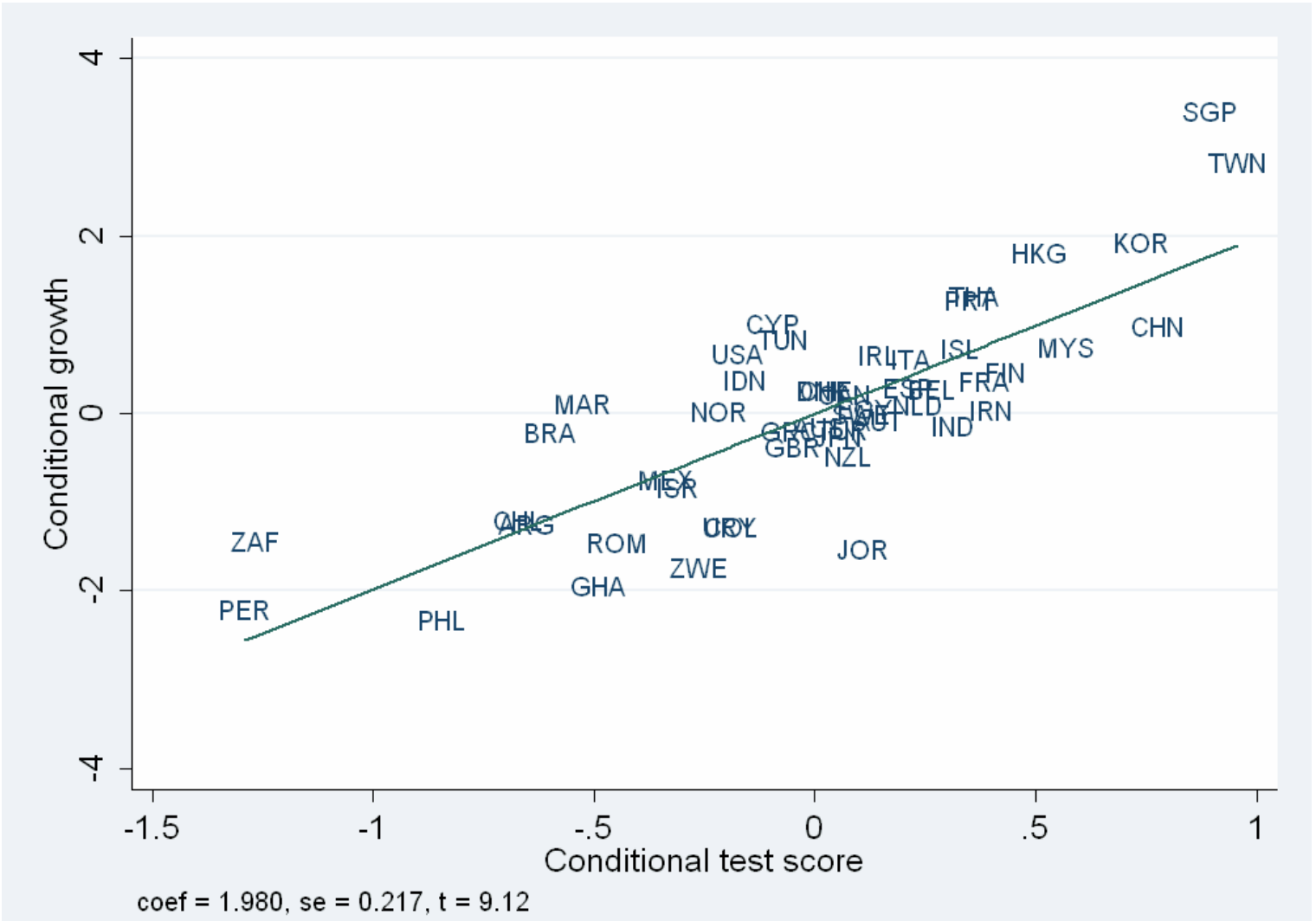
Motivation I:
Test Scores and Economic Growth



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.

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).

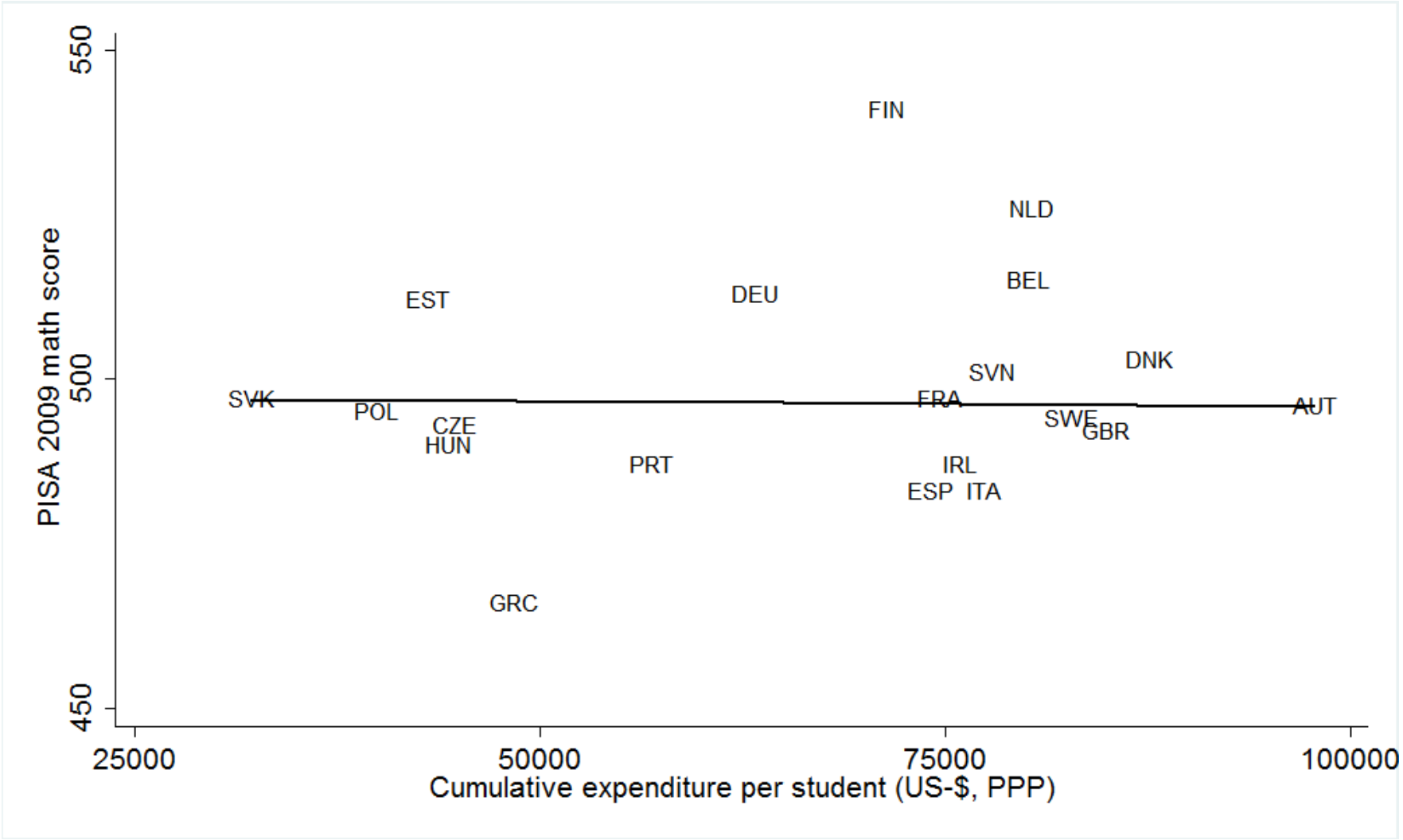
Cognitive Skills and Economic Growth



Added-variable plots 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, average test scores on international student achievement tests, and average years of schooling in 1960.

