

*Understanding causal influences  
on educational achievement  
through analysis of within  
country differences over time*

Jan-Eric Gustafsson

Department of Education

Gothenburg University



# *Comparative Studies of Educational Achievement*

- The International Association for the Evaluation of Educational Achievement (IEA)
  - Founded in 1959, by a group of social scientists (e. g., Benjamin Bloom, James Coleman, Torsten Husén).
  - ”The aim is to look at achievement against a wide background of school, home, student and societal factors in order to use the world as an educational laboratory so as to instruct policy makers at all levels about alternatives in educational organization and practice.”
  - Mathematics, science, reading, writing, science, civics, ...
  - Currently close to 70 member countries.
- The OECD Programme for International Student Assessment (PISA)
  - First survey in 2000 (reading, mathematics, science); repetition in a three-year cycle.
  - All OECD countries, along with associate countries.



# *IEA studies of mathematics*

<b>Year</b>	<b>Abbreviation</b>	<b>Name of the Study</b>	<b>No of Countries</b>
1964	FIMS	First International Mathematics Study	12
1980	SIMS	Second International Mathematics Study	20
1995	TIMSS	Third International Mathematics and Science Study	45
1999	TIMSS-R	TIMSS-Repeat	39
2003	TIMSS 2003	Trends in Mathematics and Science Study	50
2007	TIMSS 2007	Trends in Mathematics and Science Study	60



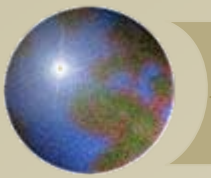
## *Purposes of the comparative studies*

- Scientific purposes (“The world as an educational laboratory”)
- Comparisons of educational achievement (“benchmarking”)
- Drift from scientific purposes to benchmarking.
- Still explanations of outcomes are essential.
  - Explanations of outcomes are expected.
  - If explanations are not provided, they will be supplied more or less spontaneously. Such explanations typically have little relation to data and findings, instead reflecting preconceived notions and political agendas.



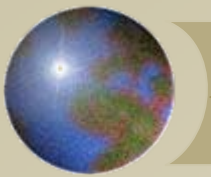
## *Lots of high-quality data, which offer great opportunities and challenges*

- Complex stratified cluster sampling designs.
- Sophisticated measurement with matrix sampling designs and item response theory (IRT) techniques.
- Hierarchical data (students within classes, classes within schools, schools within countries).
- Errors of measurement in observed variables.
- Missing data.
- Observed variables at low scale levels (nominal and ordinal variables).
- Cross-sectional data making it difficult to make correct inferences about causality because of selection bias and omitted variables (correlation is not causation).



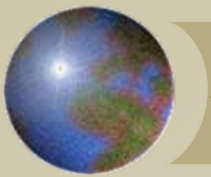
# *Selection bias*

- Mullis, Campbell and Farstrup (1993) correlated the amount of direct teaching of reading that grade 4 students had obtained with their level of reading performance. The correlation was negative, showing that students who had obtained more teaching had a lower level of reading achievement.
- Two interpretations: (1) teaching of reading hurts achievement; and (2) poor readers had obtained more teaching. Explanation 2 is the more likely one.
- Selection bias (or reversed causation or endogeneity problems)

