

Beyond descriptions: International assessments and evidence-based policy

GA-63, Split, Croatia

Oct 04, 2022

Rolf Strietholt, rolf.strietholt@iea-hamburg.de

The Purpose of the IEA

Evidence-based policy is an idea proposing that policy decisions should be based on, or informed by, rigorously established evidence.

The IEA was established to conduct international large-scale assessments (ILSAs) that provide alternatives to ideology, common sense, or intuition in education.



The fact that these comparisons are cross-national should not be taken as an indication that the primary interest was, for instance, national means and dispersions in school achievements... The main objective of the study is to investigate the “outcomes” of various school systems by relating as many as possible of the relevant input variables ... to the output assessed by international test instruments.

Husén & Postlethwaite (1967, p. 30)

Achievements and Limitations

One of the most salient findings in the field of education is that there are huge national differences in educational outcomes observed in ILSAs. Although these findings have been replicated several times, at present, the reasons for such differences are not well understood.

- Policymakers and researchers often refer to the **descriptions** of achievement differences across and within countries reported in ILSAs to motivate or justify the need for education reforms.
- However, ILSAs offer little concrete **explanatory** knowledge about how to improve educational policy and practice (Klemenčič & Mirazchiyski, 2017)



In hindsight this appears to be a rather simplistic conceptualization (...) within the framework of a cross-sectional survey approach.

Husén (1979, p. 380)

Forms of Evidence

Research aims to develop a foundation of knowledge for evidence-based decisions. It is useful to sharply distinguish between two different forms of knowledge (e.g., Hamaker et al., 2020):

- **Descriptive research** aims to summarize characteristics of groups. For example, we may be interested in the question whether gender gaps in achievement exist.
- **Explanatory research** aims to understand the underlying causal mechanism, which can then be used to develop interventions. We have a clear idea about what variables are outcomes or effects of the causal process.
- Unfortunately, the two forms of knowledge are often confused.

A funny example...

The screenshot displays the SAGE Journals website interface. At the top, the SAGE Journals logo is visible alongside navigation links for Search, Browse, Resources, and Access Options. The journal title, "Scandinavian Journal of Public Health", is prominently displayed in the header, along with its Impact Factor (3.021) and 5-Year Impact Factor (2.938). The article title, "Are pupils' Programme for International Student Assessment (PISA) scores associated with a nation's fish consumption?", is the central focus. Below the title, the authors (Volker Schmiedel, Hans Vogt, Harald Walach) and publication details (First Published November 21, 2017) are listed. The article is categorized as a Research Article and is available in PubMed. The abstract is visible, discussing the relationship between fish consumption, PISA scores, and neurodevelopment. The left sidebar contains an Article Menu with options like Full Article, Content List, Figures & Tables, and Article Metrics. The right sidebar features a SAGE Disc Hub promotion and a Related Articles section.

https://journals.sagepub.com/doi/abs/10.1177/1403494817717834

SAGE journals

Search Browse Resources Access Options: Sign In Institution Society Cart

Scandinavian Journal of Public Health

3.021 Impact Factor
5-Year Impact Factor 2.938
Journal Indexing & Metrics »

Journal Home Browse Journal Journal Info Stay Connected Submit Paper

Article Menu Close

Access Options

Full Article

Content List

Abstract

Background

Method

Figures & Tables

Article Metrics

Are pupils' 'Programme for International Student Assessment (PISA)' scores associated with a nation's fish consumption?

Volker Schmiedel, Hans Vogt, Harald Walach

First Published November 21, 2017 | Research Article | Find in PubMed | Check for updates