Motivation II: Spending and Test Scores

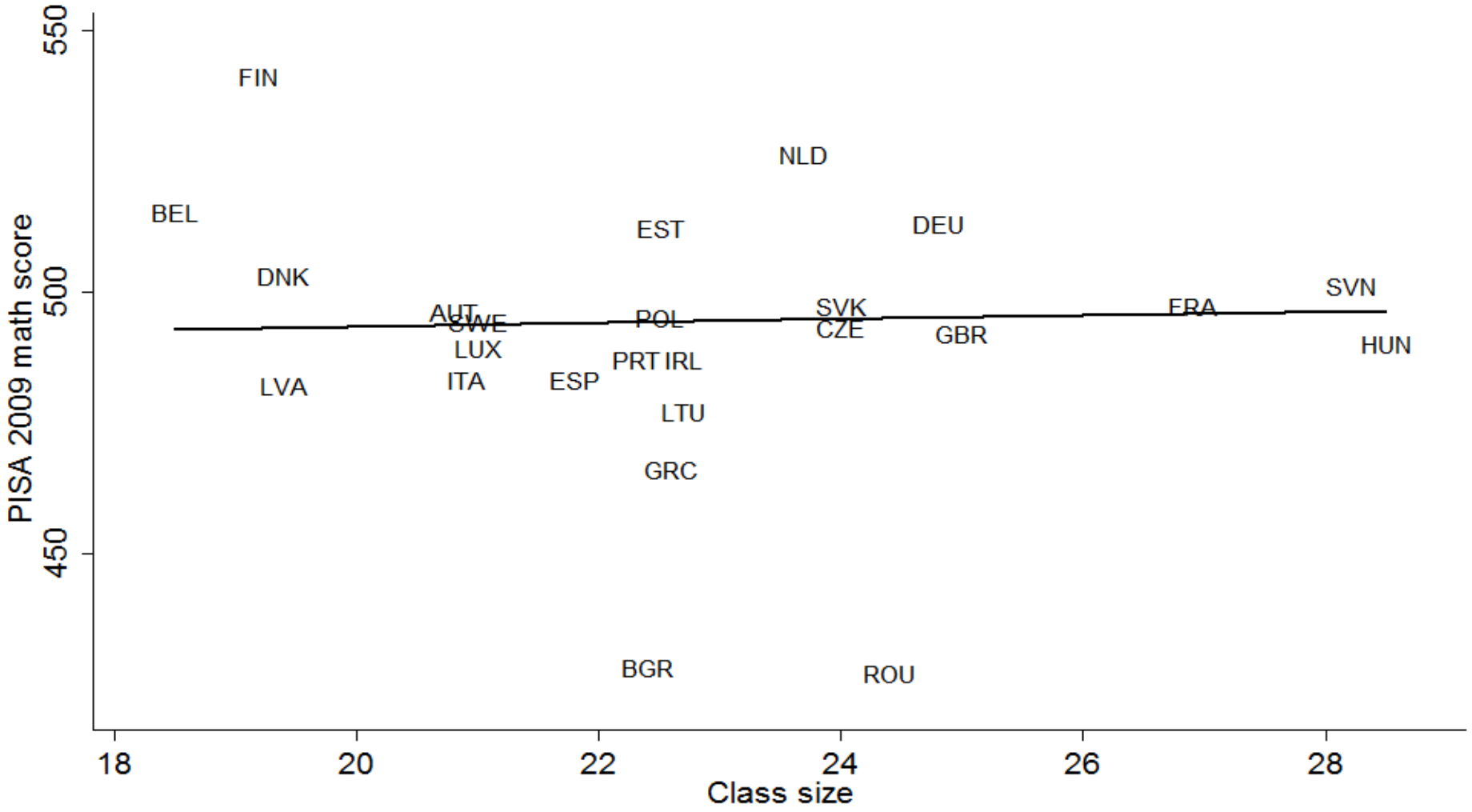
Spending and math achievement of EU countries in PISA 2009:



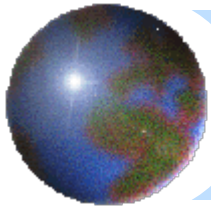
Own depiction based on PISA 2009 data. Regression line of best fit (without three outliers).

Class Size and Test Scores

Class size and math achievement of EU countries in PISA 2009:



Own depiction based on PISA 2009 data. Regression line of best fit (without three outliers).



"Curriculum-Based External Exit Examination Systems" (Bishop 1997)

1. Produce signals of student achievement that have real consequences for the student.
2. Define achievement relative to an external standard, not relative to other students in the classroom or the school.
3. Are organized by discipline and keyed to the content of specific course sequences.
4. Signal multiple levels of achievement in the subject, not only a pass-fail signal.
5. Cover almost all secondary school students.
6. Assess a major portion of what students studying a subject are expected to know.

Central Exams and Student Achievement

- Channels of influence on student achievement:
  1. Increased external rewards for learning
     - Change students' incentive structure relative to local exams
     - Improved signaling of achievement to potential employers
  2. Decreased peer pressure against learning
  3. Change student-teacher relationship
     - Teacher from "judge" to "coach"
  4. Enhanced monitoring of teachers and schools
  5. Country-wide testing creates incentives for policymakers
     - See Germany: laggard states have caught up half of the PISA lag between 2000 and 2006

Central Exams and Student Performance
– Existing Evidence –

- Accumulating evidence:
  Central exit exams strongly positively associated with students' academic performance (cf. Bishop HbEcEdu 2006)
- Cross-country studies:
  - 1991 IAEP math, science, geography (Bishop IJER 1995, AER 1997)
  - 1991 IEA reading (Bishop SEPR 1999)
  - 1995 TIMSS math, science (Bishop 1997; Woessmann OBES 2003)
  - 1999 TIMSS-Repeat math, science (Woessmann PW 2003, EduE 2005)
  - 2000 PISA reading, math, science (Fuchs/Woessmann EmpE 2007)
- Cross-regional studies:
  - Canadian provinces (Bishop 1995, 1997, 1999)
  - U.S. states (Bishop 1995; Bishop et al. EEduR 2000, BPEP 2001)
  - German states (Jürges et al. JEEA 2005, FA 2005, JPopE 2010)

Central Exams and Student Performance
– Approach and Empirical Model –

- Existence of central exit exams mostly national feature
  - ⇨ **International** data
- Background controls and heterogeneous effect
  - ⇨ **Micro** data (incl. student and school characteristics)

⇨ Evidence from extensive **international student-level data**

- TIMSS 1995, TIMSS-Repeat 1999, PISA 2000, PISA 2003

- Cross-country student-level multiple regressions:

$$T_{ilsc} = \alpha E_c + B_{ilsc}\beta + R_{lsc}\gamma + I_{lsc}\delta + a + \varepsilon_{csli}$$

- The interaction specification:

$$T_{ilsc} = \alpha E_c + (E_c I_{isc})\lambda + B_{ilsc}\beta + R_{lsc}\gamma + I_{lsc}\delta + a + \varepsilon_{csli}$$

The Underlying Studies

- TIMSS 1995:
  - Woessmann (2001): Why Students in Some Countries Do Better: International Evidence on the Importance of Education Policy. ***Education Next*** 1 (2): 67-74
  - Woessmann (2003): Schooling Resources, Educational Institutions, and Student Performance: The International Evidence. ***Oxford Bulletin of Economics and Statistics*** 65 (2): 117-170
- TIMSS 1999:
  - Woessmann (2003): 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, DC: Brookings Institution Press
  - Woessmann (2005): The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA. ***Education Economics*** 13 (2): 143-169
- PISA 2000:
  - Fuchs/Woessmann (2007): What Accounts for International Differences in Student Performance? A Re-examination using PISA Data. ***Empirical Economics*** 32 (2-3): 433-464
  - Woessmann (2007): International Evidence on School Competition, Autonomy and Accountability: A Review. ***Peabody Journal of Education*** 82 (2-3): 473-497
- PISA 2003:
  - Woessmann/Luedemann/Schuetz/West (2009): *School Accountability, Autonomy and Choice around the World*. Cheltenham: **Edward Elgar**.
  - Woessmann (2010): Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries. ***Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics*** 230 (2): 234-270

The International Databases

- Third International Mathematics and Science Study (TIMSS)
  - TIMSS 1995 + Repeat 1999 pooled: 447,089 students, 54 countries
- Programme for International Student Assessment (PISA)
  - 2000: 175,227 students, 32 countries
  - 2003: 265,878 students, 37 countries
- Representative samples of students in each country
- Individual student-level data
- Extensive background information: 40-60 variables
  - Family background (student questionnaire)
  - Resource endowment of class/school (teacher/school questionnaire)
  - Institutional features of school system (school questionnaire)

Central Exams and Student Performance
— *Basic Results: TIMSS* —

	TIMSS-95	*TIMSS-Repeat*	*Pooled*
Math	40.9 *(13.5)*	47.0 *(13.5)*	42.7 *(9.8)*
Science	39.7 *(9.9)*	35.9 *(12.9)*	35.9 *(8.3)*