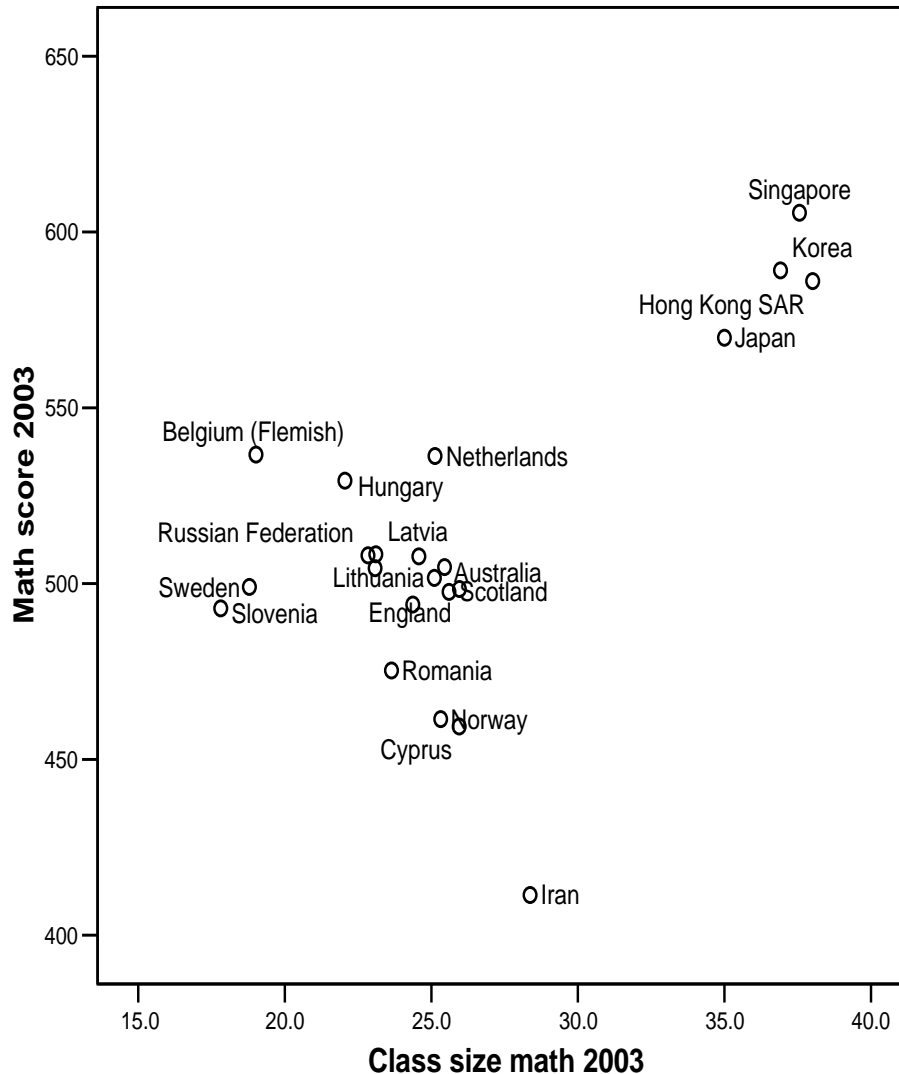


## *Omitted variables*

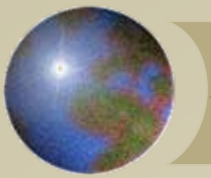
- A correlation may be misinterpreted in terms of causation, when measures of the real causes are not included.
- For grade 8 in TIMSS 2003 there is at the country level a correlation of .60 between class size and mathematics achievement (computed for the subset of 22 countries that also participated in TIMSS 1995). Thus, larger classes is related to higher achievement, but is there a causal relation?



## *Class size and mathematics achievement in TIMSS 2003*



The positive correlation between class size and achievement is due to a group of four Confucian Asian countries having high achievement and large classes. However, the reason for the high achievement is probably not the large classes, but is more likely due to cultural norms and values, quality of instruction, and other characteristics of Confucian Asian societies.



## *How can the problems of selection bias and omitted variables be solved?*

- One approach to reduce the impact of selection bias is to analyze data at a higher level of aggregation. For example, compensatory resource allocation is done at individual level, such as putting low-achieving students in small classes. This will tend to cause class size to be positively correlated with achievement. But if there are multiple classes in each school, analyses at the school level will not be affected by such bias. Neither will analyses at the country level be affected by selection bias.
- One way to reduce the problems associated with omitted variables is to investigate change over time for fixed units, thereby making the units their own controls. This is the basic idea of panel (or longitudinal) designs.



## *Controlling for selection bias and omitted variables through analysis of within country trends*

- Data aggregated to the country level should not be affected by selection bias.
- If repeated observations are made of the same set of countries many variables are more or less constant, while others change. Thus, "fixed-country" analysis which relates change in independent variables to change in the dependent variable should be a way to reduce the problem of omitted variables.
- Therefore, analysis of within country trends over time should be a way to deal with both selection bias and omitted variables.



## *Example 1: Student age and achievement*

- Many studies have shown that school achievement and intellectual performance is related to age.
- According to results from TIMSS 1995 the performance difference between grades 3 and 4 was 57 points and between grades 7 and 8 it was 30 points.
- It has been estimated that about one third of the effect is due to chronological age, and two thirds to another school year.