<https://doi.org/10.1177/1403494817717834>

Article information

Abstract

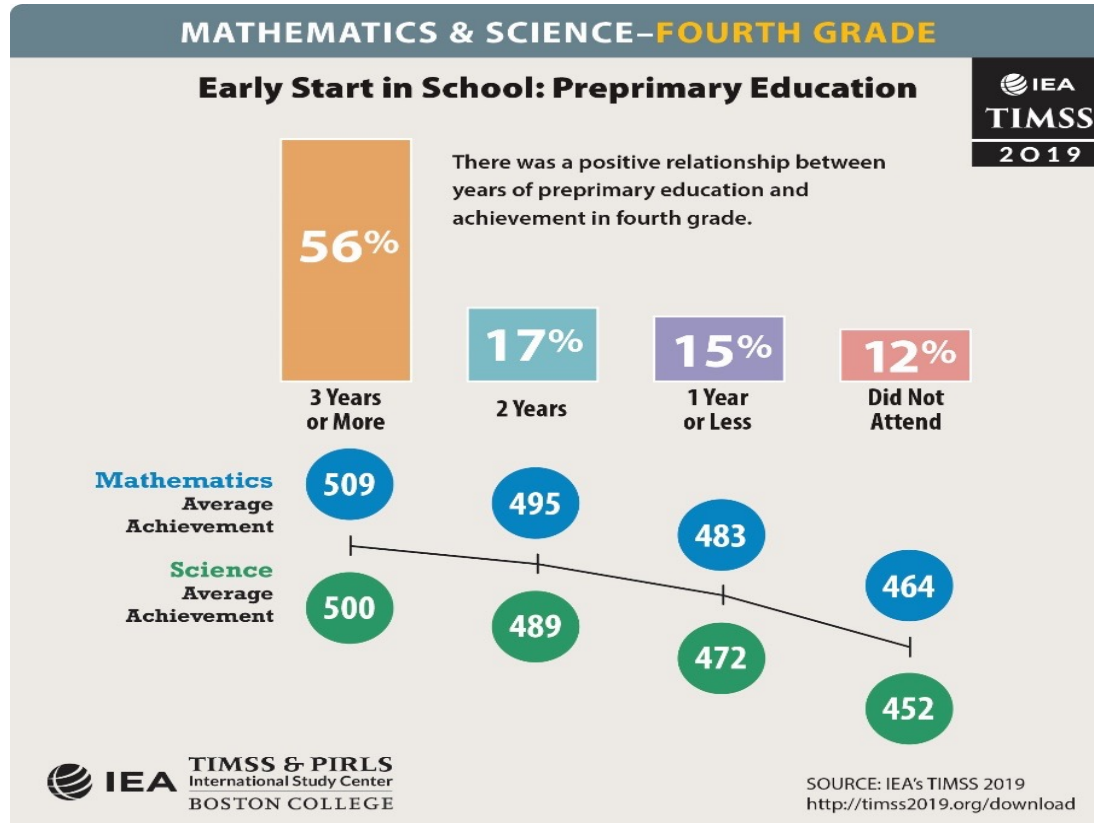
Aims: Omega-3 fatty acids are essential for neurodevelopment. Fish is a major source of these essential nutrients. We asked whether a nation's fish consumption is associated with its pupils' PISA scores as measures of school achievement, independent of economic status and breastfeeding. Method (Design and Setting): This was a regression analytic study based on published data. Results: Data on fish consumption and PISA scores were available for 64 countries. A significant regression model explained 72% of the variance in PISA scores. Breastfeeding does not enter the model. After controlling for technical and/or economic development, a nation's fish consumption remains a significant predictor, explaining an additional 4% of the variance. Discussion: This effect is likely due to the fact that fish, as the major source of omega-3 polyunsaturated fatty acids for a population, is important for the omega-3 supply to mothers and the early neurodevelopment of their children. **Conclusions: Fish consumption, as a proxy for a population's omega-3 supply, is an independent predictor of pupils' school achievement, in addition to a nation's**

Article available
Vol 46, Issue

SAGE Disc Hub
Read the latest content

Related Articles
Similar Articles

...another example

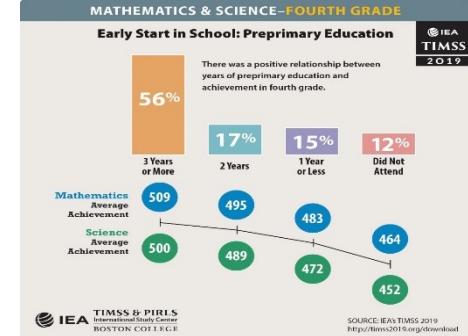


Correlation must not be Confused with Causation

Endogeneity broadly refers to situations in which an explanatory variable (e.g., preschool) is confounded with other predictors of the outcome.