- Robust to inclusion of continental fixed effects

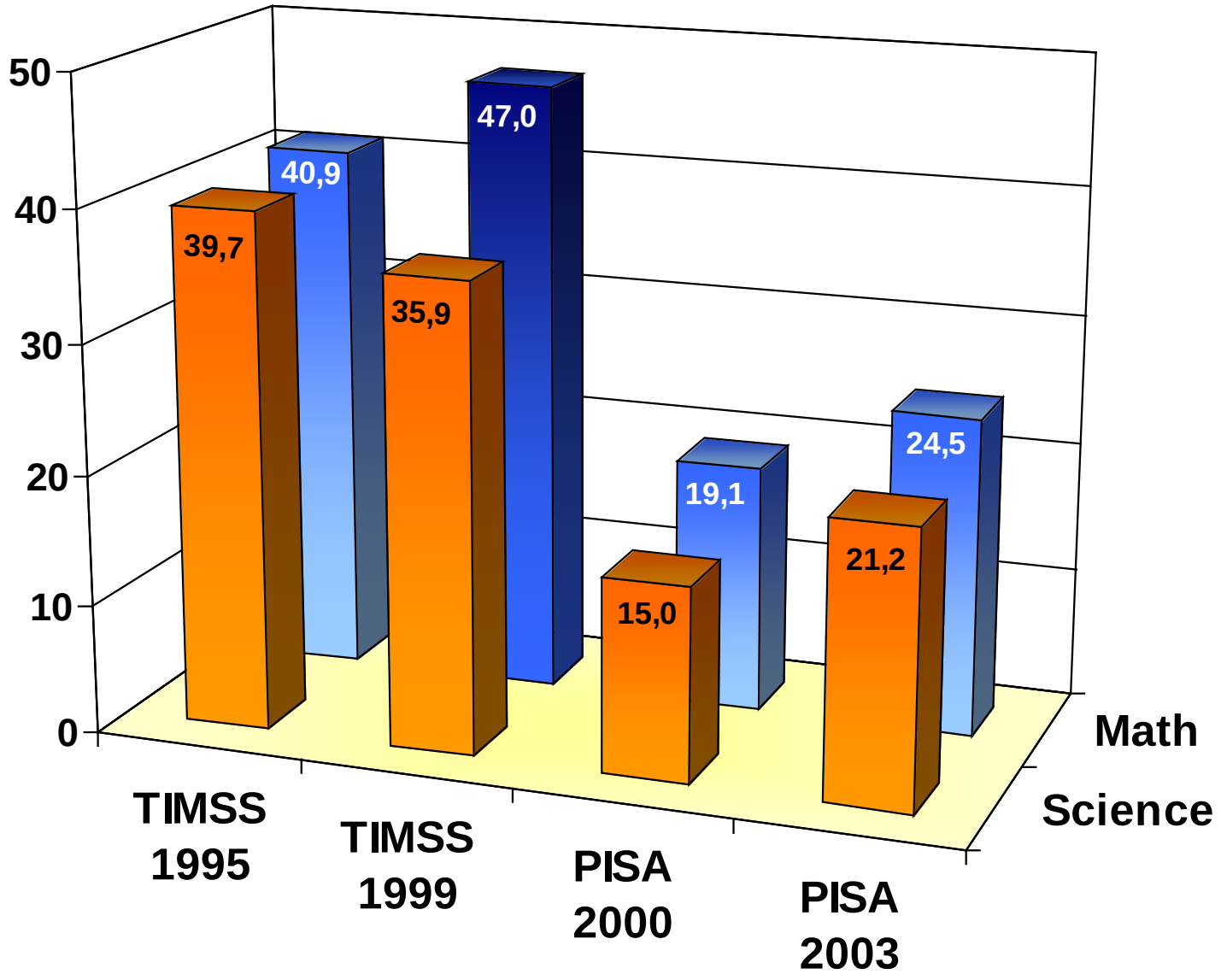
Central Exams and Student Performance

— *Basic Results: PISA* —

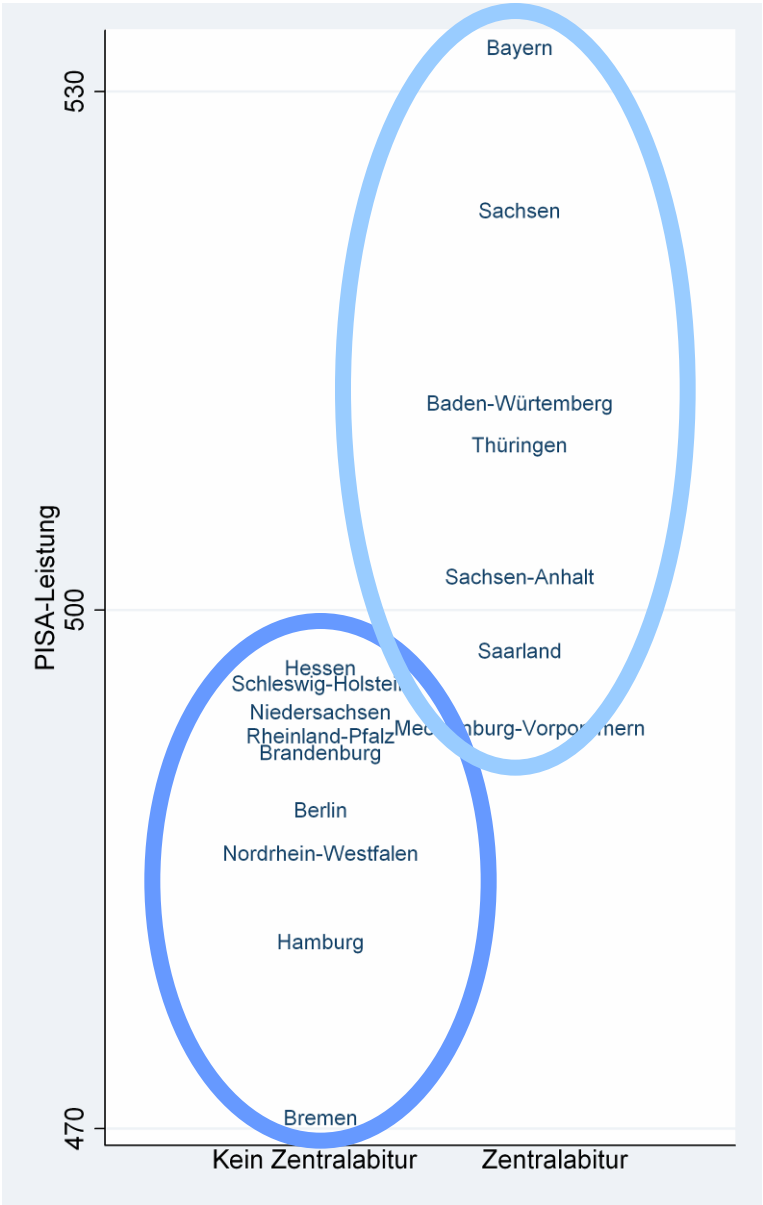
	PISA 2000	*PISA 2003*
Math	19.1 *(9.6)*	24.5 *(10.1)*
Science	15.0 *(9.1)*	21.2 *(7.6)*

Central Exams and Student Performance

— The International Evidence —



Central Exams and Student Performance across German States



Central Exams and Student Performance across German States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
External exit exam	17.86 ***	13.25 **	15.25 **	16.23 ***	18.11 ***	17.93 ***		15.42 ***	12.76 ***	19.27 ***	16.87 ***
	(3.71)	*(5.36)*	(5.48)	*(4.79)*	*(4.08)*	*(4.53)*		*(4.57)*	*(3.82)*	*(5.98)*	*(3.67)*
Socio-economic background	81.69 ***	57.67	67.34	77.14 **	84.20 **	84.65 **	82.82 **	83.67 ***	56.71 **	88.34 **	64.64 **
	(26.53)	*(33.48)*	*(38.01)*	*(28.49)*	*(30.37)*	*(29.67)*	*(36.55)*	*(26.78)*	*(23.70)*	*(34.21)*	*(28.65)*
Father employed	1.94 ***	2.99 **	2.07 ***	2.01 ***	1.87 **	1.77 **	1.79 **	1.76 ***	0.49	1.88 **	1.88 ***
	(0.50)	*(1.15)*	*(0.57)*	*(0.53)*	*(0.62)*	*(0.67)*	*(0.74)*	*(0.54)*	*(0.67)*	*(0.63)*	*(0.49)*
City state		3.83									
		(7.75)									
East German state		9.36									
		(7.74)									
GDP per capita			0.002								
			(0.46)								
Parental education			6.72								
			(14.19)								
Migration background				-0.12							
				(0.21)							
Expenditure per student					-0.88						
					(4.42)						
Class size						0.24					
						(1.25)					
Instruction time						5.62					
						(14.34)					
Comprehensive schools							-14.01 **	-4.50			
							(5.44)	*(4.88)*			
Conservative voters									89.56 **		
									(32.53)		
Conservative prime minister									2.29		
									(3.63)		
Private school enrollment										0.27	
										(0.86)	
Pre-school enrollment										-0.16	
										(0.58)	
Slope of socio-economic gradient											-0.85
											(0.63)
Constant	320.81 ***	237.15 **	223.29	318.96 ***	330.16 ***	311.16 ***	347.04 ***	338.35 ***	402.24 ***	336.63 ***	364.73 ***
	(40.34)	*(92.55)*	*(212.12)*	*(41.67)*	*(62.93)*	*(49.86)*	*(61.09)*	*(44.83)*	*(44.65)*	*(63.63)*	*(50.95)*
N	16	16	16	16	16	16	16	16	16	16	16
R^2	0.835	0.856	0.843	0.839	0.835	0.837	0.688	0.846	0.907	0.837	0.858
R^2 (adj.)	0.793	0.784	0.765	0.781	0.775	0.756	0.610	0.791	0.860	0.755	0.806