## *Data for countries in TIMSS 1995 and 2003. grade 8*

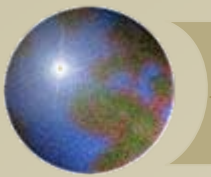
<b>Mathematics achievement and student age for countries in TIMSS 1995 and 2003</b>						
<b>Country</b>	<b>Math 1995</b>	<b>Math 2003</b>	<b>Math change</b>	<b>Age 1995</b>	<b>Age 2003</b>	<b>Age change</b>
Australia	507	505	-2	14.04	13.90	-0.14
Belgium (Flemish)	550	537	-13	14.14	14.10	-0.04
Cyprus	468	459	-8	13.74	13.80	0.06
England	498	498	1	14.05	14.30	0.25
Hong Kong SAR	569	586	17	14.18	14.40	0.22
Hungary	527	529	3	14.28	14.50	0.22
Iran	418	411	-7	14.63	14.40	-0.23
Japan	581	570	-11	14.38	14.40	0.02
Korea	581	589	8	14.20	14.60	0.40
Latvia	488	508	20	14.27	15.00	0.73
Lithuania	472	502	30	14.26	14.90	0.64
Netherlands	529	536	7	14.35	14.30	-0.05
New Zealand	501	494	-7	14.00	14.10	0.10
Norway	498	461	-37	13.88	13.80	-0.08
Romania	474	475	2	14.58	15.00	0.42
Russian Federation	524	508	-16	14.03	14.20	0.17
Scotland	493	498	4	13.70	13.70	0.00
Singapore	609	605	-3	14.55	14.30	-0.25
Slovak Republic	534	508	-26	14.26	14.30	0.04
Slovenia	494	493	-2	13.82	13.90	0.08
Sweden	540	499	-41	14.93	14.90	-0.03
United States	492	504	12	14.23	14.20	-0.03



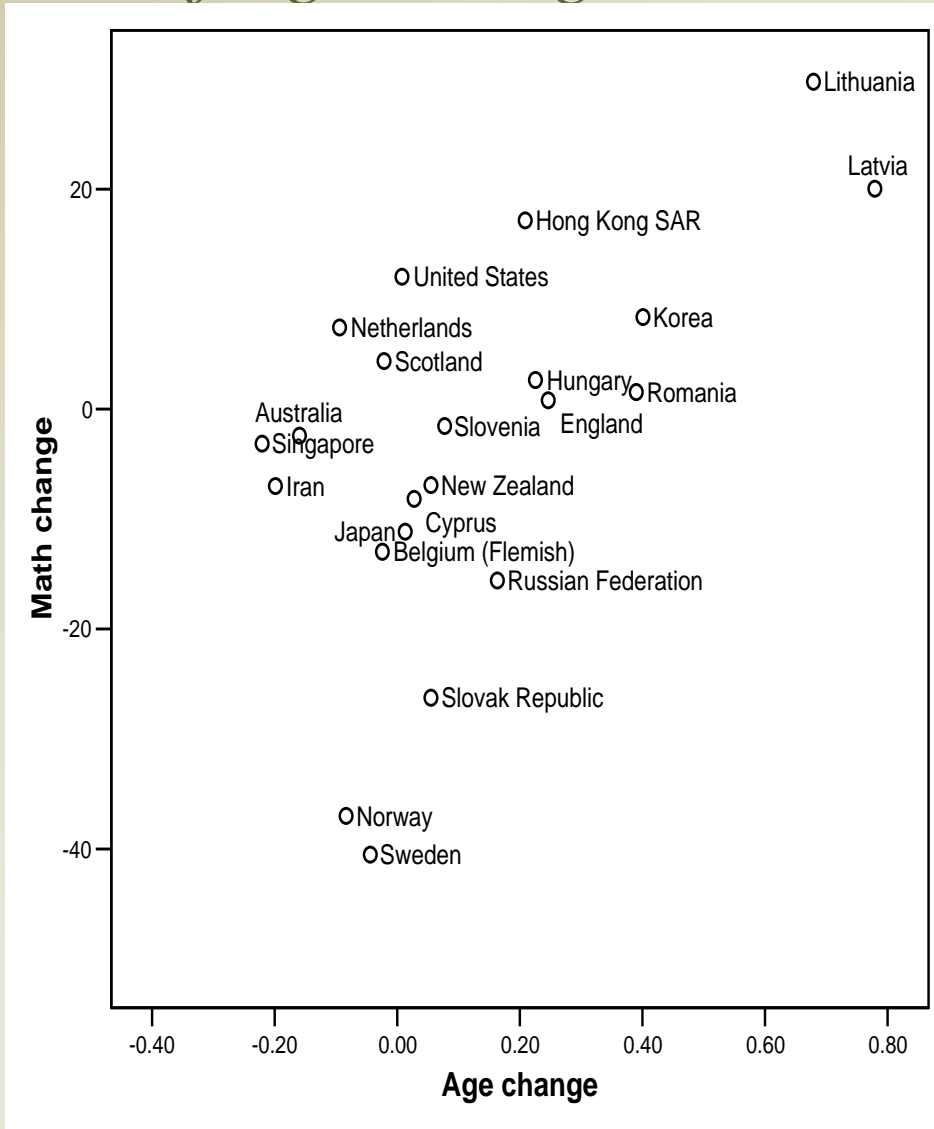
## *Correlations between age and achievement, grade 8*