Self-selection or omitted variable bias can cause endogeneity, which may occur at different levels in ILSAs:

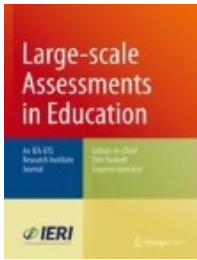
- **Country-level:** Economically developed countries may have more established preschool systems.
- **School-level:** Due to residential segregation, advanced districts may provide better preschool access *and* better schools.
- **Student-level:** Children from privileged families may enter preschool earlier.



Controlling for Confounding?

If all possible confounding factors are known and measured, conditioning is a powerful approach to address endogeneity.

- Common conditioning techniques such as regression, multilevel modeling, or matching 'control for' potential confounding variables.
- “Over the past decades, it has become increasingly obvious in the literature evaluating education policies and practices that there are myriad important factors that remain unobserved in our models, often rendering the attempts to control for all relevant confounding factors in vain.” (Schlotter, Schwerdt & Woessmann, 2009)



Nonetheless, conditioning approaches dominate the analysis of ILSA data. For example, only two out of a total of 21 effectiveness studies published in the IEA-ETS Research Institute Journal, *Large-scale Assessments in Education*, in between 2020 and 2022 did not use conditioning approaches to address the endogeneity issue.

Causal Analysis with ILSA Data

When studying causal effects, a basic distinction can be made between approaches with and without randomization:

1. Randomized experiments
2. New approaches to causal analysis



The pioneering work on methodological contributions to approaches to the analysis of causal relationships was recently recognized with Nobel Prizes awarded to economists Joshua D. Angrist (MIT) and Guido W. Imbens (Stanford).

Randomized Experiment

Randomized control trials (RCTs) are often considered the gold standard to make causal inferences. Two groups formed by randomization will be equal with respect to all characteristics except for the treatment assignment.

While randomization at the level of individuals is a preferred procedure, there are some limitations for studying effects of educational policy and practice:

- Learning takes place in classes/schools, so in many studies the classroom/school would be the unit of randomization.
- To study institutional features of educational systems (e.g., central exit exams, early tracking), the randomization would take place at the country level.

Causal analysis with observational ILSA data

In recent years, there have been important methodological developments within statistics, econometrics, and other disciplines that support causal inference from observational data. Examples are:

- Trend analysis
- Within-student-between-subject approach
- Difference-in-differences
- Instrumental variable regression
- Regression discontinuity

(e.g., Angrist and Pischke, 2009; Foster, 2010; Hernnan and Robins, 2020; Pearl, 2009a, 2009b; Rosenbaum and Rubin, 1983; Schafer and Kang, 2008; Shrouf et al., 2008; VanderWeele, 2015; VanSteelandt and Lange, 2012)

Trend Analysis



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev



Do increases in national-level preschool enrollment increase student achievement? Evidence from international assessments

Rolf Strietholt ^{a, b, c, *}, Nina Hogrebe ^d, Henrik Daae Zachrisson ^e

^a International Association for the Evaluation of Educational Achievement (IEA), Überseering 27, 22927 Hamburg, Germany

^b TU Dortmund University, Center for Research on Education and School Development (IFS), Vogelpothsweg 78, 44227 Dortmund, Germany

^c University of Gothenburg, Department of Education and Special Education, Box 300, 40530 Gothenburg, Sweden

^d University of Applied Science Hamburg, Faculty of Business & Social Sciences, Alexanderstraße 1, 20099 Hamburg, Germany

^e University of Oslo, Department of Special Needs Education, PB 1140, Blindern, 0318 Oslo, Norway

ARTICLE INFO

Keywords:

Academic achievement
Comparative analysis
Early childhood policies
Preschool enrollment
International panel data