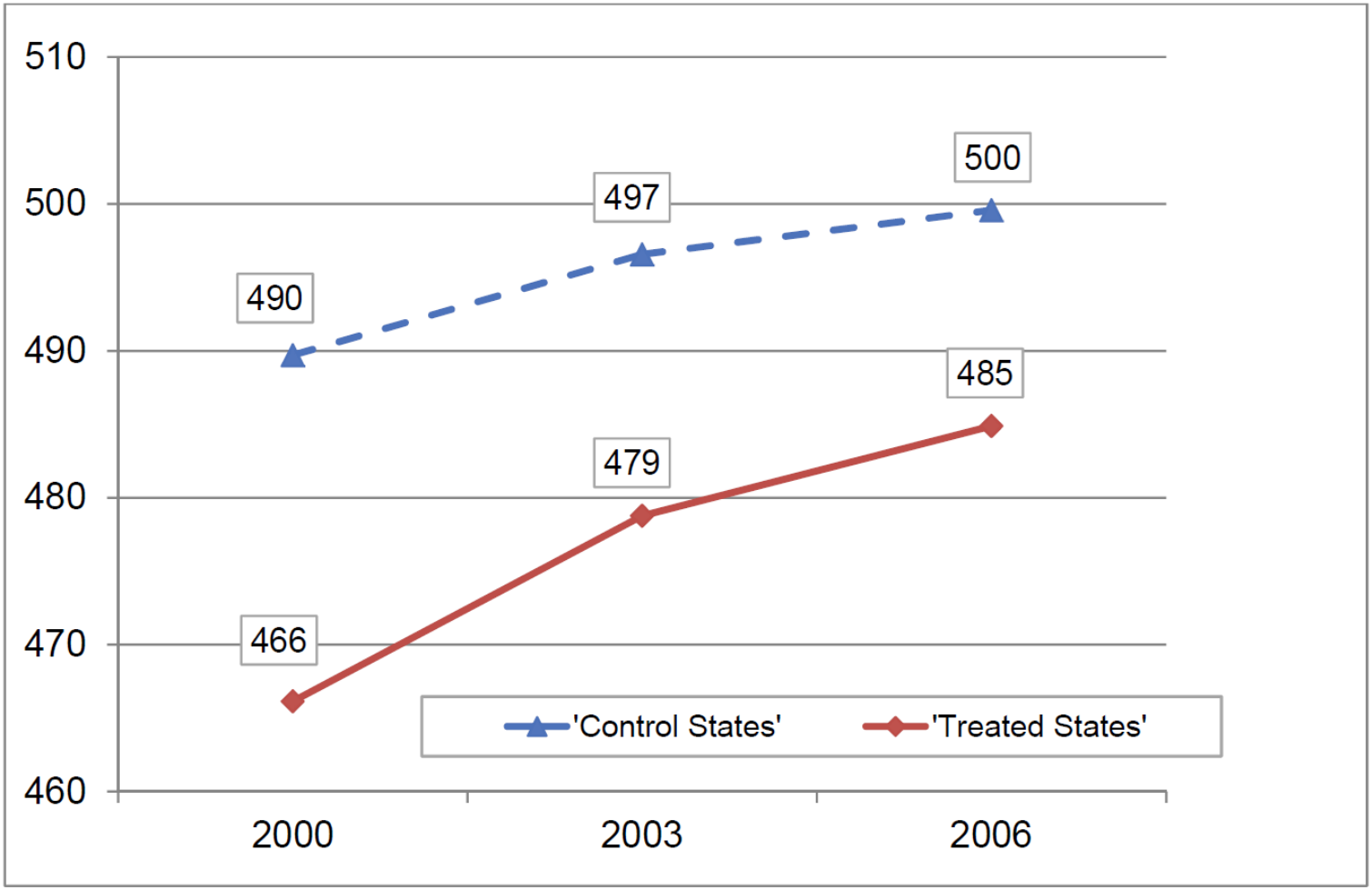
Central Exams and Student Performance across OECD Countries and German States

	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
External exit exam	22.48 **	22.48 **	21.70 ***	21.56 ***	18.48 ***	18.55 ***	19.96 ***	20.36 ***
	(10.59)	*(8.85)*	(5.95)	*(6.06)*	*(5.78)*	*(5.48)*	*(7.35)*	*(6.70)*
Socio-economic background	27.71	27.71 *	28.21 **	29.22 **	37.77 ***	38.31 ***	8.87	11.68
	(18.45)	*(15.40)*	*(12.47)*	*(13.51)*	*(12.47)*	*(11.74)*	*(13.05)*	*(12.62)*
Father employed	1.95 **	1.95 ***	1.85 ***	1.85 ***	1.49 **	1.45 ***	2.89 ***	2.64 ***
	(0.83)	*(0.69)*	*(0.57)*	*(0.57)*	*(0.55)*	*(0.52)*	*(0.55)*	*(0.50)*
Private school enrollment					0.51 ***	0.51 ***		0.50 ***
					(0.16)	*(0.15)*		*(0.16)*
Class size				0.19				
				(0.90)				
Germany (dummy)		-11.30			2.51			
		(124.64)			*(11.37)*			
External exit exam x Germany		-4.63						
		(13.52)						
Socio-economic background		53.98						
x Germany		*(74.68)*						
Father employed x Germany		-0.01						
		(1.54)						
Private school enrollment					-0.47			
x Germany					*(1.56)*			
Constant	332.11 ***	332.11 ***	338.01 ***	333.92 ***	360.88 ***	363.57 ***	259.46 ***	272.08 ***
	(67.60)	*(56.45)*	*(44.95)*	*(49.51)*	*(43.66)*	*(41.34)*	*(43.95)*	*(39.34)*
N	28	44	44	44	42	42	54	50
R^2	0.646	0.663	0.648	0.648	0.732	0.732	0.664	0.699
R^2 (adj.)	0.602	0.598	0.622	0.612	0.686	0.703	0.644	0.672
F (Germany and interactions)		0.40			0.05			
Prob. > *F*		0.805			0.953			

Identification

- Jürges/Schneider/Büchel: The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of the European Economic Association* 3, 2005.
  - Differences-in-differences approach on German TIMSS-95 data
  - Exploit that in some secondary-school tracks, states with central exit exams have them in math but not science
  - Find smaller but still substantial effects
  - Note: approach assumes that there are no spillovers between achievement in math and in science