Table 2. Correlations between achievement and age, grade 8 (N=22).

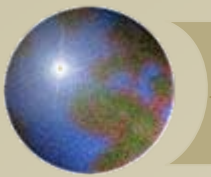
	<b>Math 1995</b>	<b>Math 2003</b>	<b>Math change</b>	<b>Age 1995</b>	<b>Age 2003</b>	<b>Age change</b>
Math 1995	1.00					
Math 2003	0.93	1.00				
Math change	-0.12	0.26	1.00			
Age 1995	0.19	0.14	-0.12	1.00		
Age 2003	0.05	0.16	0.29	0.75	1.00	
Age change	-0.14	0.08	0.58	-0.01	0.65	1.00



## *Plot of age change and math change, grade 8*



The unstandardized regression coefficient for Math change on Age change is 38. This implies that the effect of one year (a chronological year and a 12 month school year) is 38 points. For a year with a 9 month school year this corresponds to an effect of 32 points.



## *Data for countries in TIMSS 1995 and 2003. grade 4*

Table 3. Mathematics achievement and student age for countries in TIMSS 1995 and 2003, grade 4 (N=15)

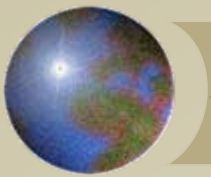
<b>Country</b>	<b>Math 1995</b>	<b>Math 2003</b>	<b>Math change</b>	<b>Age 1995</b>	<b>Age 2003</b>	<b>Age change</b>
Australia	495	499	4	9.86	9.89	0.03
Cyprus	475	510	35	9.84	9.90	0.06
England	484	531	47	10.04	10.27	0.23
Hong Kong	557	575	18	10.14	10.24	0.10
Hungary	521	529	7	10.41	10.55	0.14
Iran	387	389	2	10.50	10.40	-0.10
Japan	567	565	-3	10.39	10.41	0.02
Latvia	499	536	37	10.46	11.05	0.59
Netherlands	549	540	-9	10.26	10.23	-0.03
New Zealand	469	493	24	9.98	10.03	0.05
Norway	476	451	-25	9.87	9.81	-0.06
Scotland	493	490	-3	9.71	9.70	-0.01
Singapore	590	594	4	10.31	10.33	0.02
Slovenia	462	479	17	9.87	9.78	-0.09
United States	518	518	0	10.19	10.24	0.06