ABSTRACT

The main purpose of the study is to analyze whether globally observed trends towards preschool expansion have impacted student achievement in primary and secondary school. We use data from multiple study cycles of two international large-scale assessments that have a longitudinal component at the country level—PIRLS and PISA—and combine these data with a country-level measure of preschool enrollment rates as the main explanatory variable. Employing a multilevel regression with fixed effects for countries and years, we find that changes in preschool enrollment are unrelated to changes in average student achievement. Even after controlling for covariates on the individual and country levels, we do not find any support for the policy expectation that expanding preschool enrollment per se leads to better student achievement on country level.

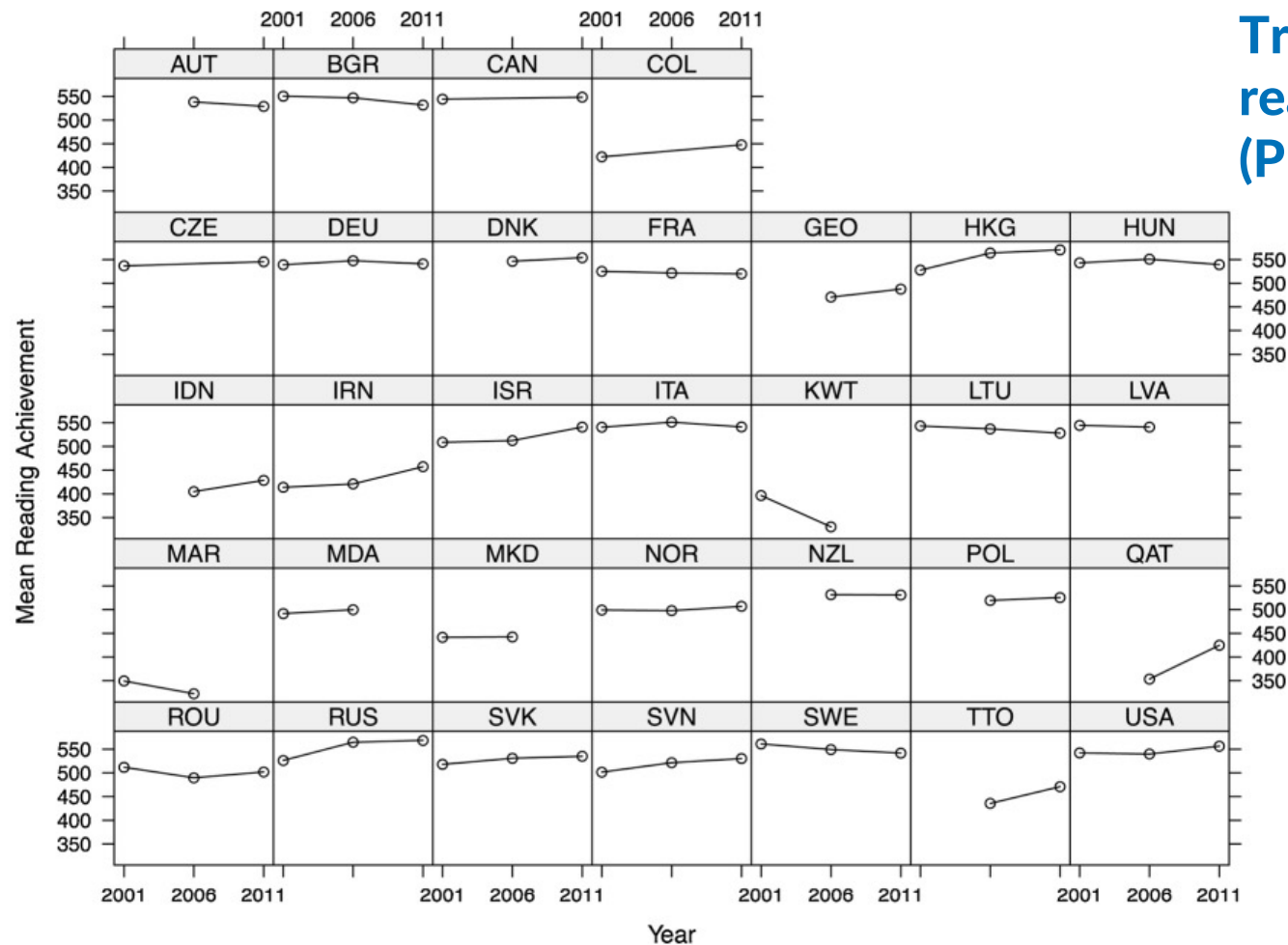
- Does preschool improve cognitive skills of children?
- There is a strong cross-sectional correlation between preschool enrollment rates and achievement.
- There is also an association between the economic level of countries and their preschool enrollment rates.
 - 83% in high-income countries
 - 44% in middle-income countries
 - 17% in low-income countries