Effects of the Introduction of Central Exams in German States



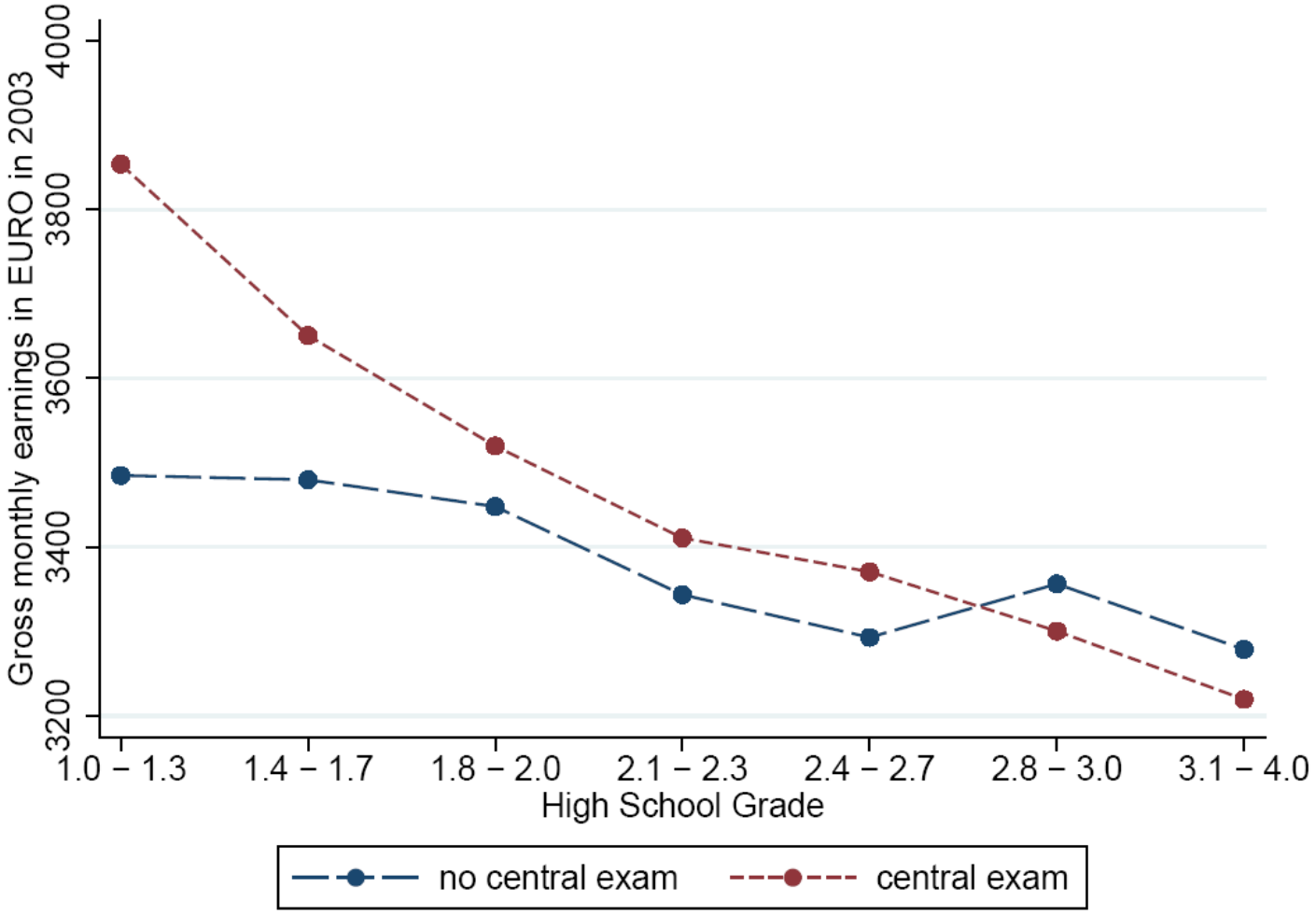
Notes: This descriptive figure shows the time trend in average reading performance between 2000 and 2006 for those states that had introduced CEEs between 2000 and 2006 for at least one type of school certificate (referred to as 'Treated States'), and those states that had not changed the CEE regime by this same time period ('Control States'). Each state is given equal weight. For the type of school certificate for which CEEs were introduced, the exact year of the introduction, as well as further details see Table 4.2 and Figure 4.2.

Source: Luedemann (2011)

Central Exams and Labor-Market Outcomes

- Effects on income
  - For students from low-track schools (11.6%) (and women from medium-track schools)
- Effects against unemployment
  - For students from low- and high-track schools (4.3/2.5pp) (Piopiunik/Schwerdt/Woessmann 2012)

School-Leaving Grade Has Signal on the Labor Market



Note: Grades are pooled in 6 categories according to the German grade classification scheme with a different grade label every .3 grade point. Grades below 3 are pooled into one category to ensure comparable numbers of observations within categories.

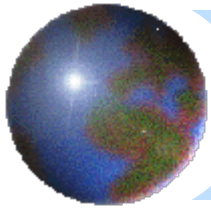
Source: Schwerdt/Woessmann (2012)

Additional Accountability Measures in PISA 2003

- Positive effects of various accountability measures:

1. Aimed primarily at students:
   - Use of assessments for decisions on student promotion and retention
2. Aimed at teachers:
   - Internal monitoring of teacher lessons by principal
   - External monitoring of teacher lessons by inspectors
3. Aimed at schools:
   - Assessments used to compare schools to district or national performance

- Combined effect:
  - Equivalent of more than one and a half grade-level equivalents on PISA test



How Autonomy Affects Student Performance
— *Depending on Given Incentives* —

- School autonomy: allows
  1. Use of superior local knowledge
     - If there are local knowledge advantages
  2. Opportunistic behavior
     - If there are opposing interests
     - *and* asymmetric information (not eased by central exams)
- School autonomy may be good or bad
  - Depending on whether in a given decision-making area,
    - there are **local knowledge advantages** and/or **incentives for opportunistic behavior**
  - and on whether there are **central exams**
    - Because central exams can ease the asymmetric information

How Central Exams Change Behavior
— Thus Changing the Effects of Autonomy —

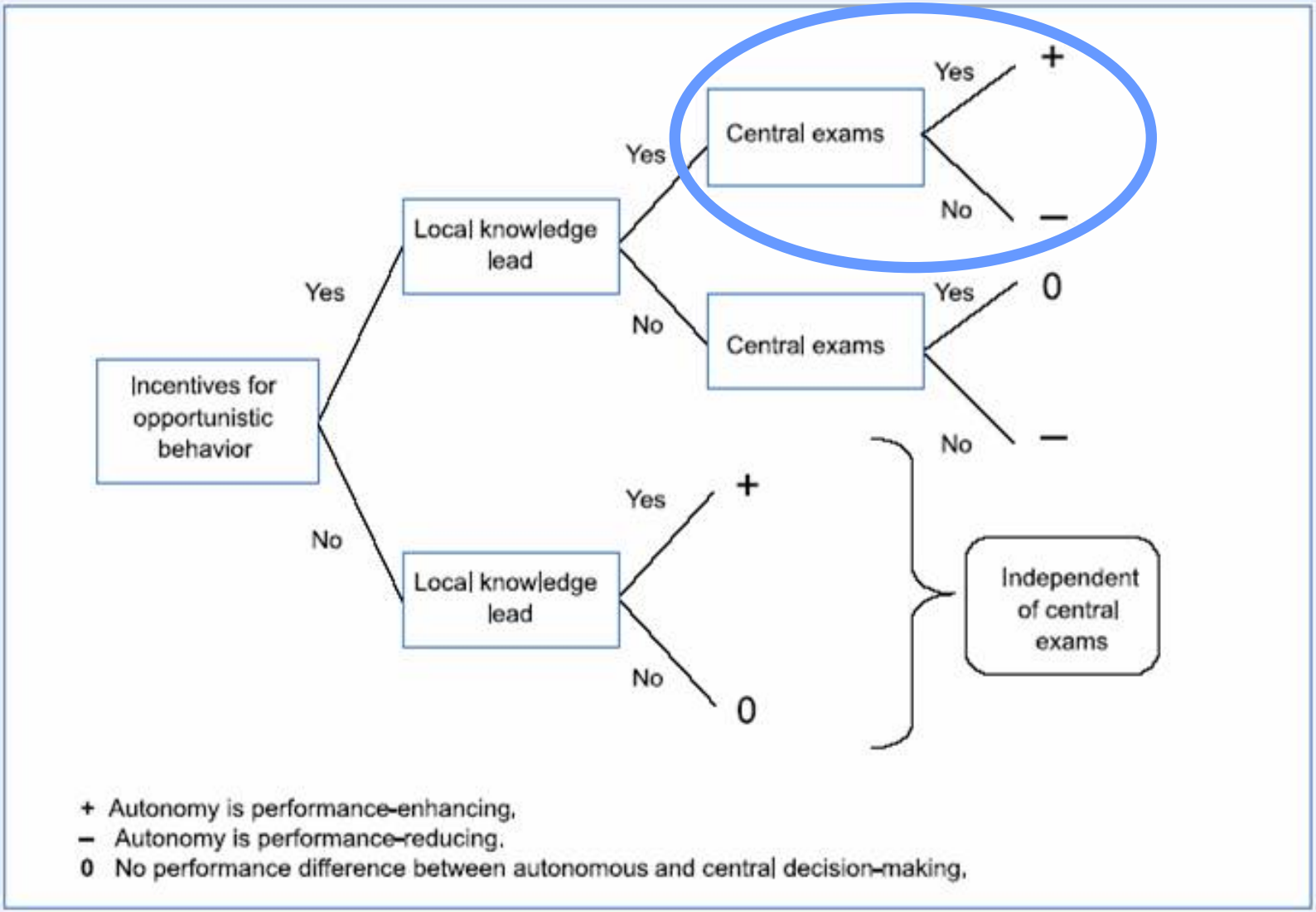
- Central exams provide *information*
  - On how individual students perform relative to the national (or regional) student population

→ Central exams ease the monitoring problems inherent in education systems
  - → Align incentives of local decision-makers with goals of the system
  - → Make it more likely that schools act according to the goals of the system if they are given autonomy

= By introducing accountability,
central exams ease the “bad” effects of autonomy,
ensuring a “good” net effect

Effects of Autonomy on Student Performance

— *With and Without Central Exams* —



Note: "Incentives for opportunistic behavior" and "local knowledge lead" are features of the respective decision-making area which can be organized either autonomously or non-autonomously.