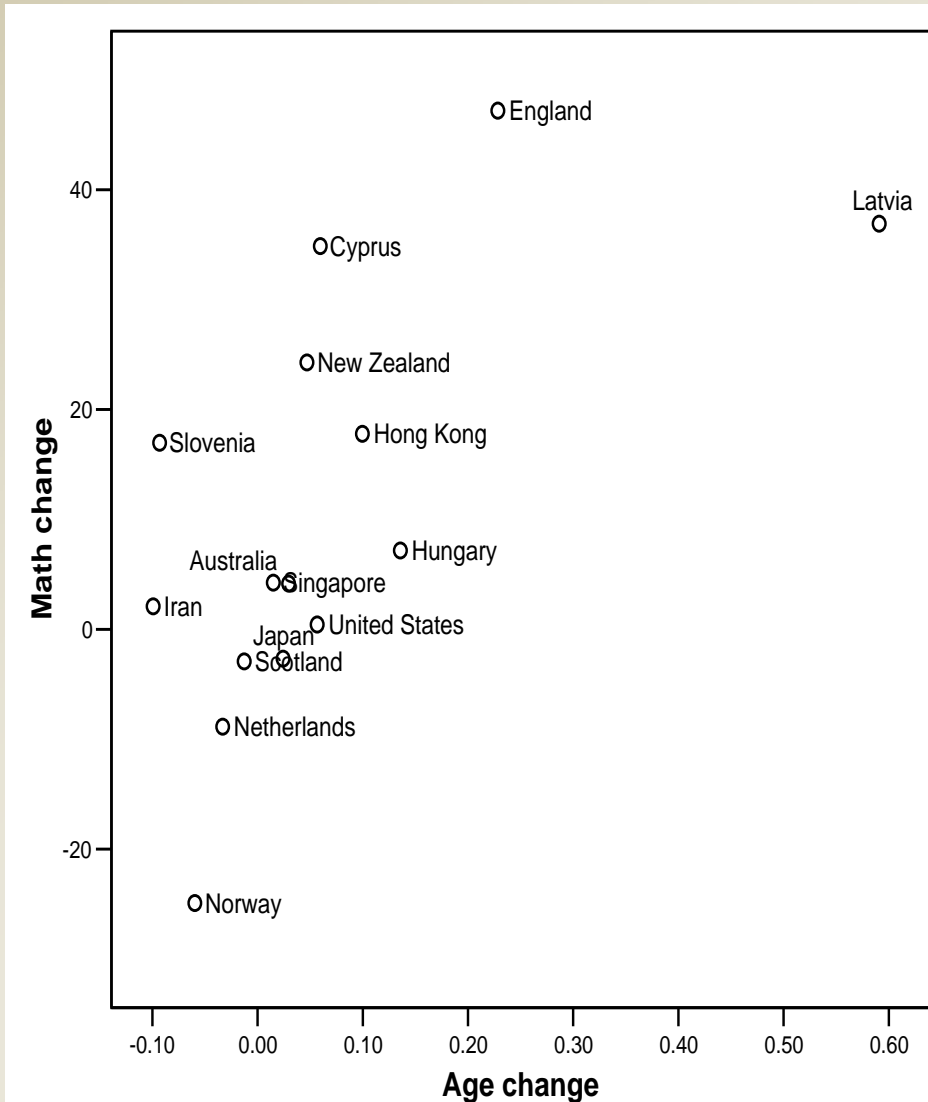
## *Correlations between age and achievement, grade 4*

Table 4. Correlations between achievement and age. Grade 4 (N=15).

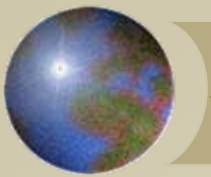
	<b>Math 1995</b>	<b>Math 2003</b>	<b>Math change</b>	<b>Age 1995</b>	<b>Age 2003</b>	<b>Age change</b>
Math 1995	1.00					
Math 2003	0.93	1.00				
Math change	-0.16	0.22	1.00			
Age 1995	0.21	0.21	-0.01	1.00		
Age 2003	0.22	0.33	0.29	0.89	1.00	
Age change	0.14	0.38	0.63	0.34	0.72	1.00



## *Plot of age change and math change, grade 4*

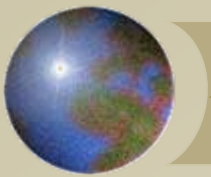


The unstandardized regression coefficient for Math change on Age change is 71. For a 9-month school year this corresponds approximately to 59 points.



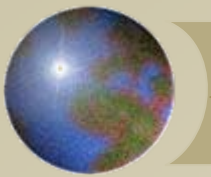
# *Conclusions about age effects*

- The estimates obtained through the analysis of within country change over time agree with what has been found with other methods.
- The age changes over time appear more or less random, The randomness solves the problems with omitted variables, so an unbiased estimate is obtained. In contrast, age differences between countries are related to many cultural and social factor, so cross-sectional correlations are biased estimates of the causal effect of age on achievement.
- The age changes over time accounted for 30 % to 40 % of the variance in the mathematics change. This suggests that measures should be taken to reduce the age differences over time when administering the assessments, or to statistically control for the age differences.