Trend Analysis

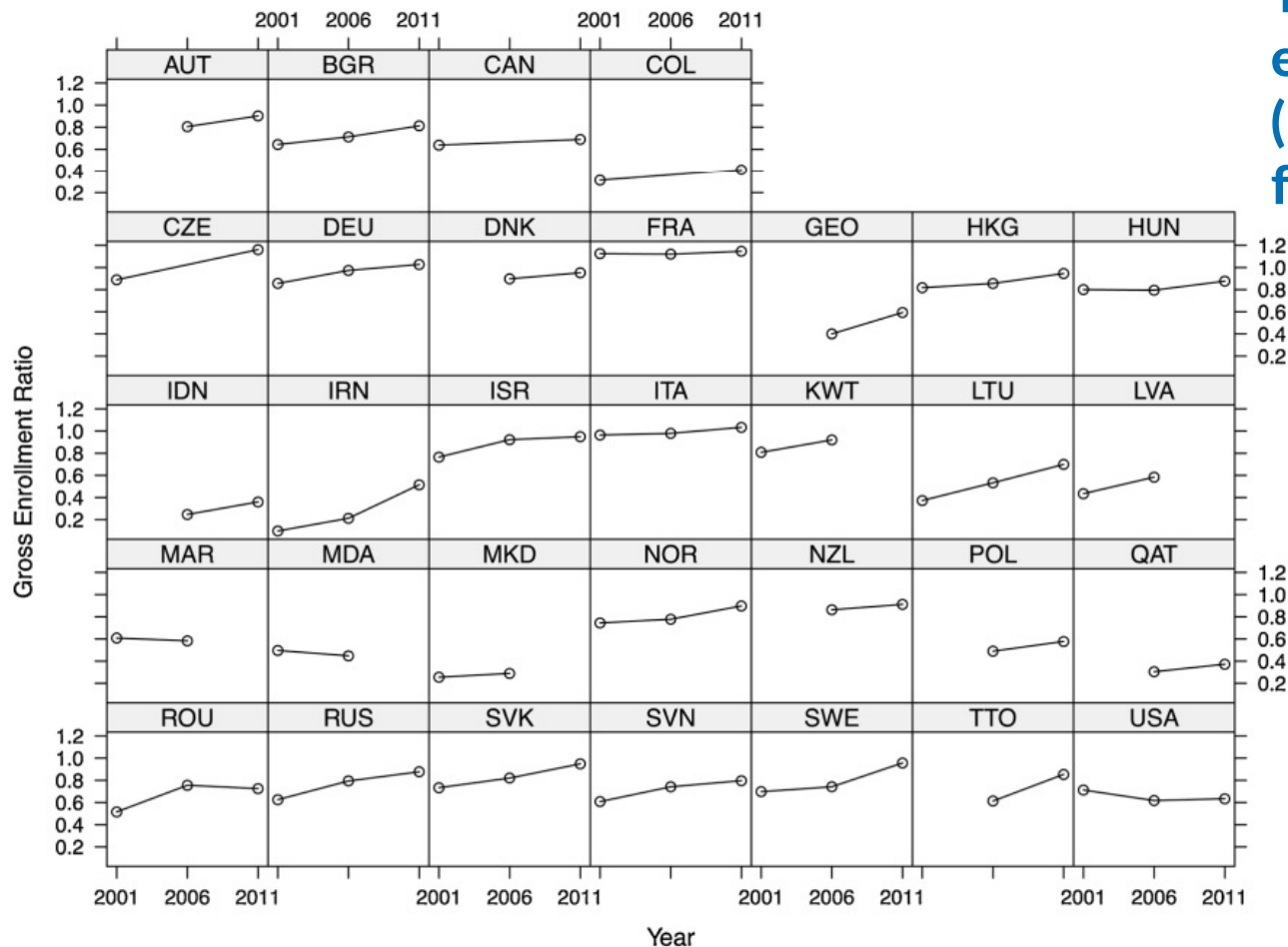
Modern IEA studies are repeated every few years and thus have a longitudinal component at the country level. The study combined data from PIRLS 2001, 2006, and 2011.