The Interaction Specification

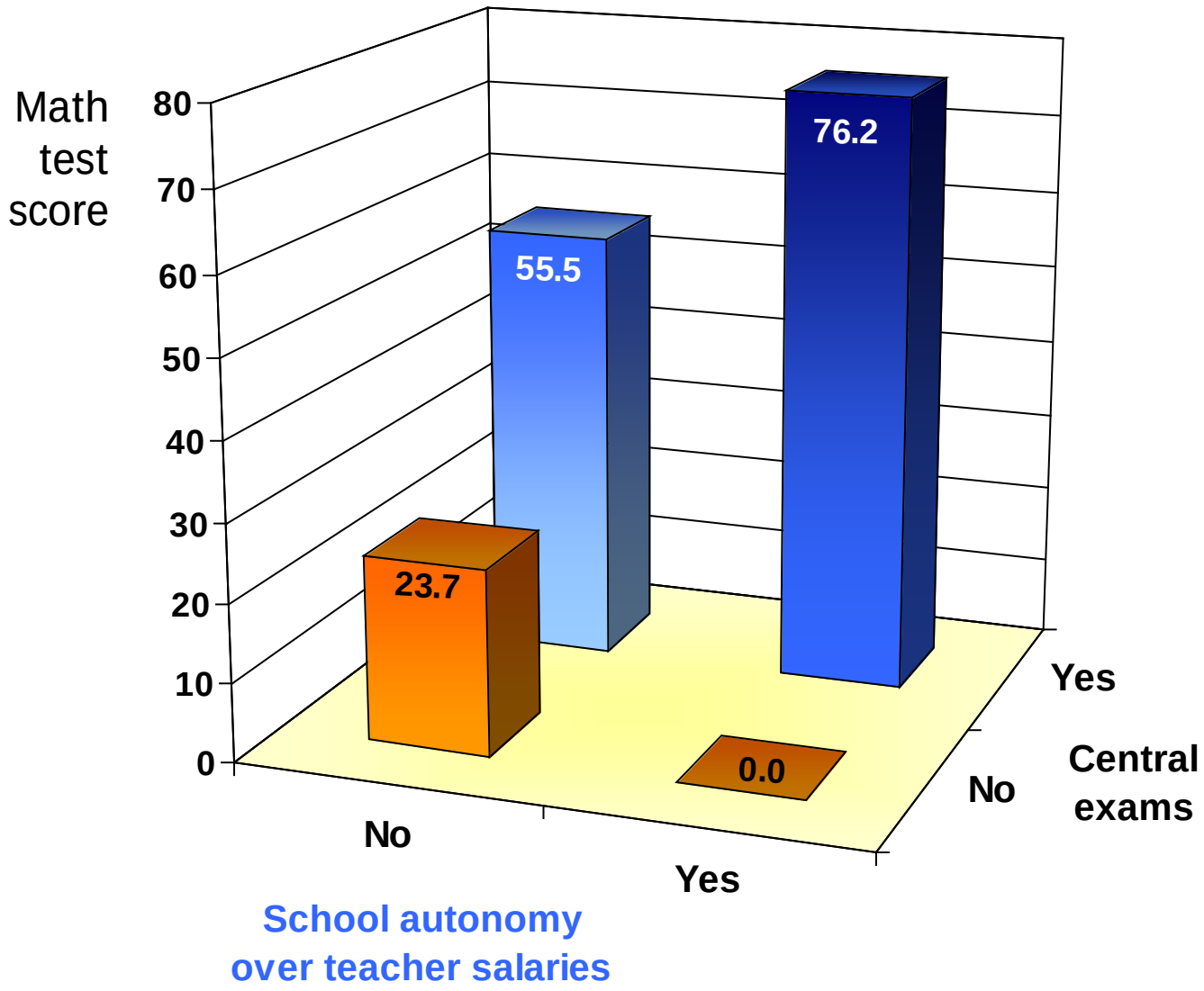
– Results: TIMSS/TIMSS-Repeat –

	Math		Science	
	Coefficient	Interaction	Coefficient	Interaction
School responsibility				
School budget	-6.9 $^{+}$ *(2.8)*	7.7 $^{+}$ *(3.5)*	-12.0 * *(2.6)*	16.1 * *(3.5)*
Purchasing supplies	7.1 $^{+}$ *(3.2)*	-5.7 *(5.0)*	15.6 * *(3.1)*	-6.2 *(5.4)*
Hiring teachers	21.6 * *(2.6)*	-20.2 * *(3.1)*	0.3 *(1.9)*	4.6 $^{\circ}$ *(2.6)*
Determining teacher salaries	-28.3 * *(3.6)*	50.2 * *(4.1)*	-8.2 * *(2.6)*	29.2 * *(3.1)*
Teachers' influence				
Class teacher has strong influence on				
Money for supplies	-24.7 * *(5.1)*	29.1 * *(6.3)*	-6.9 $^{\circ}$ *(3.6)*	13.6 * *(4.5)*
Kind of supplies	3.0 *(2.8)*	-3.5 *(3.8)*	6.0 * *(2.0)*	-3.7 *(2.9)*
Subject matter	-12.3 * *(2.3)*	8.7 * *(2.8)*	-4.6 * *(1.7)*	-0.7 *(2.2)*
Textbook	11.6 * *(3.1)*	-11.7 * *(3.6)*	6.3 * *(1.8)*	-9.9 * *(2.6)*
Strong influence on curriculum				
Teacher individually	14.6 * *(2.1)*	-3.9 *(2.7)*	14.5 * *(1.8)*	-7.4 * *(2.5)*
Subject teachers	-5.0 $^{+}$ *(2.4)*	2.8 *(3.1)*	-5.8 * *(2.1)*	8.2 * *(2.8)*
School teachers collectively	-14.7 * *(2.1)*	6.5 $^{+}$ *(2.8)*	-15.3 * *(1.9)*	14.4 * *(2.6)*
Teacher unions	-8.5 *(5.4)*	-29.5 * *(8.7)*	-6.7 *(5.1)*	-30.0 * *(9.1)*
Students (observations)	447,089		447,089	
Schools (PSUs)	12,175		12,175	
Countries	77		77	
R^2	0.296		0.266	

Coefficient: Coefficient on the dummy (= effect in systems without exit exams)

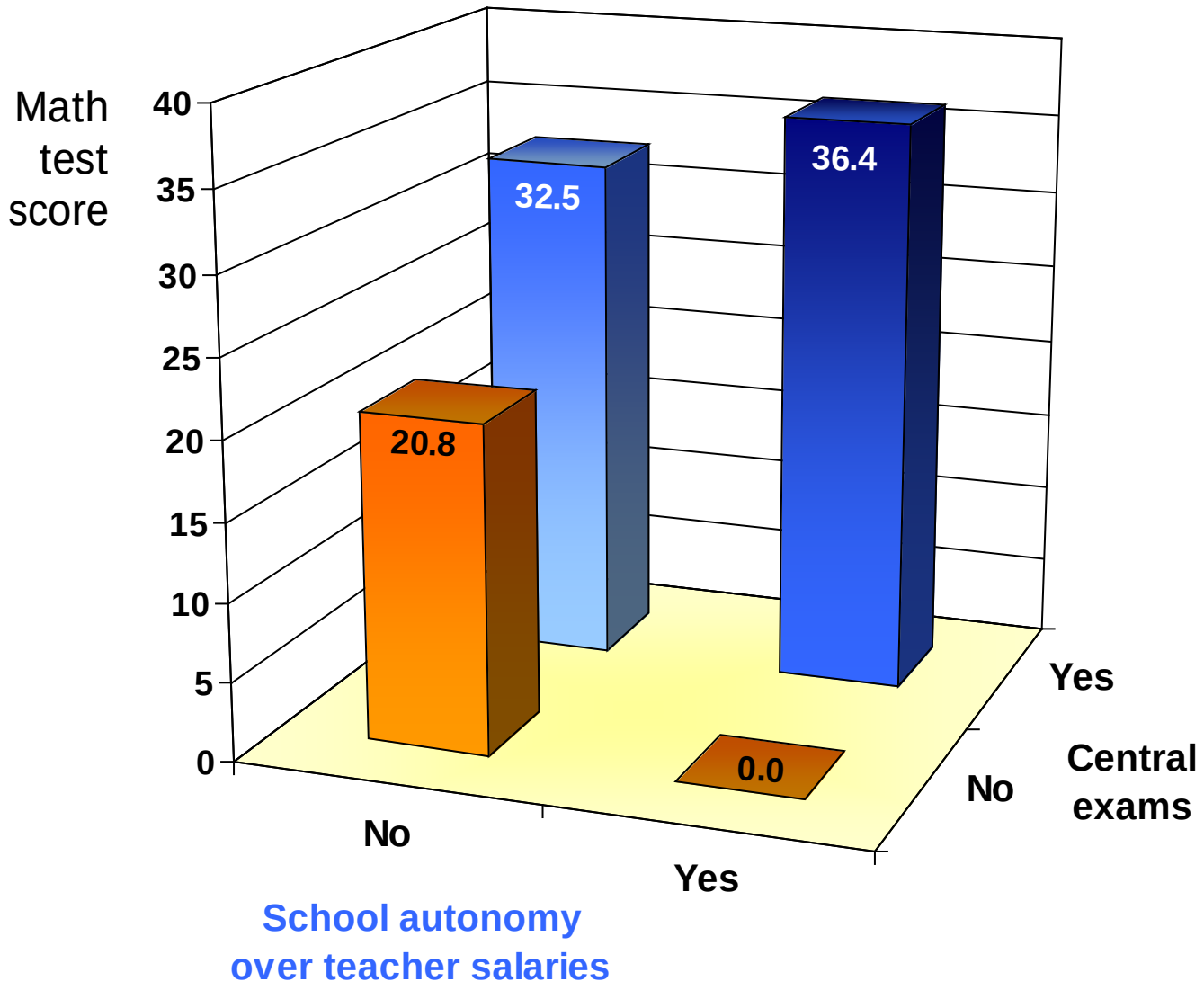
Interaction: Coefficient on interaction term between the dummy and exit exams (= difference in the effect between systems without and with exit exams)