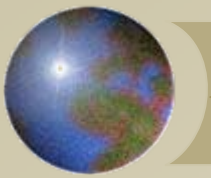
## *Example II: Effects of Class Size*

- Severe methodological problems, because typically cross-sectional data have been relied upon.
- During the 1990s large-scale randomized field-experiments were conducted, and advances were made in statistical methodology for analyzing naturally occurring variation.



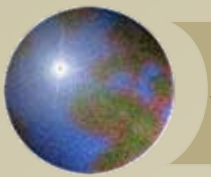
## *Student/Teacher Achievement Ratio (STAR)*

- Experiment on class size, which started in 1985 in Tennessee, USA.
- Students and teachers were randomly assigned to three class types: small classes (13 – 17 students), regular classes (22 – 25 students), and regular classes with an assistant teacher. About 100 classes of each type.
- Follow up from kindergarten to third grade in the first project phase, after which the experiment was terminated. Further follow ups throughout the school system.



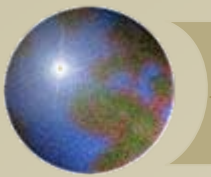
## *STAR: Results*

- Clear advantage for small classes in first grade for math and reading. Similar differences in second and third grade.
- Stronger effect of small classes for minority students.
- Follow up studies in grades 4 to 10 show lasting effects of small classes in most subject areas.
- Greater interest for going to college among students who attended a small class in early grades.



## *Studies of Natural Variation of Class Size*

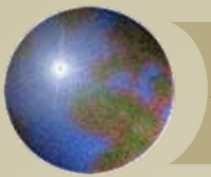
- Angrist and Lavy (1999) used the discontinuous effects on class size implied by a rule of maximum class size to investigate effects of class size. Positive effects of smaller classes in grades 4 and 5, but somewhat weaker effects than in STAR.
- Hoxby (2000) used variation in population size, as well as abrupt changes in class size caused by rules of maximum class size to study effects of class size. No effect was found.
- Hanushek and Luque (2003) analyzed TIMSS 1995 data at the school level within countries and found a tendency for smaller classes to be more effective in grade 4, but no effect of class size in grade 8.
- Meta analyses indicate that there is a positive effect of being in a small class in early grades (K-4), but that there is no effect in higher grades.