- With longitudinal data, we can use variation between repeated observations of the same unit (countries) to test if *change* in the explanatory variable predicts *change* in the outcome.
- Focus on changes within countries circumvents the bias of stable country features as they are the same at each time point. An HLM model with fixed effects for countries and years was used (plus controls at the student level).

Trends in average reading scores (PIRLS)



Trends in preschool enrollment rates (UNESCO Institute for Statistics)



Trend Analysis

Cross-sectional and trend analysis were compared by regressing achievement on national preschool enrollment rates without and with country fixed-effects.

- The cross-sectional analyses suggest that a 10% increase in preschool enrollment rates is associated with an increase of 12.5 points on the PIRLS scale.
- After adding country fixed-effects, this effects vanishes almost completely to 0.7 points.

Within-Students-Between-Subjects Approach



- Do early home learning activities (EHLA) have an effect on later academic achievement?
- The frequency of early literacy and numeracy activities is correlated with reading and math performance at the end of primary school.
- The joint administration of TIMSS and PIRLS in 2011 provides a unique opportunity to address endogeneity issues.

Within-Students-Between-Subjects Approach

A problem in identifying the effect of EHLA on achievement is that well educated parents are on average more engaged in these activities, and likely differ with respect to other determinants of student achievement from less educated parents such as poverty, health, or school choice.

- TIMSS and PIRLS in 2011 not only measure achievement in math and reading but also survey subject-specific explanatory variables.
 - Literacy activities: read books, play with alphabet toys, ...
 - Numeracy activities: count different things, play with number toys, ...
- While well educated parents are more engaged in EHLA overall, the time that is spent on subject-specific early learning activities may differ.
- This approach relies on within-student variation in the provision of EHLA to explain within-student variation in achievement.

Within-Students-Between-Subjects Approach

Both cross-sectional correlations and within-students-between-subjects estimates suggest a positive effect of EHLA on performance, however, cross-sectional correlation seems to overestimate the effect:

- The cross-sectional correlation between EHLA and achievement was about $r=.26$.
- The estimate for the within-students-between-subjects analysis is considerable smaller $r=.07$ but remains statistically significant.
 - This estimate corresponds to the conventional effect size measure of Cohen's $d=.14$.
 - The effect is quite stable in all countries.

Difference-in-Differences

The Economic Journal, **116** (March), C63–C76. © Royal Economic Society 2006. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

DOES EDUCATIONAL TRACKING AFFECT PERFORMANCE AND INEQUALITY? DIFFERENCES-IN-DIFFERENCES EVIDENCE ACROSS COUNTRIES*

Eric A. Hanushek and Ludger Woßmann

Even though some countries track students into differing-ability schools by age 10, others keep their entire secondary-school system comprehensive. To estimate the effects of such institutional differences in the face of country heterogeneity, we employ an international differences-in-differences approach. We identify tracking effects by comparing differences in outcome between primary and secondary school across tracked and non-tracked systems. Six international student assessments provide eight pairs of achievement contrasts for between 18 and 26 cross-country comparisons. The results suggest that early tracking increases educational inequality. While less clear, there is also a tendency for early tracking to reduce mean performance.