Central Exams, School Autonomy, and Student Performance



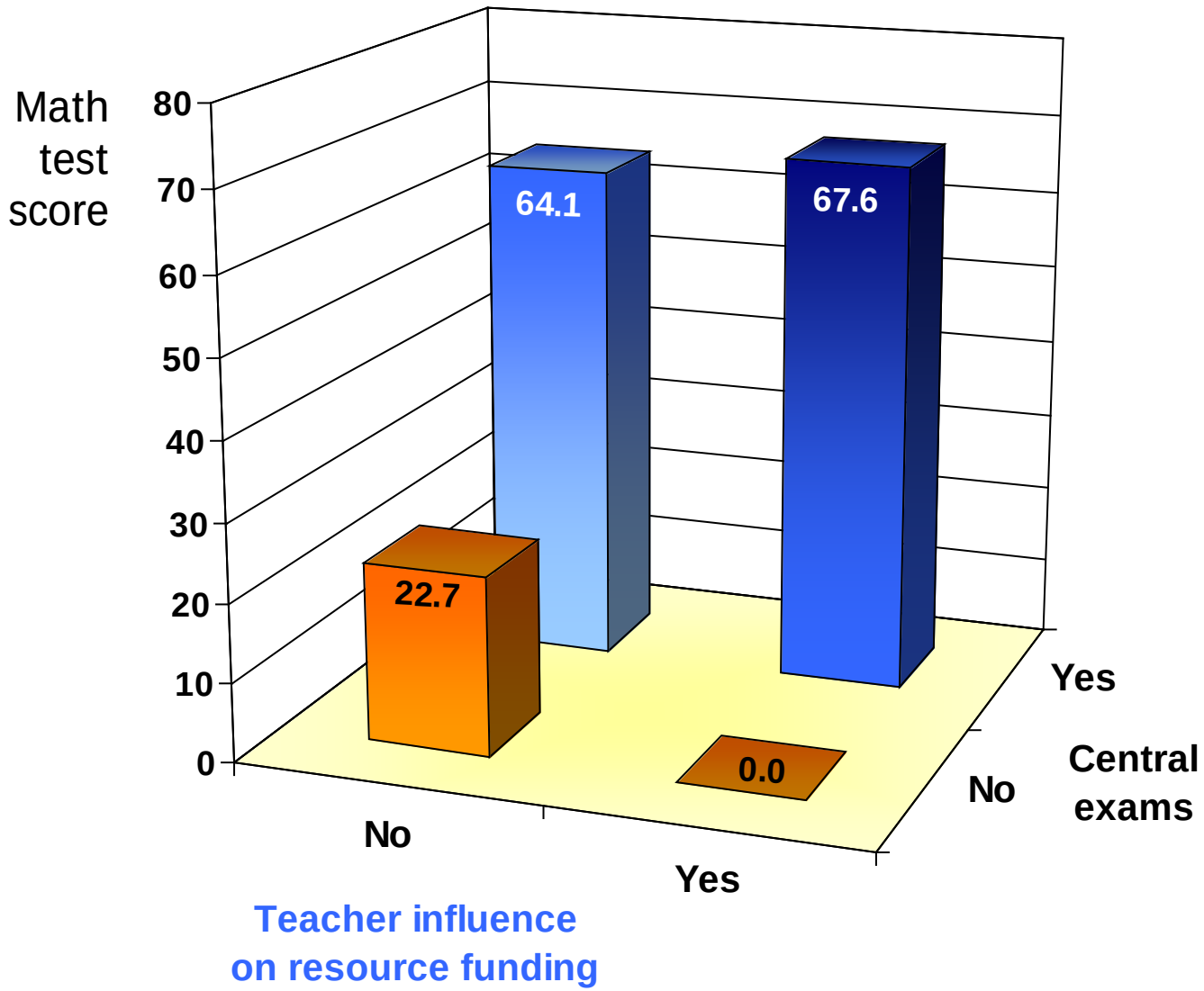
TIMSS + TIMSS-R

Central Exams, School Autonomy, and Student Performance



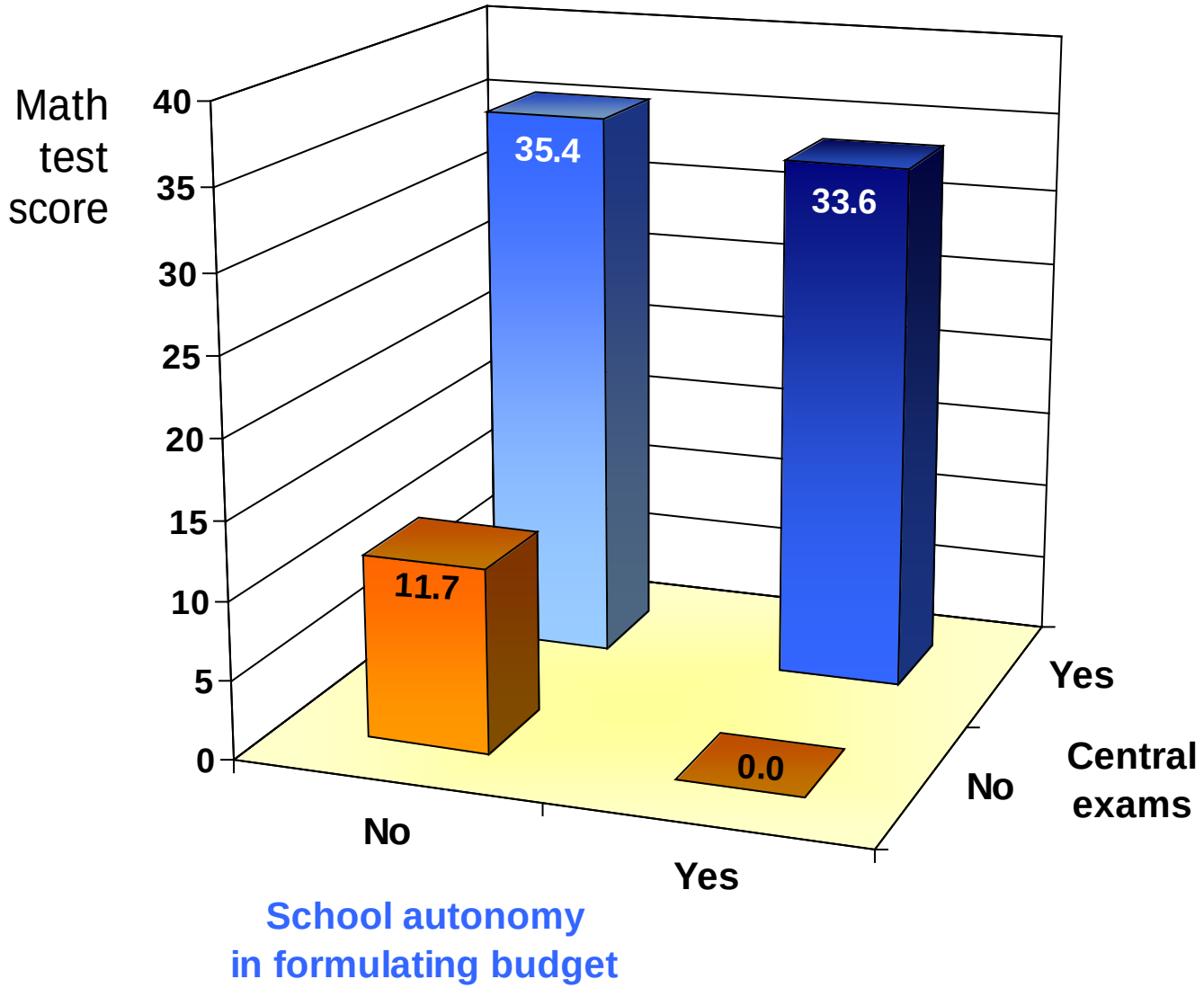
PISA 2000

Central Exams, School Autonomy, and Student Performance



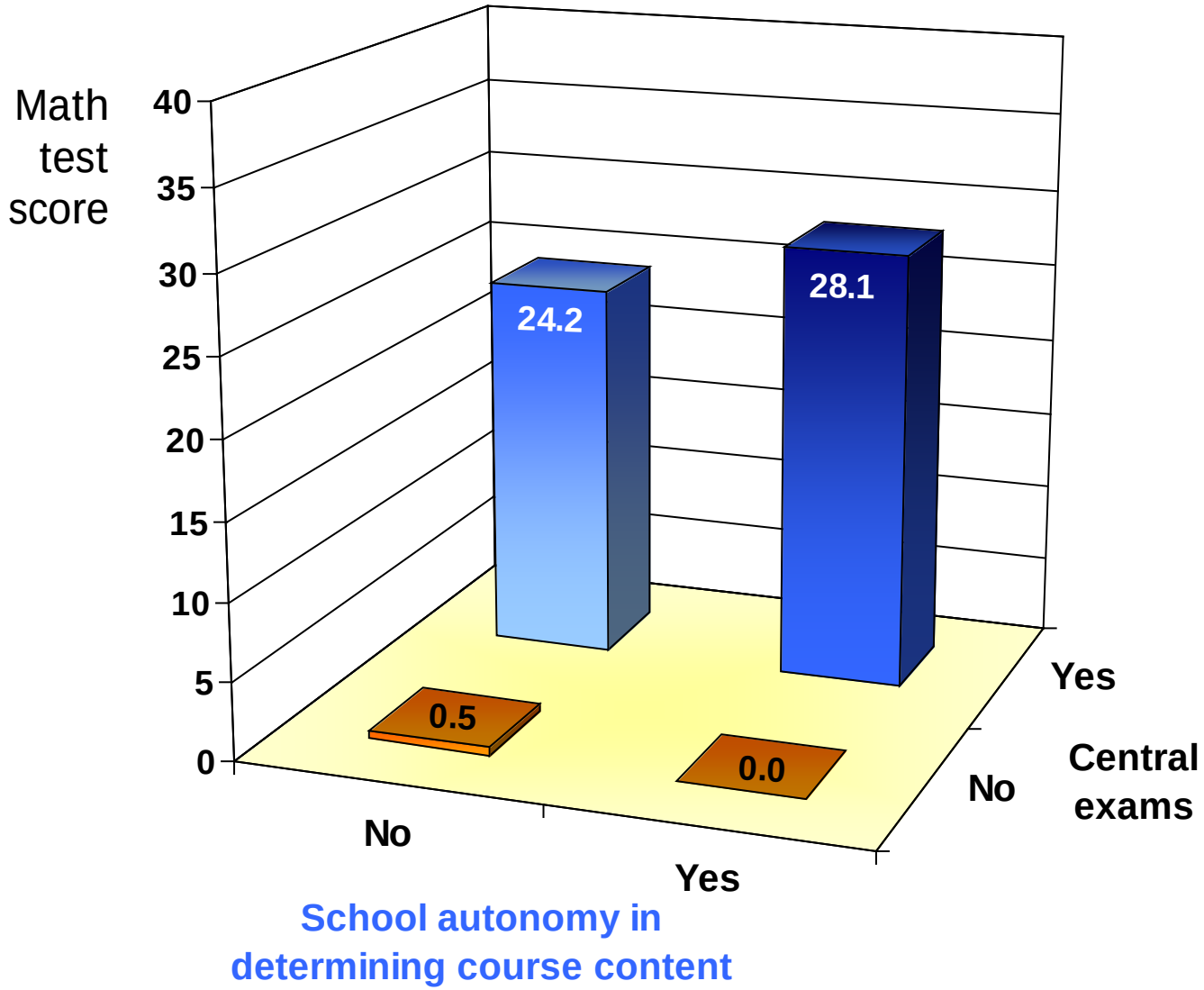
TIMSS + TIMSS-R

Central Exams, School Autonomy, and Student Performance



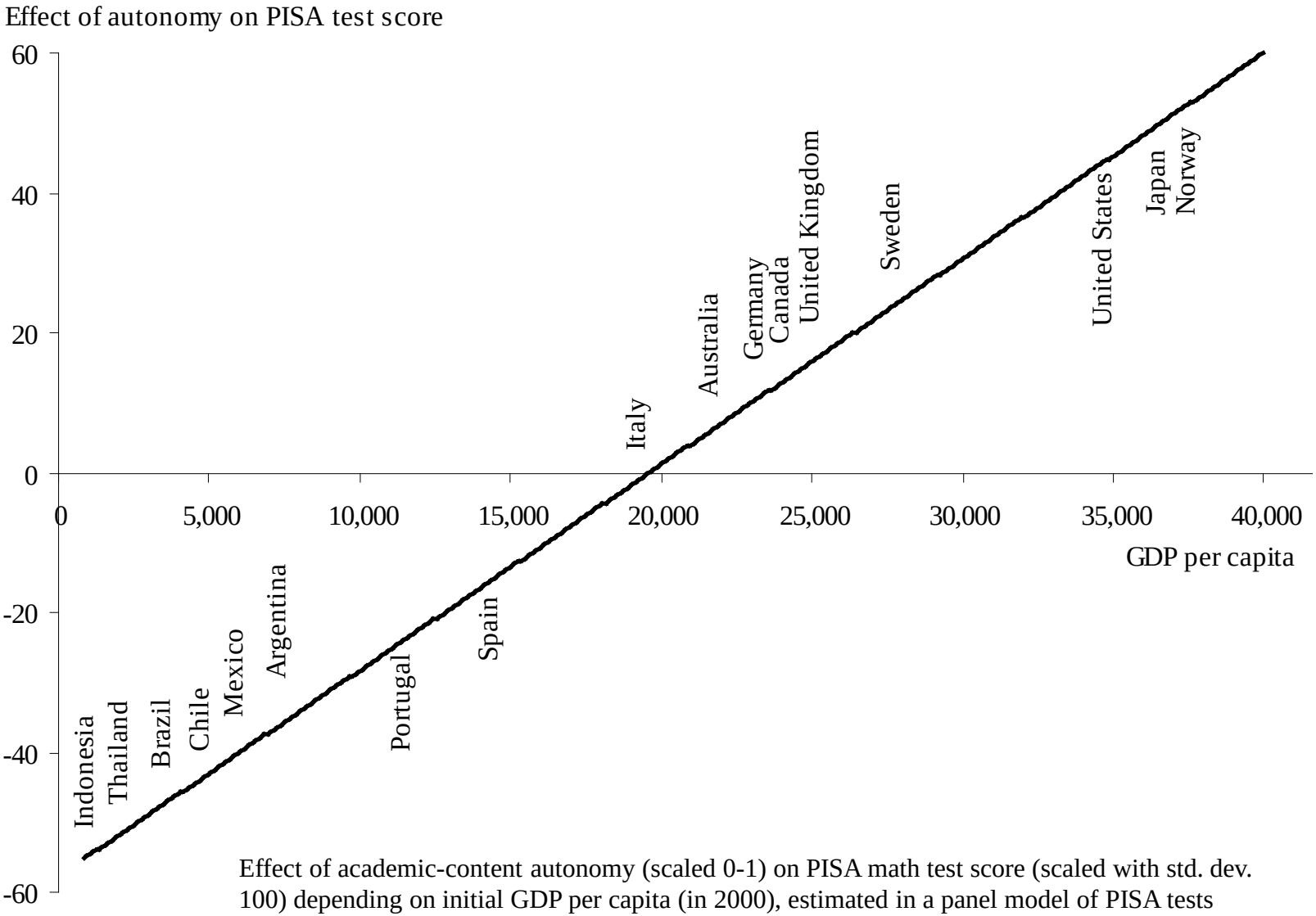
PISA 2003

Central Exams, School Autonomy, and Student Performance



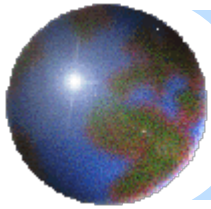
PISA 2003

Central Exit Exams, Autonomy Reforms and PISA Improvements



Effect of academic-content autonomy (scaled 0-1) on PISA math test score (scaled with std. dev. 100) depending on initial GDP per capita (in 2000), estimated in a panel model of PISA tests 2000-2009. Example countries illustrate initial level of GDP per capita.

Source: Hanushek/Link/Woessmann (2011)



Accountability in Education around the World: Lessons from International Achievement Tests

- Consistent evidence that central exams are positively associated with student achievement
  - In many international tests; across German states
- Student achievement also positively associated with additional school-level accountability mechanisms
  - Internal and external monitoring of teacher lessons
  - Assessments used to compare schools to district or nation
- As a general tendency, school autonomy interacts positively with central exams
  - Accountability as pre-requisite for autonomy reforms
- Accountability changes the behavior of students, teachers and schools