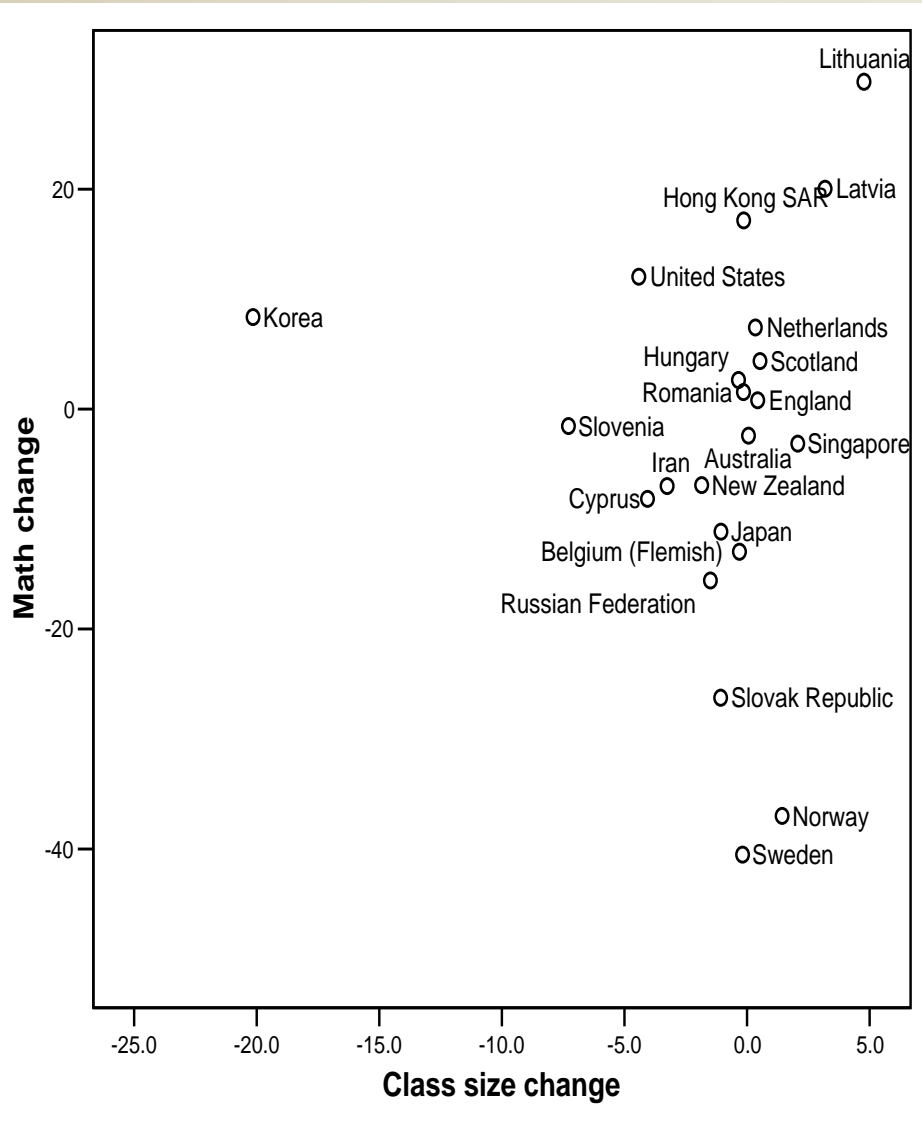
## *Correlations between class size and achievement for TIMSS 1995 - 2003, grade 8*

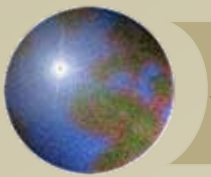
**Table 3. Correlations between achievement and class size**

	<b>Math 1995</b>	<b>Math 2003</b>	<b>Math change</b>	<b>Class size 1995</b>	<b>Class size 2003</b>	<b>Class size change</b>
<b>Math 1995</b>	1.00					
<b>Math 2003</b>	0.82	1.00				
<b>Math change</b>	-0.74	-0.23	1.00			
<b>Class size 1995</b>	0.45	0.52	-0.17	1.00		
<b>Class size 2003</b>	0.51	0.60	-0.17	0.93	1.00	
<b>Class size change</b>	0.18	0.25	-0.01	-0.12	0.26	1.00



## *Plot of class size change and achievement change, grade 8*





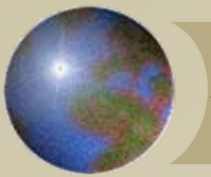
*Regressions of mathematics change on age change and class size change, grade 8 and grade 4*

Grade 8					
	B	Std. Error	Beta	t	Sig.
(Constant)	-7.94	3.68		-2.16	0.04
Age change	38.87	12.45	0.60	3.12	0.01
Class size change	-0.28	0.71	-0.07	-0.39	0.70

R-square = .38

Grade 4					
	B	Std. Error	Beta	t	Sig.
(Constant)	0.87	3.57		0.24	0.81
Age change	100.65	20.52	0.88	4.90	0.00
Class size change	-4.44	1.34	-0.60	-3.32	0.01

R-square = .68



## *Conclusions about effects of class size*

- The results show an effect of class size in the expected direction for grade 4 but not for grade 8, which is in agreement with previous findings.
- The grade 4 effect size is around .3, assuming a reduction of class size with 7 students. This is somewhat higher than what was found in STAR, but not unreasonably high.



# *Conclusions*

- The comparative studies of educational achievement do not allow simple inferences about causality, but the rich data is a gold mine for research, which should be fully exploited.
- In particular, the availability of trend data makes it possible to account for change over fixed sets of countries, and the examples presented here provide some basis for optimism that this approach may generate interesting substantive findings.
- The approach cannot be used to investigate effects of factors which keep constant over time, such as certain institutional factors.
- The trend analyses presented here are statistically unsophisticated, relying on correlations between difference scores. Other methods, such country-level growth curve analyses, could be applied.