Many countries worry about the relative merits of a selective versus comprehensive school system and the resulting system choices are surprisingly different. Some countries track students into differing-ability schools as early as at age 10 (e.g., Austria, Germany, Hungary and the Slovak Republic). By contrast, others including Canada, Japan, Norway, Sweden, the UK and the US essentially keep their entire lower secondary school system comprehensive. Parents and politicians alike would like to know whether it has consequences for the equity and efficiency of educational outcomes if a country tracks its students into different school types, hierarchically structured by performance. Such macro issues of institutional

- Does educational tracking affect performance and inequality?
- Some countries track students into differing-ability schools as early as age 10, while others keep their entire lower secondary school system comprehensive.
- Hanushek & Woessmann took a difference-in-differences (DiD) approach to identify the effect of tracking on performance and inequality.

Difference-in-Differences

Simple cross-sectional comparisons of countries with tracked and comprehensive school systems may be biased because they may also differ on other characteristics.

- Hanushek and Woessmann suggest using primary school data as a baseline, since there is no tracked primary school system.
 - Primary school ILSAs: TIMSS and PIRLS
 - Secondary school ILSAs: TIMSS and PISA
- The basic idea is to compare the differences in inequality between primary and secondary education for tracked and comprehensive systems.

Difference-in-Differences

The direction and magnitude of bias in cross-sectional correlations is difficult to predict.

- A cross-sectional perspective provides no evidence that tracking increases inequality, whereas the DiD estimate suggest a different conclusion.
- The cross-sectional perspective ignores the fact that the early tracking countries were at a much lower inequality level in primary school.



Fig. 1. *Inequality in Primary and Secondary School*

Notes. Standard deviation of test scores in the national population (difference from international average of national standard deviations in each test). Countries with a tracked school system before the age of 16 have solid lines, countries without tracking before age 16 have dashed lines.

Conclusions

The IEA was established to generate knowledge for evidence-based educational policy and practice. For this purpose, studies produce high-quality data from representative samples on important aspects of many educational systems.

- These data are observational data from cross-sectional survey studies, and do not automatically support causal inference. However, new techniques for causal inference have been developed in different disciplines, and these techniques may be useful in educational research as well.
- It's time to move beyond descriptions. Evidence-based policy needs explanations of what work.

Conclusions

- Weak empirical studies - such as simple correlations or regressions with few standard controls - offer at best descriptive but not explanatory knowledge. Descriptions do not provide evidence for the "What Works?" question.
- Multiple cycles of ILSA data as well as combinations of different studies allow for longitudinal analysis and other stronger and more credible research designs.
- Each of these designs relies on certain more or less credible assumptions. Other researchers can criticize, question, and test assumptions if they have been made explicit.

Conclusions

Strategies to promote greater use of existing IEA studies to generate explanatory knowledge on effective educational policy and practice:

- **Capacity building:** methods for causal evaluation based on observational data are still not widely used among educators. Capacity building support in these areas could help to evaluate IEA data using smart and innovative analytical strategies.
- **Dissemination:** existing research is not fostered enough across academic disciplines. Most innovative studies with sophisticated analytical strategies are published in economics journals and such journals receive little attention from educators (unfortunately, this is mutual).
- **Research:** targeted grants for innovative projects that apply the new methodological developments to IEA data could boost their use.

Conclusions

Strategies to promote greater use of future IEA studies to generate explanatory knowledge on effective educational policy and practice:

- **Harmonization of (background) instruments:** With a common set of items (e.g., bullying, school climate), different assessments could be linked.
- **When measuring change, do not change the measure:** A long-term strategy (even beyond two study cycles) for questionnaires could address different objectives in parallel.
- **Joint administration of studies:** Following the example of PIRLS and TIMSS, future studies could be administered together.
- **Design changes:** Repeated observations of students in panel studies or even experimental components could facilitate explanatory research.

Beyond descriptions: International assessments and evidence-based policy

GA-63, Split, Croatia

Oct 04, 2022

Rolf Strietholt, rolf.strietholt@iea-hamburg.de